# Proceedings of the 13'th SweCog Conference

**Book** · October 2017



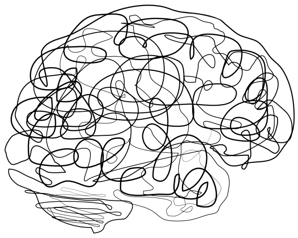
**4 authors:**

Anders Jansson
Uppsala University


Carl-Anton Werner Axelsson
Uppsala University


Rebecca Cort
Uppsala University


Erik Alexander Billing
University of Skövde



# PROCEEDINGS OF THE 13TH SWECOG CONFERENCE

Uppsala 2017
26 – 27 October

Editors
Anders Arweström Jansson
Anton Axelsson
Rebecca Andreasson
Erik Billing

SWEDISH COGNITIVE SCIENCE SOCIETY
UPPSALA UNIVERSITET

**Editors:**


| | | |
|---|---|---|
| Anders Arweström Jansson | Uppsala University | anders.arwestrom.jansson@it.uu.se |
| Anton Axelsson | Uppsala University | anton.axelsson@it.uu.se |
| Rebecca Andreasson | Uppsala University | rebecca.andreasson@it.uu.se |
| Erik Billing | University of Skövde | erik.billing@his.se |



Cover illustration and design by Anton Axelsson

Skövde University Studies in Informatics 2017:2
ISBN 978-91-983667-2-3
ISSN 1653-2325

PUBLISHED BY THE UNIVERSITY OF SKÖVDE

UNIVERSITY OF SKÖVDE

# PROCEEDINGS OF THE 13TH SWECOG CONFERENCE

# Contents

# Preface

Welcome to SweCog 2017!

This booklet contains the abstracts and short papers for all oral and poster presentations at the 2017 SweCog conference. Following the SweCog tradition, with the aim to support networking among researchers in Sweden, contributions cover a wide spectrum of cognitive science research.

This year, a large proportion of contributions discuss one of the core concepts in cognitive science – *representations*. Fundamental problems with many models of representations are discussed and analysed. Several contributions put forward embodied and enactive views of mental representations and memory, with links to meaning and the mind-matter dichotomy. Representations are also touched upon in studies of attention, concepts, critical thinking, curiosity, intentionality, imitation, optimism, and working memory. We are excited to see such a flora of contributions around this very important and sometimes problematic cornerstone in the field.

In addition to the studies mentioned above, we also see contributions in computational linguistics, distributed cognition, embodiment effects on film and video experience, e-sports, haptic communication, learning by teaching, and robot assisted therapy.

We look forward to a few exciting days in Uppsala and we thank the many people that have contributed to this conference, in particular all authors and reviewers.

*Erik Billing*

## Reviewers

| | |
|---|---|
| Jens Allwood | *University of Gothenburg* |
| Alexander Almér | *University of Gothenburg* |
| Rebecca Andreasson | *Uppsala University* |
| Anders Arweström Jansson | *Uppsala University* |
| Anton Axelsson | *Uppsala University* |
| Erik Billing | *University of Skövde* |
| Åsa Cajander | *Uppsala University* |
| Diane Golay | *Uppsala University* |
| Mikael Laaksoharju | *Uppsala University* |
| Mats Lind | *Uppsala University* |
| Jessica Lindblom | *University of Skövde* |
| Rob Lowe | *University of Gothenburg* |
| Ida Löscher | *Uppsala University* |
| Jonas Moll | *Uppsala University* |
| Gerolf Nauwerck | *Uppsala University* |
| Maike Paetzel | *Uppsala University* |
| Anders Persson | *Uppsala University* |
| Fredrik Stjernberg | *Linköping University* |
| Johan Stymne | *Stockholm University* |
| Tarja Susi | *University of Skövde* |
| Niklas Torstensson | *University of Skövde* |
| Ronald van den Berg | *Uppsala University* |
| Annika Wallin | *Lund University* |

# Conference Programme

**Thursday October 26th Assembly Hall (Building 6)**

| | |
|---|---|
| 12:00 — 13:00 | Registration in Building 6 |
| 13:00 — 13:15 | Welcome |
| 13:15 — 14:00 | **Invited speaker** - Mark Bickhard<br>*Representing is (for) What?* |
| 14:00 — 14:30 | Kata Szita<br>*Smartphones, films, and cognition* |
| 14:30 — 15:00 | Coffee |
| 15:00 — 15:30 | Erik Lagerstedt<br>*A drive through the world of functional tones, simulations and cars* |
| 15:30 — 16:00 | Ali Basirat and Marc Tang<br>*Neural network and human cognition: A case study of grammatical gender in Swedish* |
| 16:00 — 16:15 | Break |
| 16:15 — 17:00 | **Invited speaker** - John Flach<br>*Meaning Processing in a Triadic Semiotic System* |
| 17:00 — 19:00 | Poster session including refreshments |
| 19:30 | Dinner on board m/s Kung Carl Gustaf |

## Friday October 27th Assembly Hall (Building 6)

| | |
|---|---|
| 09:15 — 10:00 | **Invited speaker** - Ulrike Hahn<br>*Pessimism about optimistic belief updating* |
| 10:00 — 10:30 | Coffee |
| 10:30 — 11:00 | Joel Parthemore<br>*On the essentially dynamic nature of concepts* |
| 11:00 — 11:30 | Manuel Oliva<br>*Pupil dilation reflects the time course of perceptual emotion selection* |
| 11:30 — 12:00 | Mikael Laaksoharju<br>*Adding integral display properties to increase generalizability of a configural display* |
| 12:00 — 13:00 | Lunch |
| 13:00 — 13:30 | Anna Einarsson<br>*Composing music as an embodied activity* |
| 13:30 — 14:00 | Gabriela-Alina Sauciuc<br>*The social side of imitation in human evolution and development: Shared intentionality and imitation games in chimpanzees and 6-month old infants* |
| 14:00 — 14:30 | Coffee |
| 14:30 — 15:00 | Ronald van den Berg<br>*An ecologically rational explanation for set size effects in human cognition* |
| 15:00 — 15:30 | Christopher Summerfield<br>*Human blindness to noise in neural computation* |
| 15:30 — 16:00 | SweCog annual member's meeting<br>and conference closing |

# Invited Speakers

## Representing is (for) What?
### *Mark Bickhard*

**Abstract**: Available models of representation suffer from fatal problems, some extending back millennia in Western thought, and some introduced more recently (I have made my own contributions to this family of critiques). But representing will not go away; what is necessary is a different kind of model. I will outline an action based, pragmatist model of representing that avoids the family of problems of classic models, and show how representing emerges naturally and necessarily in the evolution of agents. As cognition and representing permeate everything mental, so also do the consequences of this shift in models of representing.

## Meaning Processing in a Triadic Semiotic System
### *John Flach*

**Abstract**: Weinberg (1975) defined "system" as "a way of looking at the world." That is, the specification of the system reflects an ontological choice that all scientists make to distinguish between the 'objects of interest' and the 'background' for their particular field of study. In this talk, I will defend my choice to identify the cognitive system as a Triadic Semiotic System that spans mind and matter. I will argue that meaning emerges from functional relations associated with a closed-loop coupling of situations and awareness. I will suggest that to fully understand this coupling, we need constructs that span the mind-matter dichotomy and that are compatible with the dynamics of circular systems. I will suggest three important constructs that I believe are essential to understanding the circular dynamics of human experience: satisfying, specifying, and affording.

## Pessimism about optimistic belief updating
### *Ulrike Hahn*

**Abstract**: Decades of research seemingly established robust evidence for an "optimism bias" whereby people think 'bad things only happen to others'. The empirical basis of this putative bias came under scrutiny with Harris and Hahn's (2011) critique that showed the standard method for showing unrealistic optimism, does this even for entirely rational, non-optimistic agents. In the same year, work by Sharon et al. (2011) introduced a potential new mechanism, and with it new evidence for unrealistic optimism, into the debate. The talk will demonstrate that this method is prone to showing 'optimism' in entirely rational agents also, and does not yield interpretable results.

# Oral Presentations

# Composing music as an embodied activity

**Anna Einarsson**[1], Tom Ziemke[2]

[1]*Royal College of Music, Stockholm*

[2]*Linköping University, Linköping*

*annaeinarssonmusic@gmail.com*

Composing music may be regarded as a rather ephemeral or disembodied activity. Quite the contrary though, drawing upon experiences from being a professional composer, and the artistic works and accounts from participating singers included in the thesis *Singing the body electric – Understanding the role of embodiment in performing and composing interactive music* (Einarsson, 2017), this presentation puts forward how musical composition may be understood as an embodied practice, relying on mechanisms of *embodied simulation* and *affective appraisal*. "You are the music, while it lasts", as the famous poem by T.S Elliot reads. Differently put, decision making while composing music can be said to happen through a process of *resonance*, where music may be seen as a sounding body to resonate with. Performing the music tacitly on an "inner stage", listening and sensing, are cornerstones in this process. For example, this concerns working with the musical structure, balancing the whole against the parts in relation to the unfolding situation. The situation encompasses not only the here and now, but also the materials and tools employed as well as the sociocultural embeddedness of institutions and peoples involved in the music performance.

The concept of resonance is a process of embodiment, incorporating emotion and feeling. Hence working with an artistic idea for a music composition can be seen as having an *affective bearing* towards which the artistic course for the work is set. This bearing constitutes an underlying principle guiding the compositional choices. The composer, as suggested in the thesis, attempts to using his/her own body as a template when shaping and listening to the work in progress, making use of embodied simulation in order to work with expectations, pivotal to musical composition, and directing the attention of as well performers as audience as the work proceeds. In a similar vein, composing opera (or other works that incorporate text) is not a matter primarily of establishing the semantics of the words, but working with the feeling the text evokes.

This presentation focuses mainly on experiences from the work *Metamorphoses* (2015) by Einarsson, a work that explores ideas of transformations and embodiment, primarily in terms of conditions for corporeality when interacting with responsive technology. It is a work for four singers and four responsive computer systems (see Einarsson 2017b). When investigating the role of the composer, the primary method has been autoethnographic – writing and later analysing process diaries from the process of composing. The process diary primarily serves as a site for reflection, contributing to an iteration between the making and the reflecting, and it only to lesser degree contains notes on more formal research procedures.

Some examples from the process diary I will discuss are:

> *"The whole time I'm checking against my experience of the sound, of singing with the patch [computer program], against the feeling it gives when I listen."*

> *"The first thing for today is to listen: try to listen through/beyond the mechanical Sibelius files. I sing in parallel with the internal scene, try to understand where the work wants to go, where it's heading, in a negotiation with how I want it to appear. What is missing? Which is the development?"*

> *"It is a matter of finding resonance between a bodily state and the sounding, and to navigate this flow of exchange between sounding material and the bodily states they induce or have originated from. It is a matter of feeling impulses that attract or repel, and being attentive towards a pendulum movement between the whole and its parts, towards the unfolding situation."*

> *"Perhaps this can be called setting off on a treasure hunt? Stumbling over a patch [computer program] that, after a bit of tweaking, generates a splashing rhythmical sound that I immediately take to. Funny - how you*

*know at once when it's right. Now, that was not what I was out after, but it was the rhythm I went for. I feel it in the body; the desire to sing, to test and sometimes to add on."*

Emphasising the embodied aspects of the compositional practice is not a matter of denying rationalisation processes or structured approaches when composing, but rather saying, "certain aspects of the process of emotion and feeling are indispensable for rationality" (Damasio 1994: xii-xiii). The method suggests that constituent of resonance is a susceptibility towards the situation, possibly a form of covert mimicry, allowing for the dynamic contours of the sounding to be apprehended. A similar notion may be found in Daniel Stern's idea of "*vitality affects"* (Stern in Johnson, 2007, p. 55). He puts forward that starting from early childhood and onwards we develop a sense for the contour or feeling of flow of our experiences. In addition, it is reasonable to believe that the process of resonance presupposes some sort of "willingness" to get engaged.

## References

Damasio, A.R. (1994). *Descartes' error: emotion, reason and the human brain*. London: Picador.

Einarsson, A. (2017). *Singing the body electric – Understanding the role of embodiment in performing and composing interactive music*. (Doctoral Thesis, Lund University, Lund, ISBN 978-91-7753-260-6, Doctoral studies and research in fine and performing arts 18, ISSN 1653-8617, http://portal.research.lu.se/portal/en/publications/singing-the-body-electric(b5d21536-08d9-42a4-a79f-4321999e371a).html

Einarsson, A. (2015). *Metamorphoses*. [Video recording]. http://www.annaeinarsson.com/#video

Einarsson, A. (2017b). Experiencing responsive technology in a mixed work: Interactive music as embodied and situated activity. *Organised Sound*, in press.

Johnson, M. (2007). *The meaning of the body: Aesthetics of human understanding*. University of Chicago Press.

# Adding integral display properties to increase generalizability of a configural display

**Mikael Laaksoharju**, Mats Lind, Anders A. Jansson

*Department of Information Technology, Uppsala University*

*mikael.laaksoharju@it.uu.se*

The two main objectives of user interfaces (UIs) are 1) to present data so that humans can make use of it and 2) to make it possible for humans to control computer-mediated processes, for instance by manipulating data or steering mechanical processes. These objectives can surely be qualified in many ways and there can be different higher-order objectives that affect the design of a computer system, like increasing safety, efficiency, productivity, quality and motivation, yet the main reason for creating a UI for a system is for that to function as an effective mediator between humans and computers. A slightly different way of expressing the above objectives is to allow humans to be in the loop in technology-mediated processes.

One critical scenario in which it is important to keep humans in the loop is when it is not possible to fully automate a process; when there is a risk that a decision needs to be made that requires knowledge – contextual, situated, synthetic, or empirical – that is not available at the time of system design. In such scenarios, it seems reasonable to claim that the humans should have the best possible conditions for acting correctly, which would imply that UI design is about creating the best possible conditions for making and implementing good decisions. In this presentation, we focus on the first objective of UIs, i.e., effective presentation of data for making good decisions.

With this in mind, let us turn to a series of studies of graphical representations of data. In 1989 Coury et al. (1989) devised an experiment to test the efficiency and effectiveness of different ways of visualizing numerical system data, for instance from sensors. In the experiment, three different displays of data – alphanumeric, bar graph, and polar graph – were compared in a basic classification task requiring systematic comparisons between four variables. The results suggest that for tasks that require integration of variables, the ubiquitous bar graph was both fastest to interpret and led to fewest errors, followed by the polar graph. Many years later, Holt et al. (2015) made a replication study in which they added a new type of display, which they named configural coordinate display (CCD, see Figure 1 for examples). The results of this study suggest that for the same task, the newly introduced display type was faster to interpret and lead to more accurate interpretations than all the other display types. Considering that the display was designed to ease the specific classification task that was given in the experiment, this should not be surprising.

Based on the introductory reasoning, we became curious about how well the CCD would work in classification tasks for which it was not specifically designed. The basis for this curiosity is the observation that many real-world systems can enter a state that is not accounted for by the designers of the system, and for a representation to be resilient, it needs to work also in unforeseen cases. After all, the reason for requiring a human to interpret data is because the monitored processes have not been possible to fully automate.

In a new replication study we compared three different displays. We used the two displays from the earlier studies that had led to best performance, i.e. the bar graph and the CCD, and introduced a new type of display, intended to combine the configural properties of the CCD while retaining generalizability by adding integral (object-forming) properties that correspond in a more unbiased way to the data (CID, see Figure 2).

26 study participants were randomly assigned to one of three conditions, i.e., one of the three types of graphical displays in Figures 1 – 3, and performed two tasks with the same type of display. In both tasks, the participants used the number keys 1 – 4 to indicate which state of a fictitious system the display represented. They first trained by judging the states in 96 displays, presented in sets of 8, during which the correct answers were provided immediately after a response. Following the training, the same 96 displays were presented one by one in a different order, without feedback. This procedure was then repeated with a new set of displays. In total, the participants trained on 192 displays for each of the two tasks, and were subsequently tested on the same 192 displays (96 training, 96 test, 96 training, 96 test per task).

## Task 1 – partial replication

With task 1 we sought both to replicate the findings of Holt et al. and to determine how much worse the performance with the CID was compared to with the CCD in the particular task that he CCD was optimized for. The states were defined by pairwise comparisons of variables (see Coury et al. 1989 for details). The relevant configural property in CCD is that the state equals the quadrant in which the dot is placed. This configural property is retained in the CID by the center of gravity of the dark-gray box representing the state in task 1. The added integral property is the shape of the dark-gray rectangle, which represents different relations between variables.

### Task 2 – test of generalizability

To test the generalizability of the different types of displays, a different relational state definition was chosen for task 2. The state was again determined by pairwise comparisons of variables: first judging which combination had the greater sum, and then whether the difference was small or large. The four different possible states are represented in Figures 1, 2 and 3, which also illustrate that the placement of the dot in the CCD is not sufficient to determine the state in task 2. The introduction of ambiguity between states, where states 2 and 4 are defined as one of the sums being more than twice the size of the other, was intended to test whether a numerically imprecise visualization like the CID would lead to worse accuracy in magnitude assessment than the more precise bar chart.

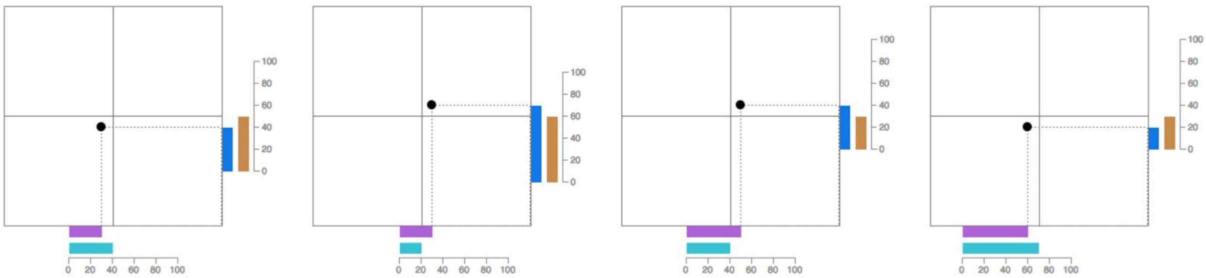



*Figure 1: The configural coordinate display (CCD) introduced by Holt et al. These examples show the states for task 2, ordered from left (1) to right (4). The color scheme has been inverted for print.*

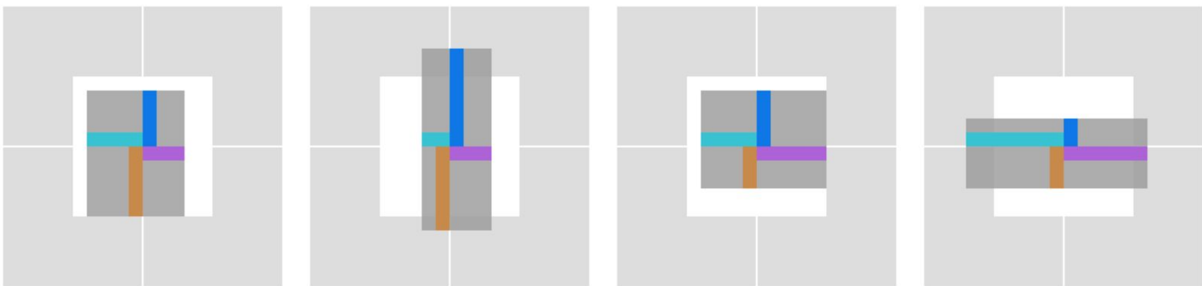


*Figure 2: The configural integral display (CID). These examples show the states for task 2, ordered from left (1) to right (4). The color scheme has been inverted for print.*

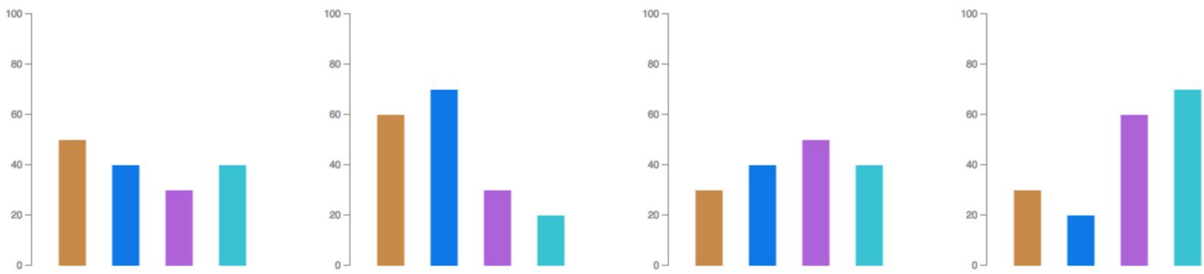



*Figure 3: The bar chart display used in the studies. These examples show the states for task 2, ordered from left (1) to right (4). The color scheme has been inverted for print.*

## Results

Despite a slightly different statistical analysis (using medians instead of means), the results in task 1 are largely consistent with the results from the replicated studies (see Figure 4). The median latency and accuracy for the CCD are very close to the corresponding averages in the study by Holt et al. The median latency for the bar chart display is lower than in both previous studies while the accuracy is higher than i Holt et al., both differences likely due to increased training. The latency for CID is slightly higher than for CCD (significant difference at .05 level).

The results in task 2 show that the greatest performance loss happened for participants using the CCD. This supports the suspicion that the configural optimization of the CCD for the first task indeed has consequences for its generalizability. The bar chart and the CID had latencies comparable to task 1 but worse accuracy. An analysis of responses reveal that most of the decrease of accuracy can be attributed to the ambiguity that was introduced, i.e., to the magnitude assessment between states 1/2, and 3/4.

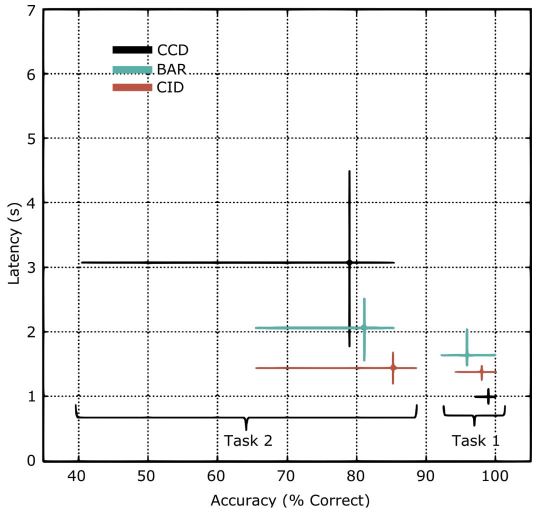



*Figure 4: Accuracy vs. median latency for the two experimental tasks. The lines are drawn from the 25th to the 75th percentiles.*

## Discussion

The CID is similar to a polar graph display with four variables, as the bars originate from a shared origo and are displayed along the horizontal and vertical axes. However, we argue that an important difference is that the gestalt of the CID is simpler to perceive than the gestalt of the polar graph. Instead of relations being represented by irregularly slanted diamond shapes where angles are meaningful, the CID conveys the same information by combining transformation and translation of rectangles. Instead of having to learn the interpretation of angles, two independent properties can be combined to interpret equally many states. The shape of the rectangle can be stretched horizontally and/or vertically, and displaced both horizontally and vertically. When searching visually in a map of graphs, the observer can focus on finding either graphs with a certain elongation or graphs with a certain displacement.

Task 2 may appear as if it was deliberately chosen because it breaks the CCD's shortcut to determine the state of the system. This, however, happens for all state classification tasks except for the one that it was specifically designed to simplify. The important point here is that if the heuristic shortcut can break down, it means that the designer has to know in advance exactly what is important for an observer. If this is the case – that it is possible to use a simple, fail-safe heuristic to assess the system state – one may wonder why the resulting decision task is delegated to humans at all instead of being fully automated. The human contribution becomes valuable when a decision does not follow deterministically from an observation, which for a designer of a display can be translated to when she does not know in advance what data is important to accentuate. This uncertainty alone motivates the adoption of display types that represent data in an unbiased fashion.

## Acknowledgement

The study that this presentation builds on was conducted in collaboration with Johanna Löfvenberg and is included as a case study in her Master's thesis (Löfvenberg, 2016).

## References

Coury, B. G., Boulette, M. D., & Smith, R. A. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of system status. *Human Factors*, *31*(5), 551-569.

Holt, J., Bennett, K. B., & Flach, J. M. (2015). Emergent features and perceptual objects: re-examining fundamental principles in analogical display design. *Ergonomics*, *58*(12), 1960-1973.

Löfvenberg, J. (2016). *Poietic design : Heuristics and applications* (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-297084

# A drive through the world of functional tones, simulations and cars.

**Erik Lagerstedt**[1], Henrik Svensson[2]

[1,2]: *Interaction Lab, School of Informatics, University of Skövde*

*erik.lagerstedt@his.se*

In this presentation we explore the relation between perception and cognition by exploring the relation between the concepts of functional tones (von Uexküll, 1957) and simulation (e.g Barsalou, 1999; Hesslow, 2002; Möller, 1999). Additionally we argue that the automotive domain provides a unique testing ground for these theories.

Von Uexküll (1957) promotes the idea that perception and action are fundamentally connected, especially through functional cycles where perception and action continuously and mutually adjust and react to each other. The properties that are perceived are called functional tones, and dependant on not only the morphology of the perceiver, but also on their history, experiences and even mood. As an example, von Uexküll uses a hermit crab which is perceiving a sea anemone. When the hermit crab is depraved of food or its shell, the sea anemone assumes a "feeding tone" or a "dwelling tone" respectively. Functional tones are in some regards similar to affordances (cf. Gibson, 1979), but are also different in some ways (Susi & Ziemke, 2005), which are relevant for the current purpose. For example, "[t]he affordance of something does not change as the need of the observer changes. The observer may or may not perceive or attend to the affordance, according to his needs, but the affordance, being invariant, is always there to be perceived. An affordance is not bestowed upon an object by a need of an observer and his act of perceiving it. The object offers what it does because it is what it is" (Gibson, 1979, p. 130).

The view of perception as simulation shares a similar emphasis on the interconnectedness of perception and action in perception. Perception as simulation depicts perception as based on anticipatory simulations (Gross, Heinze, Seiler, & Stephan, 1999; Möller, 1999; cf. also Jordan, 1998). Perception and action generation are suggested to be part of one and the same (neural) process, rather than making the traditional division of perception, cognition, and action (Gross et al., 1999; Möller, 1999). Perception is not merely a passive transformation of information but an active effort to control the inputs or stimuli of the agent (cf. e.g. Varela, Thompson, & Rosch, 1991). According to this view of perception as simulation, perception is supposed to be a process that generates sequences of sensorimotor hypotheses. The sensorimotor hypotheses themselves are internal simulations that anticipate future situations that would result from the execution of different motor commands, without actually executing these actions (Möller, 1999)[1]. From the set of internally generated sensorimotor sequences the action associated with a favourable outcome is selected (Gross et al., 1999; Möller, 1999). Furthermore, in a real neural system, the same neurons will be involved in the representation of real sensory and motor signals and the sequences of sensorimotor hypotheses (Gross et al., 1999; Möller, 1999). Möller (1999, p. 171) summarised the main points of perception through anticipation as follows: "Perception of space and shape is based on the anticipation of the sensory consequences of actions that could be performed by the agent, starting from the current sensory situation. Perception and the generation of behaviour are two aspects of one and the same (neural) process". Hence, just as emphasised by von Uexküll, only a small part of an agent's control system can be characterised as "purely sensory" or "purely motoric" (cf. Möller, 1999). Instead, the main part of the system integrates information from different sensory modalities and motor information. From the perspective of perception as simulation it is obvious that motor "information" is not only the output of the system, but is as much input to the system as perception is since it is an essential part of the simulation process.

A possible difference from non-representational approaches such as Gibson (1979) and his notion of affordance is, in the words of Möller, that this "approach does not deny the existence of representation in general, but only replaces sensory with sensorimotor representations. The 'utility' of objects is not directly 'offered' by the external world, but determined by the generation of sensorimotor hypotheses based on the sensory input" (Möller, 1999, p. 186). While Gibson's notion of affordance might have been more directed to the cues offered by the external world, von Uexküll's somewhat similar concept of functional tones, however, puts emphasis on the sensorimotor capabilities of the agent. As mentioned above, the functional tones that an agent will perceive are inseparably dependent on the particularities of that agent, and they fill a role in the functional cycles of the agent. The sensorimotor capabilities of the agent will also determine its Umwelt (approximately the subjective surrounding of the agent). To quote von Uexküll; "all animals, from the simplest to the most complex, are fitted

[1]These sensorimotor hypothesis or simulations have also been suggested to be a general principle of cognition (e.g. Hesslow, 2002, 2012) able to support other phenomena of cognition such as mental imagery, prospection, problem solving, and dreams (for a comprehensive review Svensson, 2013).

into their unique worlds with equal completeness" (von Uexküll, 1957, p. 11). Furthermore von Uexküll (1957) argues that an "(…) animal grasps its objects with two arms of a forceps, receptor, and effector". That is, the animal "sees" the objects through its perceptions of and actions toward this object, although, as pointed out above, it is not the actual action or perception but the perceptual and action signs/cues that shapes the object for the animal. We believe that an important claim that von Uexküll makes is that the perceptual world and the effector world sometimes influences the meaning of the other, at least that the effector signs/cues sometimes changes the perceptual cues. This is illustrated by the anecdote where a man, who does not know what a ladder is, instead sees a collection of rods and holes until he is shown how to properly use the ladder (von Uexküll, 1957, p. 48). We propose here the hypothesis that the mechanism behind functional tones could very well be the same as that suggested by simulation theories. By the anticipatory simulations the agent automatically executes in parallel several possible interactions with the external world that adds what von Uexküll calls a functional or effector tone. The object acquires a new meaning, but it might be an additional meaning, the other meanings provided by the earlier perceptual cues might still be visible (in other functional circles) e.g., the man previously ignorant of ladders might still be able to see the rods and holes as well as the ladder. The examples of functional cycles that von Uexküll provides are mainly reactive systems, however, through simulations they should be generalisable to anticipatory systems. Worth noting is that even though functional cycles are reflexive in nature, and that they can chain together into more complex behaviours, von Uexküll was very clear in his opposition of the idea that all animals simply are some kind of reflexive machines (von Uexküll, 1957). He pointed out that functional cycles are only triggered under a select few and very specific conditions; when the animal perceived the correct functional tone.

Cars can be seen as artificial agents already existing in the "real world". They are, however, still controlled to a large degree by human drivers. To improve their potential autonomy, it is worth enhancing them with existing solutions found in natural agents (i.e. animals). We have proposed that functional tones and cycles, in combination with simulation theory, can be useful when describing behaviour of natural agents. They are thus candidates for guiding the development of cars with higher autonomy. Dreams4Cars is a three year EU Horizon 2020 project (2017-2019) that aim for increasing the abilities of "self-driving" cars by constructing an offline simulation mechanism in which cars, by recombining aspects of real-world experience, can produce a simulated world, with which they can collectively interact to safely develop and improve their behaviour. The project takes inspiration directly from simulation theories, especially the notion of agents using predictions to improve their behaviour in different ways. The simulation theory has previously been tested in simulated robot experiments (e.g. Hoffman & Möller, 2004; Ziemke, Jirenhed & Hesslow, 2005) with simple robots, environments, and tasks that only superficially resemble the everyday human tasks and environments. The road vehicle domain, however, can serve as an appropriate setting for testing hypotheses regarding agents' interaction with their environment. The autonomous vehicles are, similar to animals, situated in an environment where the morphological particularities determine aspects, such as, what terrain is possible to cross and what features will serve as obstacles. It shares this environment, which is dynamically changing (e.g. temperature and light conditions), with other independent agents, which may or may not be relevant for the car. A difference between cars and animals, is that cars, and to an extent their environment, are designed and created by humans. This provides more options and control when it comes to hypothesis testing.

In more detail, the dream4cars architecture will consist of an on-line and off-line mode system integrating several different cognitively inspired mechanisms, including simulation. In the online mode a "dorsal loop" mechanism proposes several action possibilities in parallel at various levels of abstraction from the perceptual situation (cf. Cisek, 2007; Windridge, 2017), which in combination with a mid- and low level controller determines longitudinal and lateral control. While Cisek (2007) used the concept of affordance, we believe that the concept of functional tones may be the better concept to describe interactions by integrating more agent centric aspects such as morphology. The online system also includes forward models (Porrill, Dean & Anderson, 2012), which also figures prominently in the view of perception/cognition as simulation. Forward models are internal models that predict future states of the system. For example, the cerebellum is thought to implement forward models that based on copies of motor commands predicts the sensory consequences of those motor commands. Forward models in the on-line mode can be used to detect sensor failures and to identify salient situations to be re-used for learning in the off-line mode (for the possible functions of forward models in the human brain, see Porrill et al., 2012). In the off-line mode previous experiences will be recombined such that the car can "dream" of its behaviour in these situations (the final dream scenarios are realised in the OpenDS car simulator). The driving agent will be trained using a combination of learning schemes, such as "motivated learning" (cf. Gurney, et al., 2013) and "motor babbling" (cf. Windridge, 2017). The design of the Dreams4Cars architecture is thus related to several aspects of human perception, for example, how do higher-level (more abstract) simulations relate to lower-level sensorimotor simulation mechanisms, and how do sensorimotor capabilities of autonomous agents affect its perceptual abilities.

In summary, we have proposed that perception is not the passive information passing from a stimulus to response, but perception, action, and cognition are intertwined in complex ways. They should thus be viewed as different perspectives of the same overall process of generating behaviour, in which sensorimotor processes are part of the cognitive process and the cognitive/motor systems are part of the perceptual process itself. We acknowledge that many details of the relation between functional tones and simulations are yet to be discovered, but we have indicated that it is a line of inquiry worth pursuing.

*Acknowledgement. This work has been supported by funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731593.*

**References**

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(04), 577–660.

Cisek, P. (2007) Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 362* (1485), 1585–1599.

Gross, H.-M., Heinze, A., Seiler, T., & Stephan, V. (1999). Generative character of perception: a neural architecture for sensorimotor anticipation. *Neural Networks*, *12*(7–8), 1101–1129. https://doi.org/10.1016/S0893-6080(99)00047-7

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, *6*(6), 242–247. https://doi.org/10.1016/S1364-6613(02)01913-7

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research*, *1428*, 71–9. https://doi.org/10.1016/j.brainres.2011.06.026

Hoffmann, H. & Möller, R. (2004) Action selection and mental transformation based on a chain of forward models. In: S. Schaal, A. J. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam & J.-A. Meyer (Eds.), *From Animals to Animats 8* (pp. 213-222). Cambridge, MA: MIT Press.

Gibson, J. J. (2015). The ecological approach to visual perception. Psychology Press. (Original work published 1979)

Gurney, K., Lepora, N., Shah, A., Koene, A., & Redgrave, P. (2013). Action Discovery and Intrinsic Motivation: A Biologically Constrained Formalisation. In G. Baldassarre & M. Mirolli (Eds.), *Intrinsically Motivated Learning in Natural and Artificial Systems* (pp. 151–181). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32375-1_7

Jordan, J. S. (1998). Recasting Dewey's critique of the reflex-arc concept via a theory of anticipatory consciousness: implications for theories of perception. *New Ideas in Psychology*, *16*(3), 165–187. https://doi.org/10.1016/S0732-118X(98)00009-9

Möller, R. (1999). Perception Through Anticipation. A Behaviour-Based Approach to Visual Perception. In A. Riegler, M. Peschl, & A. von Stein (Eds.), *Understanding Representation in the Cognitive Sciences* (pp. 169–176). Springer US. Retrieved from http://dx.doi.org/10.1007/978-0-585-29605-0_19

Porrill, J., Dean, P., & Anderson, S. R. (2012). Adaptive filters and internal models: Multilevel description of cerebellar function. Neural Networks, 47, 134–149.

Susi, T., & Ziemke, T. (2005). On the subject of objects: Four views on object perception and tool use. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, *3*(2), 6–19.

Svensson, H. (2013). *Simulations*. Linköping: Linköping University Electronic Press. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:658266

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied mind : cognitive science and human experience*. Cambridge: MIT Press.

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed.), Instinctive behavior – the development of a modern concept (pp. 5–80). New York: International University Press, Inc. (Original work published 1934)

Windridge, D. (2017). Emergent Intentionality in Perception-Action Subsumption Hierarchies. Frontiers in Robotics and AI, 4. https://doi.org/10.3389/frobt.2017.00038

Ziemke, T., Jirenhed, D.-A. & Hesslow, G. (2005) Internal simulation of perception: A minimal neuro-robotic model. Neurocomputing, 68, 85-104.

# Pupil dilation reflects the time course of perceptual emotion selection

***Manuel Oliva***[1], *Andrey Anikin*[1] *& Christian Balkenius*[1]

[1]*Lund University, Cognitive Sciences*

*manueloliva@gmail.com*

The processing of emotional signals usually triggers an increase in pupil size in comparison to emotionally neutral stimuli, and this effect was usually attributed to the emotional arousal elicited by the stimuli (Bradley, Miccoli, Escrig, & Lang, 2008; Partala & Surakka, 2003). Changes in pupil size have also been associated to decision making processes during visual perceptual rivalry (Einhäuser, Stout, Koch, & Carter, 2008), however, little is known about the role of pupil dilation during emotional selection. Therefore, in this study we investigated the relationship between pupil dilation and perceptual selection during the recognition of human nonverbal vocalizations. For such purpose, participants (N = 33) had to listen to human nonverbal vocalizations and indicate whether the stimuli had positive or negative emotional valence. The results show that the pupil dilation of the listener reveals the time course of emotional perceptual selection, where the peak pupillary response coincides with the time of emotion selection (Figure 1).

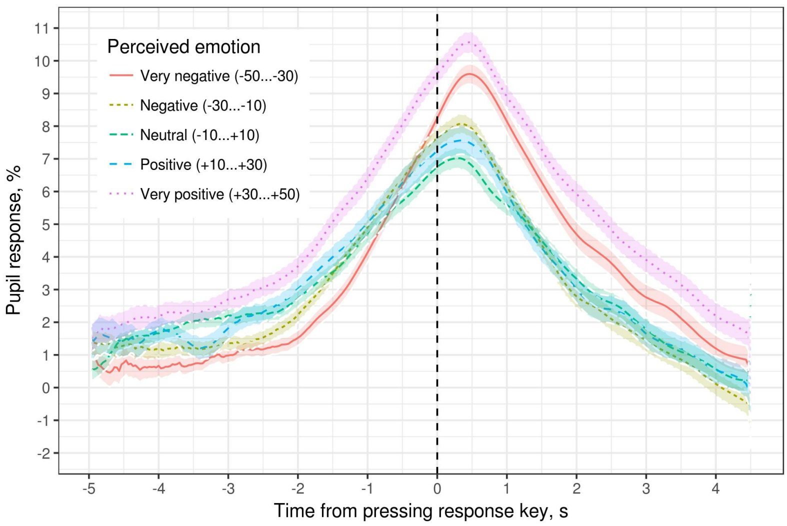



Figure 1. Pupil size time-locked to response times. The dashed line represents the moment participants responded about the emotional valence of the stimulus. Pupil size increased throughout the selection process until just after participants indicated the emotional valence of the stimuli. This pattern was consistent across a wide range of stimuli that varied in arousal intensity, ambiguity, and duration.

In addition, pupil dilation revealed properties associated to the perceptual decisions, where responses reported with lower confidence and/or higher perceived arousal (Figure 2) triggered larger pupil dilations.

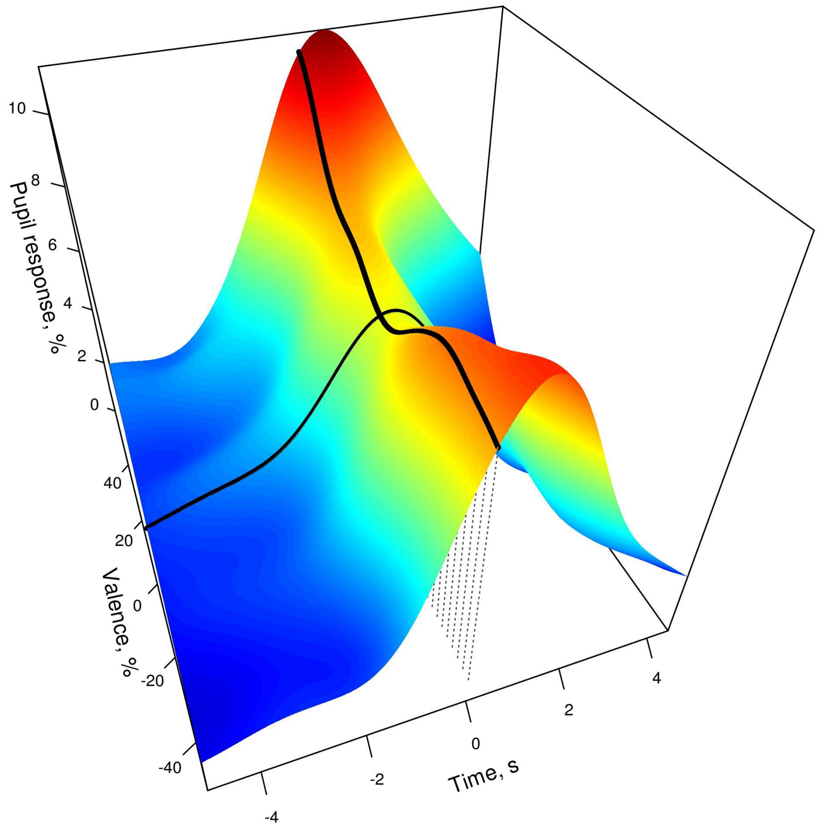



Figure 2. Time-locked pupil responses and stimulus emotional valence. Sound stimuli could be identified as negatively valenced (eg., sadness, cry) or positively valenced (eg., laughs). Pupil responses were moderated by the perceived emotional arousal of the stimulus. The dotted vertical lines and the upper black line indicate the moment a response was made (0 s). The results show that pupil responses not only reveal the time course of emotional selection but also characteristics associated to the decisions, such as the perceived emotional valence of the stimuli.

We interpret the results as suggesting that pupil responses to emotional stimuli do not only betray autonomic responses caused by arousing stimuli. Instead, we argue that during the recognition of emotions pupil dilation seems to be dominated by cognitive mechanisms in such a way that the emotional selection process can be traced through pupil dilation. These results contrasts with views of automatic emotion processing that assume little or no attentional mediation.

Because changes in pupil dilation (under isoluminance conditions) are believed to be caused almost exclusively by the release of norepinephrine from the locus coeruleus (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Joshi, Li, Kalwani, & Gold, 2016), the results suggest an important regulatory role of the LC-NE system during emotion recognition.

**References**

Bradley, M. B., Miccoli, L. M., Escrig, M. a., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and automatic activation. Psychophysiology, 45 (4), 602.

Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. Proceedings of the National Academy of Sciences of the United States of America, 105 (5), 1704–9.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010, may). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cognitive, affective & behavioral neuroscience, 10 (2), 252–69.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron, 89 (1), 221–234.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. International Journal of Human Computer Studies, 59 (1-2), 185–198.

# On the Essentially Dynamic Nature of Concepts

**Joel Parthemore**

*guest researcher, University of Skövde; postdoctoral researcher, Eindhoven University of Technology*

*j.e.parthemore@tue.nl*

It is commonly assumed that one of the essential characteristics of concepts – regardless of their referent – is their stability, tending toward stasis; indeed, it can be hard to see how concepts can be systematic and productive, in the way they are conventionally taken to be, unless they are so. Even the question has been raised whether concepts can change; on some prominent accounts emerging from the rationalist tradition, they cannot. The Unified Conceptual Space Theory (UCST) (Parthemore 2013, Parthemore 2011) – an extension of Peter Gärdenfors' (2004) Conceptual Spaces Theory (CST) – makes the controversial claim that concepts not only are subject to change – over an iterative lifecycle of "birth", "maturation", and "death" – but that, at an underlying level and from a certain critical perspective, they are in a state of continuous motion and must be to function as they do. Mere openness to change is not enough. Even the most seemingly fixed of concepts – mathematical concepts are the paradigmatic example – can be seen to evolve and continually be evolving. UCST suggests that concepts possess an intrinsic tension that might at first appear to present a contradiction: to be able to apply in more or less the same way across unboundedly many contexts (systematicity) and to be able to combine coherently with other concepts (productivity), they must be relatively stable; and yet, since each new application context is, in some nontrivial way, different from every previous context in ways that do not fit within neat conceptual boundaries, they must adapt to fit. In a physical world we have reason to view as ultimately one of fluidity, of processes and motion rather than stable objects, concepts – as the means by which we structure our understanding of that world and so the primary means by which we encounter it – should have a similar nature.

Theories of concepts are attempts, within cognitive science and philosophy of mind, are attempts to say what concepts are. They seek to lay out the ground rules for the organization of "higher-order" minds capable of stepping back from the present moment to consider it and its contents in light of moments past and moments yet to come. Among the contemporary theories being debated one finds Jerry Fodor's (1998) Informational Atomism and Jesse Prinz's (2004) Proxytypes Theory, along with the aforementioned CST and UCST. CST sees conceptual spaces as the analogue to physical spaces, with a different space for each conceptual domain, its geometry determined by the integral dimensions of that domain (in the case of the colour domain, those may be taken to be *hue*, *saturation*, and *brightness*; UCST attempts to show how all the different conceptual spaces described by CST come together in a single, unified "space of spaces" defined by integral dimensions common to all concepts. UCST comes with a toy software program for generating *mind maps*: visual descriptions of a given conceptual domain (Parthemore 2011, Ch. 8). The present paper is largely set within the framework of CST and UCST, though the claims it makes should resonate far beyond.

For sake of working definition (one that should be acceptable within all the theories mentioned above), let us take concepts to be either the building blocks of systematically and productively structured thought or the abilities by which a certain class of agents – call them *conceptual agents* – are able to engage cognitively with their environment in a systematically and productively structured fashion, one that affords them a flexibility of response to that environment akin to that afforded by consciousness. Although various researchers have offered their largely similar lists of the defining properties of concepts (see e.g. Chrisley & Parthemore 2007, Prinz 2004, Laurence & Margolis 1999, Fodor 1998) – which generally if not universally include systematicity and productivity – I'm not aware of anyone listing stability or as one of those properties. Nevertheless, most researchers would appear to assume that concepts are, at least most of the time, stable entities – and some (notably Fodor 1998) go so far as to argue that (at least most) concepts do not and cannot change. If the majority of researchers would allow that (at least most) concepts can change – within limits[1] – they would also generally

[1] That is because, unlike Fodor, they allow that beliefs are partly constitutive of concepts and not just (as everyone, including Fodor, allows) concepts are constitutive of beliefs. Beliefs change; therefore concepts change. Fodor takes the position he does because he is assuming a realist metaphysics.

hold that, most of the time, concepts do not. Both the resistance to and boundaries of change are important: after all, what use would a concept of "grizzly bear" be if it applied to a type of mammal one day and a kitchen utensil the next? On reflection, it may seem as if this tending-toward-ultra-stable nature is obligatory on concepts, and that nothing more needs to be said. Nevertheless, the following seems safe, on most accounts, to allow:

> **Theorem 1.** Concepts – to function as concepts – must be open to change.

This implies that any concepts that completely cease to be open to change are, metaphorically speaking, dead – at least if one allows that that which they are meant to reflect can, in principle, change. (Even concepts like parity conceivably could change if e.g. number theory were revised or expanded; after all, at no point can one comfortably announce that one has arrived at the "right" number theory.) The notion of conceptual "death" in turn implies the following, which I likewise take to be non-controversial:

> **Theorem 2.** Concepts may be seen to follow a life cycle of birth, maturation and (at least in certain cases) death.
>
> **Corollary.** The death of one concept is often the birth of another, or of several others.
>
> **Example.** When the concept of phlogiston was discarded in the late 18th Century, the concept of oxidation may be seen to have taken its place. Although the "birth" of the oxidation concept preceded the "death" of the phlogiston concept (except as a matter of historical interest), nevertheless the former may be seen as the natural heir of the latter.

This paper takes the far stronger position – strongly implied by UCST but, so far as I know, not defended elsewhere in print – that concepts are, by their nature, and from a critical and irreducible perspective, in a state of *continuous* (if often only incremental) change. The claim proceeds from what might be observed as a central (albeit paradoxical) tension in the nature of concepts. On the one hand, concepts – to function as concepts – must be both stable[2] and general ("context free") enough to apply across unboundedly many contexts; systematicity and compositionality would seem to require if not outright presuppose this. On the other, concepts always are applied within specific contexts – each of which is, seemingly unavoidably, different in some substantive way from any that have gone before. That implies that concepts *must* be sensitive to context (i.e., "context sensitive"), adapting to fit each new context as needed. The following seems safe, again, to allow on most accounts:

> **Theorem 3.** Concepts must be just stable enough, but not too stable!

That still is not, of course, sufficient to require continuous change. To get there requires two further ideas: first, that concepts are one thing when we self-reflect on them as concepts – in which case one can agree that they appear as stable representations (often called *mental representations*); and logically quite another when we simply get on with possessing and employing them non-reflectively as, seemingly, most of the time we must do – in which case they might seem to be something else, something non-representational – and action-based (for we are using them, not reflecting on them). Actions are, by nature, things in motion; and motion implies (if not requires) change. These two contrasting perspectives do a great deal, I think, to explain and resolve the debate over whether concepts are "really" representations or abilities. In truth, both perspectives are needed, and neither can be reduced to or otherwise reconciled with the other. If one allows for these two perspectives, then one will at least allow that the apparent stability of concepts, when we reflect on them, may not reflect their full nature.

The second requisite idea, which I take to be largely uncontroversial, at least until considered in its full implications, is that concepts are massively interconnected and – with care to avoid too close of a dictionary metaphor – inter-defining. Of course, not every agrees with such a view: Fodor, pointedly, views concepts (which he understands as atomic symbols) as strictly independent of one another, whereby (1998, p. 54) "it's plausible prima facie that 'a' might refer to a even if there are no other symbols". The idea is that knowing about weddings or funerals may require, inter alia, knowing about flowers, which may require understanding what a rose is, which may require recognizing red, which may also be connected to one's understanding of fire hydrants,

[2] By "stable" I simply mean "resistant to change".

which may be connected with one's concept of dogs (at least in the popular American imagination), which requires understanding what a pet is, which is why – even if one has never heard or used the expression before – one knows immediately what Fodor means when he uses his favourite conceptual example of a pet fish.[3] Fish is a type of seafood, which relates to sushi, which – if one has had a bad experience – may connect with one's concept of food poisoning. At no point does one reach the point where one may safely stop; the regress goes on as long as patience and cognitive resources allow. Think of concepts like a spider's web: pull on any one strand of the web, and the entire web vibrates; or consider each new experience like a pebble in a body of water, sending ripples to the farthest reaches. *If concepts and conceptual frameworks are massively connected in this way, then a change anywhere in the system will produce at least marginal movement throughout the system.* To avoid this, one must argue either that concepts in general or some concepts or clusters of concepts in particular[4] are more weakly connected: that is, that they are substantially context-free.

The central claim of this paper is that concepts – to function as concepts – are, at least when we are not looking at them, moving targets. The claim further is that this should be true on every level on which one may talk about concepts – individual, group, society, even species – albeit on different time scales (Parthemore, 2014). Even the most seemingly fixed of concepts – say, mathematical concepts of primeness or parity – may be seen to evolve and continuously be evolving, for the individual and society. Failure to be aware of change should not be taken as lack of change – not if the circumstantial evidence in favour of (continuous) change is sufficiently strong, as I will attempt to convince the audience that it is. I close with consequences – both theoretical and practical – for research in philosophy of mind, cognitive science, and related fields.

## References

Chrisley, R. & Parthemore, J. (2007). Synthetic phenomenology: exploiting embodiment to specify the nonconceptual content of experience, *Journal of Consciousness Studies*, **14**(7): 44-58.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Gärdenfors, P. (2004 [2000]). *Conceptual Spaces: The Geometry of Thought*. Bradford Books.

Laurence, S. & Margolis, E. (1999). Concepts and cognitive science. In Margolis, E. & Laurence, S. (eds.), *Concepts: Core Readings* (3-81). Cambridge, Massachusetts, USA: MIT Press.

Parthemore, J. (2014). Conceptual change and development on multiple time scales: From incremental evolution to origins. *Sign Systems Studies*, **42**(2-3): 193-217.

Parthemore, J. (2013). The unified conceptual space theory: An enactive theory of concepts. *Adaptive Behavior* **21**(3): 168-177.

Parthemore, J. (2011). *Concepts Enacted: Confronting the Obstacles and Paradoxes Inherent in Pursuing a Scientific Understanding of the Building Blocks of Human Thought* (DPhil thesis). Falmer, Brighton, UK: University of Sussex. http://sro.sussex.ac.uk/6954/1/Parthemore%2C_Joel.pdf.

Prinz, J. (2004 [2002]). *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge, Massachusetts, USA: MIT Press.

[3] Fodor takes the pet fish as proof that concepts are not prototypes, for a pet fish is neither a prototypical pet nor a prototypical fish, never mind the intersection of the two. I prefer to take a different lesson: that concepts compose in a different way and along different paths from language, and that one does not arrive at the concept of pet fish simply by combining the lexical concept of pet with the lexical concept of fish. In short: an overly linguistic view on concepts seriously distorts one's view of them.

[4] *Natural kinds* concepts, if they exist, would be the obvious example. Needless to say, I have argued that they do not (see e.g. Parthemore, 2011: 58-59).

# The social side of imitation in human evolution and development: Shared intentionality and imitation games in chimpanzees and 6-month old infants

**Gabriela-Alina Sauciuc**[1], Tomas Persson[1], Elainie Alenkaer Madsen[1]

[1]*Lund University, Cognitive Science*

*Gabriela-Alina.Sauciuc@lucs.lu.se*

Imitation is generally acknowledged as a key mechanism of social learning, foundational to the emergence of human culture. By enabling quick and high-fidelity copying of others' actions, imitation mediates the cross-generational transfer of knowledge and skills (e.g. Nielsen, 2009). Besides this 'learning' (or 'cognitive') function, imitation accomplishes also important social-communicative functions, by facilitating social interaction and promoting prosociality (e.g. Duffy & Chartrand 2015; Eckerman, Davis, & Didow, 1989; Užgiris, Benson et al., 1989). The social function of imitation is understudied in the field of comparative psychology, or even claimed to be absent in nonhuman primates. This claim, however, is grounded on how nonhuman apes (henceforth 'apes') perform in imitation learning experiments compared to human children. More specifically, chimpanzees exhibit lower levels of joint attention and gaze at the experimenter's face (Carpenter & Tomasello, 1995). Moreover, children - but not chimpanzees - exhibit 'over-imitation', i.e. they show a propensity for faithfully copying demonstrated actions, even when these actions are irrelevant for achieving a demonstrated outcome. Such differences, it has been argued, derive from the fact that, in imitation contexts, children are motivated by a need to belong, to engage socially and to promote shared experiences (Carpenter & Call 2009; Nielsen 2009). In turn, these differences in social motivation are taken to account for the profound differences that exist between human and nonhuman primate cultures (Over & Carpenter 2012).

Based on evidence from social, developmental and comparative psychology, we have recently proposed a broader definition of the social-communicative function of imitation (Persson, Sauciuc, & Madsen, 2017), that encompasses reactive and non-intentional phenomena (e.g. nonconscious mimicry, imitation-induced prosociality), as well as proactive and arguably intentional phenomena, such as social conformism or the communicative imitation documented in preverbal toddlers (e.g. Eckerman, Davis, & Didow, 1989; Eckerman & Stein, 1990). All these phenomena have been documented in nonhuman primates: nonconscious mimicry in the form of postural congruence (Jazrawi, 2000), facial mimicry (Scopa & Palagi, 2016), interactional synchrony (Yu and Tomonaga, 2016) and contagious yawning (Madsen, Persson, et al., 2013), imitation-induced prosociality expressed by increased levels of attention, proximity and object exchange after exposure to being imitated (e.g. Paukner, Suomi et al., 2009), social conformism in the form of a preference for a group-adopted procedure even when it went against a prefered or more efficient one (Hopper, Schapiro, et al., 2011), and communicative imitation in the form of familiar-action imitation used to engage or maintain interaction (Persson, Sauciuc, & Madsen, 2017).

In this presentation, we address the presence of shared intentionality in imitative contexts with evidence from four experimental studies that our team has conducted with 6-month old infants (Sauciuc, Madsen, et al., *in prep*), as well as with enculturated (Sauciuc, Persson, & Madsen, *in prep*) and non-enculturated (Madsen, Sauciuc, & Persson, *in prep* a, b) chimpanzees of various ages (infants, juveniles, adults). Common to all these studies is that the participants have been exposed to an imitation condition in which the experimenter imitated all their actions, as well as to a number of control conditions that varied in agreement with the specific aims of each study. In Sauciuc, Madsen et al. (*in prep*), to establish if 6-month old infants discriminate being imitated from contingent responding, and to examine likely mechanisms that mediate this process, infants interacted with an experimenter who (i) imitated all infant's action ipsilaterally; (ii) imitated all infant's actions contralaterally; (iii) imitated with a still-face, i.e. imitated bodily but not facial actions; or (iv) responded with the infant's actions contingently but with different actions. In Madsen, Sauciuc, & Persson (*in prep a*), to track the ontogenetic course of imitation recognition in chimpanzees, we replicated Haun and Call's (2008) study on imitation recognition in adult apes and exposed infant and juvenile chimpanzees to four types of interaction in which the experimenter either (i) imitated all chimpanzee's actions; (ii) responded to the chimpanzee's actions with temporally contingent but different actions; (iii) produced actions that were not related to the chimpanzee's actions; (iv) sat still. In Sauciuc, Persson, & Madsen (*in prep*) four additional control conditions were administered in order to ascertain that behavioural indicators of shared intentionality (e.g. imitation games, laughter) could not be attributed to alternative factors known to increase playfulness in chimpanzees, including non-play species-specific behaviours, species-specific play forms (chase) or facial expressions that accompany play. Finally, in Madsen, Sauciuc, & Persson (*in prep* b), chimpanzees were exposed to bouts of (i) imitation, (ii) non-imitative play and (iii) no action in order to investigate the effects of imitation and non-imitative play on subsequent intentional imitation of non-instrumental actions and nonconscious mimicry (such as contagious yawning).

To examine the presence of shared intentionality in the studied populations, we focused on the presence of testing behaviours and imitation games, as well as on the presence of smiling and laughter during such responses. ‘Testing behaviours’ are generally acknowledged as an indication of explicit imitation recognition, i.e. that the imitated individual is aware of the imitator’s intention to copy his/her behaviours (Whiten & Suddendorf, 2001). They may take the form of ‘behavioural repetitions’ (the imitated individual repeatedly reproduces a previously imitated action), ‘testing sequences’ (the imitated individual produces rapidly a series of different actions) or ‘testing poses’ (the imitated individual suddenly freeze in a posture). Such ‘testing behaviours’ are generally regarded as a mean by which the imitated individual actively tests the contingent correspondence between own actions and those of the imitator. The presence of testing behaviours is thus considered to be an indication that the imitated individual is aware of this action correspondence, as well as of the impact that his/her actions has on the behaviour of the imitator (e.g. Bates & Byrne, 2010;). Testing behaviours have been documented in human infants as early as 9-months of age (Agnetta & Rochat, 2004) and in apes (e.g. Haun & Call, 2008), but not in monkeys. Unlike human infants, however, apes do not seem to exhibit shared intentionality in such imitative contexts, i.e. they do not show signs of enjoyment and playfulness (laughter, imitation games) when being systematically imitated (Nielsen, 2009).

Contrary to this view, our studies bring evidence that both enculturated and non-enculturated chimpanzees show enjoyment and playfulness when being imitated. Indeed, laughter and imitation games were present in both young and adult chimpanzees, in both enculturated and non-enculturated populations. We have also found that human infants produced testing behaviours as early as 6 months of age, and that they engaged in imitation games accompanied by smiling regardless of whether the experimenter imitated them ipsilaterally, contralaterally or with a still-face. In all the studied populations, testing behaviours were primarily expressed by behavioural repetitions, but testing sequences accompanied by smiling / laughter and careful monitoring of the experimenter’s actions were also present.

We conclude that the social side of imitation in its proactive form emerges early in human development, and has ancient evolutionary roots, i.e. it was likely present in the common ancestor of humans and chimpanzees. Since both enculturated and non-enculturated chimpanzees evidenced enjoyment, playfulness and strong social engagement when being imitated, it is unlikely that lack of shared intentionality and social motivation accounts for chimpanzees’ poorer performance in imitation learning tasks when compared to human children.

## References

Agnetta, B, & Rochat, P. (2004). Imitative Games by 9-, 14-, and 18-Month-Old Infants. *Infancy, 6,* 1-36.

Bates, L.A., & Byrne, R.W. (2010). Imitation: what animal imitation tells us about animal cognition. *WIREs Cognitive Science 1*, 685–695

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. *Revue de Primatologie, 1:6*

Carpenter, M., & Tomasello, M. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development, 4,* 217–237

Eckerman, C.O., & Stein, M.R. (1990). How imitation begets imitation and toddlers’ generation of games. *Developmental Psychology, 26*, 370–378

Eckerman, C.O., Davis, C.C., & Didow, S.M. (1989). Toddlers’ emerging ways of achieving social coordination with a peer. *Child Development*, *60*, 440-453.

Haun, D.B.M., & Call, J. (2008). Imitation recognition in great apes. *Current Biology, 18,* R288–R290.

Hopper, L.M., Schapiro, S.J., Lambeth, S.P., & Brosnan, S.F. (2011). Chimpanzees' socially maintained food preferences indicate both conservatism and conformity. *Animal Behavior, 81*, 1195–1202.

Jazrawi, S. (2000). *Postural congruence in a captive group of chimpanzees* (*Pan troglodytes*) (MA dissertation). Department of Anthropology, University of Calgary, Calgary.

Madsen, E.A., Persson, T., Sayehli, S., Lenninger, S., & Sonesson, G. (2013). Chimpanzees show a developmental increase in susceptibility to contagious yawning: a test of the effect of ontogeny and emotional closeness on yawn contagion. *PLoS ONE*, *8*, e76266.

Madsen, E.A., Sauciuc G.A., & Persson, T. (*in prep*. a). The development of imitation recognition in young chimpanzees. *PLoS ONE*.

Madsen, E.A., Sauciuc G.A., & Persson, T. (*in prep*. b). The chameleon effect in young chimpanzees. *Scientific Reports*.

Nielsen, M. (2009). The imitative behaviour of children and chimpanzees: a window on the transmission of cultural traditions. *Revue de Primatologie 1*, 5.

Over, H, & Carpenter, M. (2012). The social side of imitation. *Child Development Perspectives*, *7*, 6-11.

Paukner, A., Suomi, S.J., Visalberghi, E., & Ferrari, P.F. (2009). Capuchin monkeys display affiliation towards humans who imitate them. *Science, 325*, 880–883.

Persson, T, Sauciuc, G.A., & Madsen, E.A. (2017). Spontaneous cross-species imitation in interactions between chimpanzees and zoo visitors. *Primates*, First online: 16 august 2017.

Sauciuc, G.A., Madsen, E.A., Persson, T., Zlakowska, J., & Lenninger, S. (*in prep*). Being imitated: explicit recognition and prosocial effects in 6-month old infants. *Scientific Reports*.

Sauciuc, G.A., Persson, T, & Madsen, E.A. (*in prep*). Imitation games and laughter in human-chimpanzee interaction and their affiliative effects. *International Journal of Primatology*.

Scopa, C. & Palagi, E. (2016). Mimic Me While Playing! Social Tolerance and Rapid Facial Mimicry in Macaques (*Macaca tonkeana* and *Macaca fuscata*). *Journal of Comparative Psychology*, *130*, 153-161.

Užgiris, I.Č., Benson, J., Kruper, J., & Vasek, M. (1989). Contextual influences on imitative interactions between mothers and infants. In: J.L. Lockman & N.L. Hazen (eds) *Action in social context: perspectives on early development* (pp. 103-127). New York, Springer.

Whiten, A, Suddendorf, T. (2001). Meta-representation and secondary representation. *Trends in Cognitive Science, 5*, 378.

Yu, L., & Tomonaga, M. (2015). Interactional synchrony in chimpanzees: examination through a finger tapping experiment. *Scientific Reports*, *5*, 10218.

# Human blindness to noise in neural computation

**Christopher Summerfield**

*Department of Experimental Psychology, University of Oxford*

*christopher.summerfield@psy.ox.ac.uk*

Humans typically make near-optimal sensorimotor judgments but show systematic biases when making cognitive decisions. We tested the hypothesis that this gap in optimality arises because of differential metacognitive sensitivity to two independent sources of uncertainty: encoding noise (limiting stimulus perceptibility) and integration noise (arising when multiple pieces of discordant information are combined).

In five psychophysical experiments, human observers judged the mean orientation of an array of gratings. We independently manipulated the contrast (encoding noise) and orientation variability (integration noise) in the array. Participants adapted optimally to increased levels of encoding noise but not integration noise: under higher orientation variability, they failed to use probabilistic prior information to compensate for increased errors, reported excessive confidence, and refrained from selecting an option to "opt out" of more difficult trials. These findings are captured by a Bayesian model which is blind to integration noise, providing a computationally grounded explanation of human suboptimal inferences.

# Smartphones, films, and cognition

**Kata Szita**

*Department of Cultural Sciences, University of Gothenburg*

*kata.szita@gu.se*

The smartphone gains an increasing amount of attention not only as a multimedia tool, but also in a number of academic disciplines. What is, however, yet to be exploited is its role in reforming the consumption of a wide range of audiovisual content by its portable and interactive nature. In this paper, I use an interdisciplinary approach to turn the spotlight on how smartphones can affect viewing strategies and the overall film and video experience, and point toward potential new content producing strategies.

Smartphone film experience differs from the cinematic or home video experience in three key factors: (1) the viewing space, in which stimuli from the virtual, filmic world collide with those of the physical one, and induce attention oscillation; (2) the size of the screen, which may cause changes in gaze patterns by masking some visual details which would otherwise be highlighted on a large screen; and (3) the relationship between the spectator's body and the interactive interface, and the possibility to adjust the distance between the screen and one's eyes to adjust the amount of external visual distraction in the film experience.

Active environments, in which an abundance of sonic and visual information divide the spectator's attention, function according to specific physical and social properties, which means that they fail to cohere with the viewing activity. However, immersion into the movie reserves cognitive processes and somewhat counterbalances the environmental stimuli. For instance, no matter if the source is part of the movie stimulus or not, when a sonic referent is related to what is seen on the screen, the spectator can interweave the sensation with the film narration, thus the stimulus becomes an acoustic cue. From another angle, visual referents are understood in relation to sonic information, which results in the shift of the spectator's gaze towards the (expected) source of the sound. (Smith & Henderson, 2008) Sonic cues, such as voice changes, words, and music or sound effect onsets guide attention and extend each other. Michel Chion (1994) explains that the system of sound and image directs attention, affects emotional responses, generates anticipations, and influences narrative meaning. Hence, a sonic distraction—or, a sound effect that the spectator is exposed to while watching—can be classified by its ecological connection to the film soundtrack and visual language and its neutrality, that is, whether the sound or visual effect evokes an urge for immediate action (e.g. telephone ringing, sight of a written text) or is neutral to the observer (e.g. abstract images, music).

Being in a physical proximity to the screen adds an additional factor of distraction, while it also suggests a broader influence on the image and sound presentation, and perhaps a greater sense of immersion. The ability to alter the properties of the screening is twofold: on the one hand, one can apply changes in the screening pace, luminance, volume, and the like, on the other, because of the size of the device, there is a possibility to vary the perceptual relationship with the stimulus by moving the screen closer or further from one's eyes. These options are presumed to be in a tight interplay with external distractors, and provide the possibility to adjust the viewing conditions to that of the surrounding space.

Visch, Tan, and Molenaar (2010) claim that higher immersive conditions, with a screen being closer to the subject intensifies both the appreciation of the film (artifact emotions) and the emotional engagement to the characters (fictional world emotions), that is due to the VR stimulus being perceived as an "alternative to reality" (p. 1444). In addition, emotional responses correlate with the level of immersion and emotional engagement is the result of empathy toward the characters' personalities and acts. (M. D. Slater & Rouner, 2002) Thus, immersion depends on the exclusion of the environmental stimuli which, consequently, suggests the decrease of the emotional involvement in the case of environmental distraction; the number of sensory modularities and the proximity thereof; and the feeling of being present in the diegetic space. (M. Slater & Wilbur, 1997; Wirth et al., 2007) The medium properties also have an impact on the feeling of presence, namely the number of the senses involved in perception (M. Slater & Wilbur, 1997), and the image size and viewing angle (among others, Troscianko, Meese, & Hinde, 2012).

To analyze viewer behavior and to identify the direction toward the optimization of mobile-distributed movies and videos and to compare smartphone film experience to more traditional, larger screen experience, 38 volunteers were asked to watch two, approximately 10 minutes' clips from an English-speaking commercial movie, *The Walk* (Rapke, Starkey, Zemeckis, & Zemeckis, 2015) on different screens and in different viewing conditions, while their physiological reactions were measured and subjective ratings were obtained. Unlike for instance natural scenes (Smith & Mital, 2013), Hollywood-style movies (see Bordwell, Thompson, & Staiger, 1985) are proved to induce a high level of exogenous (externally-driven) impact on the viewing process, that makes them suitable for measuring viewer behavior and information acquisition, and the role of media. In order to maintain exogenous control, the sequences were chosen to bear elements that guide the attention in an analogous way for all viewers and present semantically meaningful objects, such as faces or landmarks. The two sequences contain these elements to the same extent and are set mostly at the same locations, however depict different episodes of the main protagonist's act, which make them suitable for measuring the same kinds of reactions, while eschewing biases due to sequential effect.

Participants were randomly assigned to a group with a predetermined order of film sequences and viewing conditions. While cinema, television, and other screens are not only bigger, but also most typically used in spaces set to enhance viewing experience, smartphones users often consume moving-image content on their devices as a secondary activity and within spaces which are not specifically designed for movie watching. To consider these factors, two types of screens were used: a five-and-a-half-inches' smartphone and a projector, projecting on a fifty inches' canvas. The conditions were labeled as *mobile*, *mobile V*, *projector*, and *projector V*, where the "V", variant, conditions contained additional stimulation with light sonic and visual effects of varying ecological validity, source location, duration, and neutrality at determined points of time in correlation with the movie sequence. These effects were designed to recreate stimuli from the physical space, and were replayed unannounced. Each of the 38 participants was tested in two of the conditions (*mobile* and *projector V* or *mobile V* and *projector*) in a randomized, but counterbalanced order, making it a total of 76 trials.

The *projector* condition was used to provide the baseline data of viewer behavior and to identify the areas of interest (AOIs) for the later stages of the analysis. In this setup, participants were seated in a shielded, dimmed experiment room in a fixed distance of 180 centimeters from a 47.3" canvas, where the eye level was approximately in the middle of the screening area. For sound presentation headphones were used. The *mobile* condition is designed to represent the typical smartphone viewing settings, therefore varying viewing angles were used giving subjects the liberty to adjust the screen distance to their personal preferences, visual abilities, and to the environment, which, although, still not exceeded the visual angle of the larger screen condition, as the screen size and the coverage of the visual field by the film image affects the gaze and the perception of the visual narrative cues. (Hochberg & Brooks, 1978a, 1978b)

During the free viewing task, two sets of data were monitored in a noninvasive way: physiological reactions (eye movements and electrodermal activity, EDA) and correlation thereof with the external stimuli, that is, sonic and visual distractors and the participant's potential interaction with the device (recorded through screen capture in *mobile* and *mobile V* conditions). For measuring eye movements both on and off screen, mobile eye tracking glasses were used. Eye tracking measurements would not provide the sufficient information about viewing and information seeking strategies on small screens, and therefore have to be complemented by other methods. Due to the center bias (the gaze shifts to the center of the screen if there is nothing attractive elsewhere, as well after cuts), in case of rapid changes on the screen spectators have a tendency to fixate at the center, and their gaze rarely shift to the edges of the screen. (Tseng, Carmi, Cameron, Munoz, & Itti, 2009; Mital, Smith, Hill, & Henderson, 2011) In addition, smaller screens tend to decrease the amplitude of eye movements, and spectators are more likely to move their eyes close to the center than on a bigger screen as the sharp area of vision covers a bigger proportion of the screen. (see Smith, 2014) Complimentary measurements would not limit the findings to determine *where* spectators look, but give a broader overview to what extent they acquire the necessary narrative information and what elements of the narrative affect their understanding. As a sensitive marker for emotional engagement, electrodermal activity provides valuable information on emotional arousal, reactiveness, attention, and immersion by measuring autonomic changes in sweat glands at given times of interest. (Boucsein, 2012)

After watching the assigned movie clip, participants were asked to fill up a questionnaire that targeted to measure the subjective feeling of emotional engagement and presence on the one hand, and narrative understanding on the other. Besides the above measurements, the questionnaire was also used to focus subjects' attention to the viewing task, engage them in the activity, and motivate them to explore the film as much as possible. The first section used a 10-point Likert-type scale with values ranging from true to not at all true, where participants were asked to rate their film experience by the following factors: presence in the diegetic space, empathy towards the characters, level of feeling scared, moved, and nauseated. The second section of the questionnaire contained questions about the narrative elements of each sequence with the possible answers of "yes", "no", and "I don't know". The "I don't know" option was necessary to avoid obliging the subject to aim for the right answer even when he or she did not actually pay attention to that particular segment of the film clip. The answers were classified and analyzed as "correct answer", "incorrect answer", and "I don't know".

The analysis aims to define the potential viewing and content-making strategies for smartphone viewership and concern whether viewing experience on smartphones induces different level of attention, presence-feeling, and influences the overall narrative understanding than other, widely-used screens of commercial consumption, such as cinema. The study hypothesizes that in the case of smartphones and environmental distractions, subjects' eye movements are less synchronized, and more likely follow individual patterns; the gaze is more focused towards the center of the screen and saccade amplitude decreases; and that films watched on a smartphone in an active environment induces more moderate changes in skin conductance at scenes with high emotional content than uninterrupted watching on the bigger screen.

## References

Bordwell, D., Thompson, K., & Staiger, J. (1985). *The Classical Hollywood Cinema: Film Style & Mode of Production to 1960*. New York: Columbia University Press.

Boucsein, W. (2012). *Electrodermal Activity*. Boston: Springer.

Chion, M. (1994). *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Hochberg, J., & Brooks, V. (1978a). Film Cutting and Visual Momentum. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye Movements and the Higher Psychological Functions* (pp. 293-313). Hillsdale: Erlbaum.

Hochberg, J., & Brooks, V. (1978b). The Perception of Motion Pictures. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of Perception* (Vol. 10, pp. 259-304). New York: Academic Press.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing Is Predicted by Motion. *Cognitive Computation*, 3(1), 5-24.

Rapke, J. P., Starkey, S. P., Zemeckis, R. (Producers), & Zemeckis, R. (Director). (2015). The Walk [Motion picture]. United States: Sony Pictures Entertainment.

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (Five): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 6(6).

Slater, M. D., & Rouner, D. (2002). Entertainment Education and Elaboration Likelihood: Understanding the Processing of Narrative Persuasion. *Communication Theory*, 12(2), 173–191.

Smith, T. J., & Henderson, J. (2008). Attentional Synchrony in Static and Dynamic Scenes. *Journal of Vision*, 8(6), 773.

Smith, T. J., & Mital, P. K. (2013). Attentional Synchrony and the Influence of Viewing Task on Gaze Behavior in Static and Dynamic Scenes. *Journal of Vision*, 13(8), 1-24.

Troscianko, T., Meese, T. S., & Hinde, S. (2012). Perception While Watching Movies: Effects of Physical Screen Size and Scene Type. *I-Perception*, 3(7), 414-425.

Tseng, P.-H., Carmi, R., Cameron, I., Munoz, D., & Itti, L. (2009). Quantifying Center Bias of Observers in Free Viewing of Dynamic Natural Scenes. *Journal of Vision*, 9(7), 1-16.

Visch, V. T., Tan, E. S., & Molenaar, D. (2010). The Emotional and Cognitive Effect of Immersion in Film Viewing. *Cognition and Emotion*, 24(8), 1439-1445.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., . . . Jäncke, P. (2007). A Process Model of the Formation of Spatial Presence Experiences. *Media Psychology*, 9(3), 493-525.

# Neural network and human cognition: A case study of grammatical gender in Swedish

**Ali Basirat**, **Marc Tang**

*Department of Linguistics and Philology, Uppsala University*

*ali.basirat@lingfil.uu.se, marc.tang@lingfil.uu.se*

We target grammatical gender in Swedish and compare the errors in grammatical gender recognition produced by the artificial neural network model versus errors of L1 and L2 Swedish speakers attested in the existing literature. This research thus follows the call of SweCog 2017 “creating a strong interdisciplinary cluster of cognitive science oriented research within Sweden”. First, we combine cognitive science with computational linguistics and language acquisition. Second, our analysis is based on the Swedish language. Third, as mentioned by the call of SweCog: “describing on-going work are especially welcome”. Our study is also an on-going project which will benefit from the feedback at the conference.

“Language offers a window into cognitive function, providing insights into the nature, structure and organization of thoughts and ideas.” (Evans & Green, 2009, p. 5). Thus, linguists are interested in systems of nominal classification, i.e. how languages classify nouns of the lexicon, due to their diverse lexical and pragmatic functions as well as cognitive and cultural correlates. As stated by Lakoff & Johnson (2003, pp. 162–163), “In order to understand the world and function in it, we have to categorize, in ways that make sense to us, the things and experiences that we encounter”. This need is fulfilled in language via various means of categorization, one of the most prominent being grammatical gender (Aikhenvald, 2000; Corbett, 1991). In gender system (also known as noun class system) of languages, all nouns of the lexicon are assigned to a specific number of classes. Saying that a language has two genders implies that there are two classes of nouns which can be distinguished syntactically by the agreement they take with other syntactic units in their contexts (Bohnacker, 2004, p. 198; Senft, 2000). As demonstrated in Figure 1, nouns in Swedish are divided into neuter and uter (common). The two categories are thus reflected on the determiners and adjectives respectively.

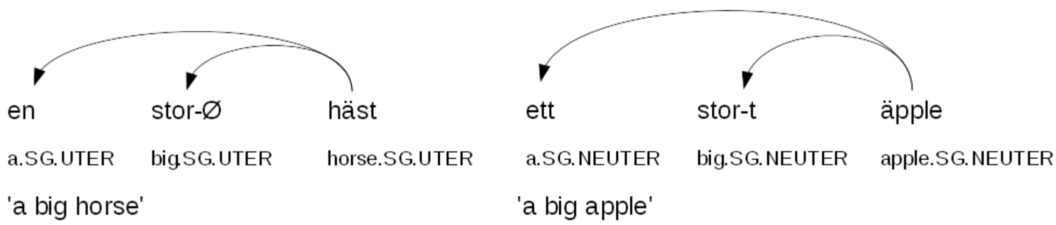



*Figure 1. The grammatical gender agreement in Swedish*

We selected Swedish due to the following reasons. Gender assignment across languages is underlined by cognitive and sociocultural principles (Aikhenvald, 2012; Corbett, 1991, p. 57; Kemmerer, 2017). By way of illustration, long-shaped objects tend to be affiliated to masculine grammatical gender and round-shaped objects are generally associated with feminine grammatical gender. However, contradictory observations are made in Swedish. First, grammatical gender assignment is generally viewed as arbitrary (Andersson, 1992; Teleman, Hellberg, & Andersson, 1999), but some semantic regularities are still attested in Swedish. Animate nouns, especially “all non-pejorative, classificatory nouns denoting adult human beings, a qualified majority of all other human nouns and a majority of all other animate nouns” strongly tend to be affiliated to the uter gender (Dahl, 2000, pp. 586–587), while other inanimate nouns are affiliated to the neuter gender. This distribution is also reflected in terms of count/mass division. Nouns referring to concrete and countable entities are more likely to be uter while abstract or collective meanings are associated to neuter, e.g. “possible people containers” such as nouns denoting location and organization are perceived as collective units. Thus, they tend to be neuter (Fraurud, 2000, pp. 191–203). Second, the L1 and L2 acquisition of Swedish grammatical gender are controversial and differ significantly from other languages. Monolingual children acquire the Swedish gender system with nearly no errors (Andersson, 1992; Bohnacker, 1997; Plunkett & Strömqvist, 1990), which is considered rare in comparison to other gender languages, for which “children's acquisitional paths have been reported not to be quite so error-free” (Bohnacker, 2004, pp. 214–217). Moreover, gender assignment on Swedish nouns via their phonological form or semantics is generally considered as unpredictable (Andersson, 1992; Teleman et al., 1999), which makes this observation even more unexpected. Nevertheless, while L1 acquisition display a lack of errors, L2 (child) learners do encounter difficulties, suggesting that different strategies are employed (Bohnacker, 2004, p. 218).

We chose a computational approach since the use of computer simulations of neurons and neural networks are one of the most important tools in computational cognitive neuroscience (Parks, Levine, & Long, 1998). We simulate the learning process of the brain with neural networks, which "have become a subject of intense interest to scientists spanning a broad range of disciplines including psychology, physics, mathematics, computer science, biology and neurobiology" (Gopal, 1996, p. 69). In this study, we apply the feed-forward neural network (Haykin, 1998) to classify between the Swedish grammatical genders of nouns. The architecture of the network is as follows: We first feed a 50-dimensional vector representation of words, called word vector, in the input layer of the network. The word vector is then processed in the only hidden layer of the network consisting of 100 neurons. In the output layer, the network generates two weight values corresponding to the grammatical gender of the input word. The input word vectors are extracted by the RSV (Real-valued Syntactic Word Vectors) word embedding model (Basirat & Nivre, 2017). Our model relies on two main sources of data, which both originate from the Swedish Language Bank (Språkbanken) located at the University of Gothenburg: a corpus of Swedish raw sentences and a list of nouns affiliated to grammatical genders. The corpus originates from Swedish Wikipedia available at Wikipedia Monolingual Corpora, Swedish web news corpora (2001-2013) and Swedish Wikipedia corpus collected by Språkbanken. The list of nouns and their affiliated grammatical gender was extracted from the SALDO (Swedish Associative Thesaurus version 2) dictionary. The dictionary contains in total 85,970 nouns of uter and neuter genders. These nouns are then filtered by criteria of frequency in the corpora. We eliminate with a frequency lower than 100. As a result, we obtain 21,670 words which are partitioned in a standard way into three parts with no overlap, so that the results can be generalized to the entire lexicon of the language. We use 80% of words (17,338) to train the neural network, 10% of words (2,166) as the development set, and the remaining 10% as test set. The dictionary data is divided into three sets so that the performance of the neural network may be enhanced and re-measured between the development test and the test. All words are randomly selected in their base format with no morphological inflection and all three training sets contain an equivalent distribution of uter and neuter nouns.

Our preliminary results on the development set show that the neural network can correctly detect the grammatical gender of the nouns with the overall accuracy of 93.46%. In other words, when being presented a new word, the neural network can interpret correctly the grammatical gender of the noun in 93.46% of the time. It is worth mentioning that the only source of information used by the neural network is the contextual information encoded into the word vectors with no information about the possible morphological inflections of words. We have also observed that the precision for the uter nouns (97.1%) is significantly higher than the precision for the neuter nouns (84.6%), i.e. most of the difficulties encountered by the neural network are related to the fact that it could not identify the gender of neuter nouns. This correlates with the findings of human acquisition where learners tend to overgeneralize the uter gender (Bohnacker, 2004, p. 218) due to the lack of balance in terms of quantity between uter and neuter nouns in Swedish (71.06% vs 28.94% according to SALDO). As shown in Figure 2, the errors generated by neural network are mostly cases where the noun is located in the space of the opposite gender, e.g. a neuter noun misinterpreted as uter nouns (red) is more likely to be found in the cloud of neuter nouns (green).

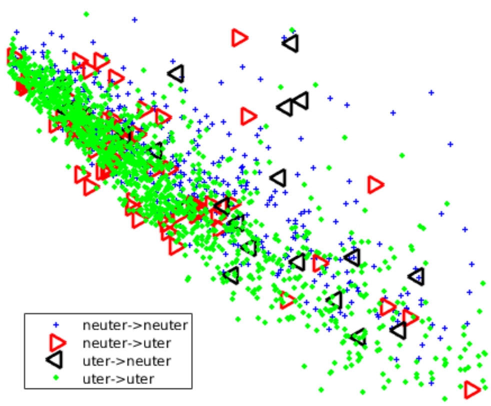



*Figure 2: PCA representation of the word vectors classified by the neural network with respect to their grammatical genders. "X→Y" means the noun belonging to category X is classified as Y.*

We will provide in our presentation a preliminary analysis, which further demonstrate that the distribution of errors made by both human and the artificial neural network is similar. We observe that both humans and the artificial neural network have difficulties interpreting the gender of nouns in cases of syntactic ambiguity and polysemy, e.g. participles such as *flyttande* 'moving' are attested as nouns in SALDO but they may also appear as participles in modifier structures. The same case occurs between verbs and nouns, e.g. *delta* can link to a verb 'participate' or a noun 'delta'. Moreover, nouns such as *kaffe* 'coffee' may be interpreted as uter or neuter depending on whether it is referring to the entity as count or mass. Thus, when two entries are competing for the same word vector, the vector is biased toward the most frequent reference which will negatively affect the decision making of the neural network. Finally, the actual limitations and future prospects of this study include an expansion of first-hand data from Swedish L1 and L2 speakers, i.e. we intend to gather data specific to our study via empirical experiments with human speakers. Furthermore, a cross-language analysis is also required to verify the stability of the neural network in various linguistic environments.

## References


Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Aikhenvald, A. Y. (2012). Round women and long men: Shape, size, and the meanings of gender in New Guinea and beyond. *Anthropological Linguistics*, *54*(1), 33–86.

Andersson, A.-B. (1992). *Second language learners' acquisition of grammatical gender in Swedish* (PhD dissertation). University of Gothenburg, Gothenburg.

Basirat, A., & Nivre, J. (2017). Real-valued Syntactic Word Vectors (RSV) for Greedy Neural Dependency Parsing. In *Proceedings of the 21st Nordic Conference on Computational Linguistics, NoDaLiDa* (pp. 21–28). Gothenburg: Linköping University Electronic Press.

Bohnacker, U. (1997). Determiner phrases and the debate on functional categories in early child language. *Language Acquisition*, *6*, 49–90.

Bohnacker, U. (2004). Nominal phrases. In G. Josefsson, C. Platzack, & G. Håkansson (Eds.), *The acquisition of Swedish grammar* (pp. 195–260). Amsterdam: John Benjamins.

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

Dahl, O. (2000). Elementary gender distinctions. In B. Unterbeck & M. Rissanen (Eds.), *Gender in grammar and cognition* (pp. 577–593). Berlin: Mouton de Gruyter.

Evans, V., & Green, M. (2009). *Cognitive linguistics: an introduction* (Repr). Edinburgh: Edinburgh University Press.

Fraurud, K. (2000). Proper names and gender in Swedish. In B. Unterbeck, M. Rissanen, T. Nevalainen, & M. Saari (Eds.), *Gender in Grammar and Cognition* (pp. 167–220). Berlin: Mouton de Gruyter.

Gopal, S. (1996). Neural network models of cognitive maps. In J. Portugali (Ed.), *The Construction of Cognitive Maps* (pp. 69–85). Dordrecht: Kluwer Academic Publishers.

Haykin, S. (1998). *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall PTR.

Kemmerer, D. (2017). Categories of object concepts across languages and brains: the relevance of nominal classification systems to cognitive neuroscience. *Language, Cognition and Neuroscience*, *32*(4), 401–424.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. London: The University of Chicago Press.

Parks, R. W., Levine, D. S., & Long, D. L. (Eds.). (1998). *Fundamentals of neural network modeling: neuropsychology and cognitive neuroscience*. Cambridge, Mass: MIT Press.

Plunkett, K., & Strömqvist, S. (1990). *The acquisition of Scandinavian languages*. Gothenburg: University of Gothenburg.

Senft, G. (2000). *Systems of nominal classification*. Cambridge: Cambridge University Press.

Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska Akademiens grammatik. Vol. 2: Ord. [The Swedish Academy Grammar, Part 2: Words]*. Stockholm: Norstedts.

# An ecologically rational explanation for set size effects in human cognition

**Ronald van den Berg**[1], Wei Ji Ma[2],

[1]*Department of Psychology, University of Uppsala, Uppsala, Sweden*

[2]*Center for Neural Science and Department of Psychology, New York University, New York, USA*

*ronald.vandenberg@psyk.uu.se*

## Background and motivation

Human cognition is strongly constrained by set size effects in working memory and attention: the precision with which these systems encode information rapidly declines with the number of items, as observed in for example delayed estimation experiments (Fig. 1A-B). In spite of good descriptive models of this phenomenon, no normative models exist. Most models explain set size effects by postulating a fixed total amount of encoding resources (e.g., (Bays, 2014; Palmer, 1990; Sims, Jacobs, & Knill, 2012; Zhang & Luck, 2008). These models predict that precision is inversely proportional to set size, which is often found not to be the case: it declines more slowly in some tasks and faster in others (van den Berg, Awh, & Ma, 2014). Moreover, a conceptual problem with these models is that they do not explain why a highly evolved system would be subject to such a restrictive constraint on cognitive behavior: why would the brain not recruit more encoding resources as set size increases? We propose that the key to understanding set size effects is to take into consideration that stimulus encoding is costly (Attwell & Laughlin, 2001). The brain must thus balance behavioral benefits of high encoding precision against the neural loss it induces: while increasing encoding precision reduces behavioral loss (fewer task errors), it also increases neural loss. We hypothesize that the brain encodes stimuli with a level of precision that minimizes the *combined* behavioral and neural loss and that set size effects may be a consequence of this ecologically optimal encoding strategy.

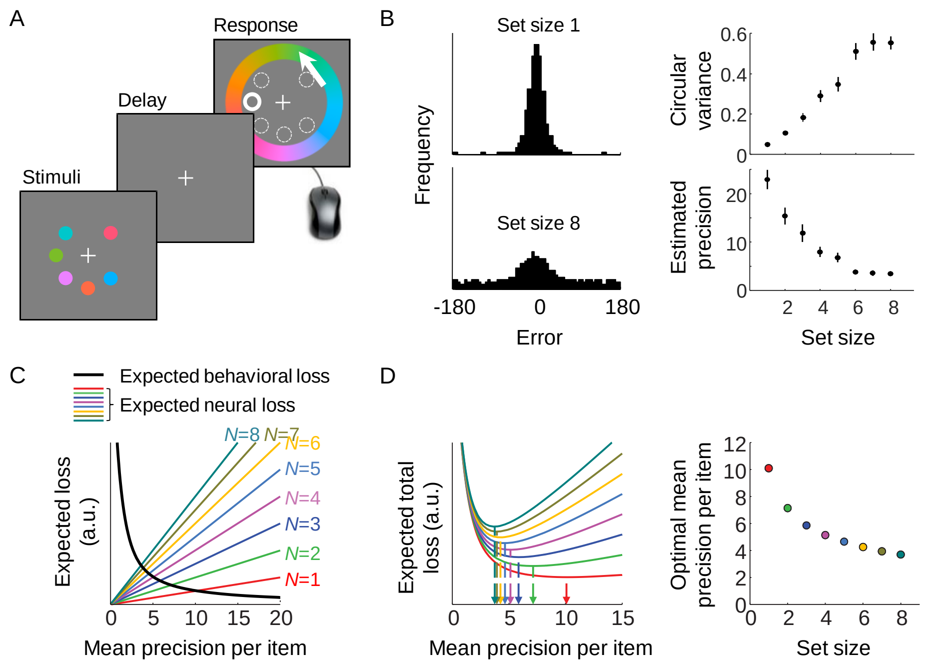



**Figure 1.** An ecologically rational model of set size effects in delayed estimation. (A) Example of a delayed-estimation experiment (set size 6). The subject is briefly presented with a set of stimuli and, after a short delay, reports the value of a randomly chosen target item. (B) Estimation error distributions widen with set size, suggesting a decrease in encoding precision (data from (van den Berg, Shin, Chou, George, & Ma, 2012)). (C) We assume that there are two kinds of loss in this task: a behavioral loss that decreases with encoding precision and a neural loss that is proportional to both set size and precision. In this task, the expected behavioral error loss is independent of set size. (D) Total expected loss has a unique minimum that decreases with the number of remembered items.

## Key theoretical idea

The total expected loss is assumed to be a linear combination of the expected behavioral and neural loss,

$$\langle L_{\text{total}} \rangle = \langle L_{\text{behavioral}}(J,N) \rangle + \lambda \langle L_{\text{neural}}(J,N) \rangle,$$

where $\lambda$ is a free parameter, $N$ denotes set size, and $J$ precision. As in previous work (van den Berg et al., 2012), we quantify encoding precision as Fisher Information. We assume that neural loss is proportional to both precision and set size: $L_{\text{neural}}(J,N)=\alpha JN$, where parameter $\alpha$ is absorbed in $\lambda$. The exact form of $L_{\text{behavioral}}(J,N)$ is dictated by the experimental task. In delayed estimation, we assume that the behavioral loss induced by an estimation error $\varepsilon$ is $|\varepsilon|^{\beta}$, where $\beta$ is a free parameter. Under these assumptions, the expected behavioral loss is a monotonically decreasing function of precision and subject to a law of diminishing returns (Fig. 1C). At each set size, there is a unique encoding precision that minimizes the total loss (Fig. 1D, left). Importantly, optimal encoding precision decreases with set size (Fig. 1D, right), which is qualitatively consistent with set size effects in empirical findings (cf. Fig 1B). To test whether this rational theory can also quantitatively account for set size effects in human cognition, we apply it to data from 4 different experimental tasks.

## Model fits to empirical data

We analyzed data from 6 delayed-estimation experiments (Fig. 1A) that are part of a benchmark data set, comprising a total of 67 subjects (van den Berg et al., 2014). With only three parameters per subject, the maximum-likelihood fits provide an excellent account of the error distributions across all set sizes (Fig. 2). We examine the generality of our theory by testing whether it can also explain set size effects in 3 previously published categorical decision tasks: change detection (Keshvari, van den Berg, & Ma, 2013), change localization (van den Berg et al., 2012), and visual search (Mazyar, Van den Berg, Seilheimer, & Ma, 2013). To do so, we combine the rational encoding strategy presented here with a Bayesian decision model for each task. Single-trial behavioral loss is assumed to be 0 (error) or 1 (correct). We again use maximum-likelihood estimation on individual data sets to fit the models and find that they account well for all data (Fig 3).

## Conclusion

Our results demonstrate that set size effects in visual working memory and attention may be explained as the result of optimally balancing behavioral performance against neural costs. The approach that we took here shares both similarities and differences with the concept of bounded rationality (Simon, 1957), which states that human behavior is guided by mechanisms that provide "good enough" solutions rather than optimal ones. The main similarity is that both approaches acknowledge that human behavior is constrained by various cognitive limitations. However, a crucial difference is that in the theory of bounded rationality, these limitations are treated as postulates or axioms, while we explain them as the rational outcome of an ecological optimization process.

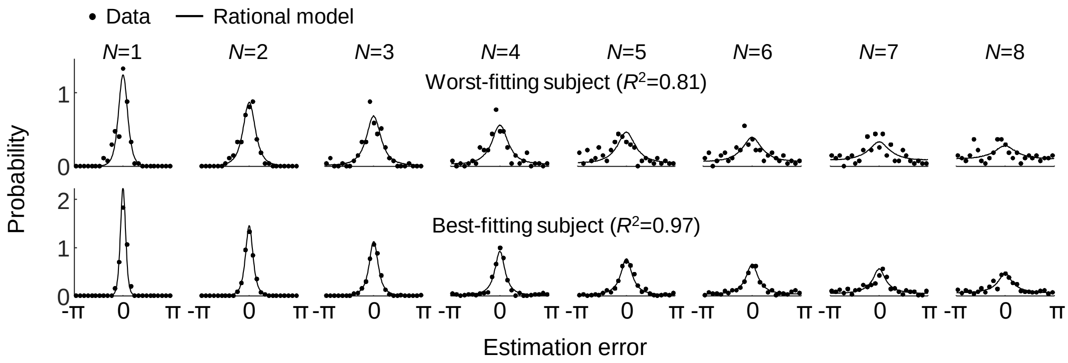



**Figure 2.** Maximum-likelihood fits to raw data of the worst-fitting and best-fitting subjects in the delayed-estimation benchmark dataset. Goodness of fit was measured as $R^2$, computed for each subject by concatenating histograms across set sizes.

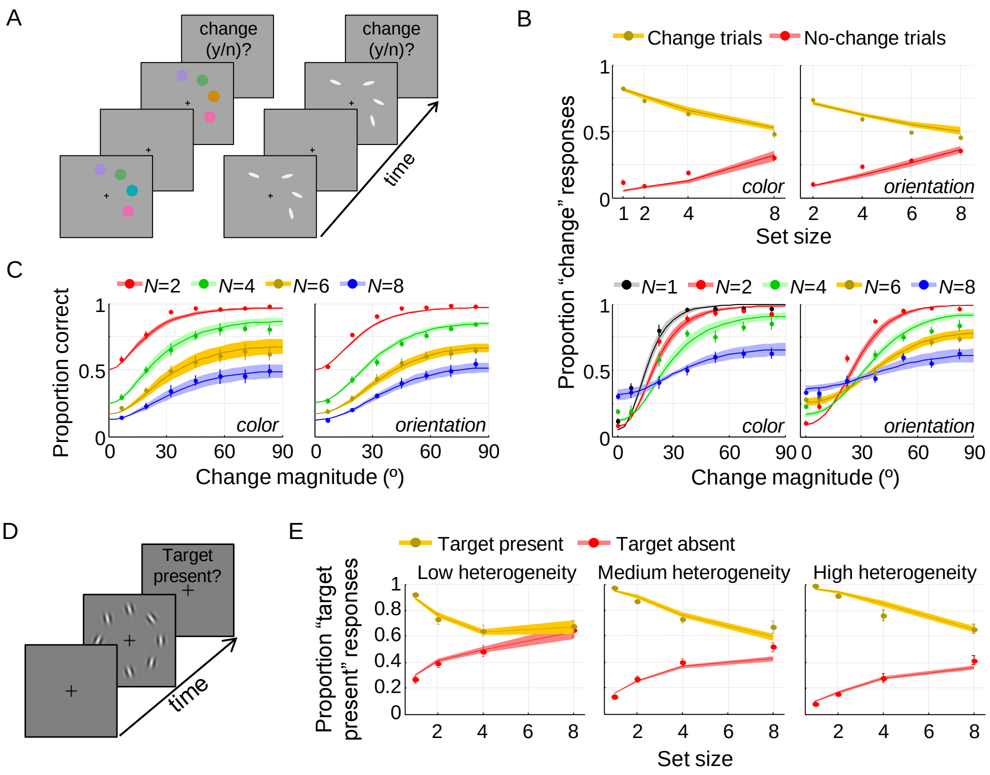



**Figure 3.** Model fits to 3 categorical decision-making tasks. (A) Paradigm in the change-detection experiments. The paradigm for change localization was the same, except that a change was always present and subjects reported the location of change. (B) Model fits to change-detection data. (C) Model fits to change-localization data. (D) Paradigm in the visual-search experiment. (E) Model fits to visual-search data.

## Acknowledgements

R.v.d.B. was supported by grant 2015-00371 from the Swedish Research Council and grant INCA 600398 from Marie Skłodowska-Curie Actions. W.J.M. was supported by grant R01EY020958 from NIH and grant W911NF1410476 from the Army Research Office.

## References

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *21*(10), 1133–1145.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(10), 3632–45.

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No Evidence for an Item Limit in Change Detection. *PLoS Computational Biology*, *9*(2).

Mazyar, H., Van den Berg, R., Seilheimer, R. L., & Ma, W. J. (2013). Independence is elusive : Set size effects on encoding precision in visual search. *Journal of Vision*, *13*(5), 1–14.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(2), 332–350.

Simon, H. A. (1957). *Models of Man*. *Operations Research* (Vol. 5).

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, *119*(4), 807–30.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–49.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.

# Poster Presentations

# Towards a distributed cognition perspective of the Swedish train traffic system

**Rebecca Andreasson**[1], Anders A. Jansson[1]

[1]*Department of Information Technology, Uppsala University*

*rebecca.andreasson@it.uu.se*

In Sweden, train traffic on the operational level is dependent on two main actors: train traffic controllers and train drivers. Train traffic controllers are engaged in a remote control process, monitoring train paths, points, and signals and when necessary, reschedule the current traffic plan with respect to perturbations and disruptions. The remote control process includes two different time frames which in turn require two different ways of working: one in the form of planning ahead and the other in the form of acting directly on feedback from the monitoring system. The main task of the train drivers is to operate their trains by following signals and the current plan set by the traffic controllers. The importance of a well-functioning collaboration between train drivers and train traffic controllers has lately received interest in the Human Factors & Ergonomics (HF&E) research (e.g. Tschirner, Sandblad & Andersson, 2014). This has led to updated work strategies for the traffic controllers that can enable them to continuously re-plan the traffic by projecting the traffic situation instead of the previous strategy to identify and handle conflicts as they occur. The new strategy is called *control by re-planning* (Kauppi, Wikström, Sandblad, & Andersson, 2006) and Kauppi et al. describe that it supports the traffic controllers' need to plan ahead and works as an efficient tool for solving traffic conflicts.

In several studies about the train drivers work situation, their use of information and how information and its availability affect driver behaviour, it is revealed that train drivers suffer from a lack of information (Jansson, Olsson, & Erlandsson, 2006; Jansson, Olsson, & Kecklund, 2005). To obtain relevant information is described as the main challenge in the work role of a train driver and it is concluded that "*…the drivers sometimes found themselves driving in an informational vacuum*" (Jansson et al., 2005, p. 40). Accordingly, while train traffic controllers have access to information about the overall traffic situation, the drivers have to do with the limited information provided by their surroundings, i.e. the trackside signals, and the driver advisory system (DAS) that usually holds static information about speed, current position of the train, and the original time plan. As the traffic controllers get improved possibilities to re-plan the timetable when necessary to adapt it to the current traffic situation, the drivers are facing a more dynamic situation which forces them to plan their driving behaviour based on information that might be outdated. One example of this can be when the traffic controllers decide to lower the speed limit for a specific section of the tracks due to issues with the infrastructure. This will affect all drivers in the area and their possibility to reach the end station on time. However, the static information that the drivers have access to will not display changes in real-time and the driver will know of the change only when he reaches the section in which the permitted speed has been lowered. With earlier access to this information, the driver can adapt his driving behaviour and potentially arrive at the end station on time by adjusting (in this case increasing) the speed wherever the time-table and infrastructure allow it.

The new work strategy in terms of control by re-planning is based on the idea that there can be only one plan for the whole train traffic management system, and that this plan must be made available, at any time, for all actors in the system. If there are perturbations or delays, a new plan is immediately created by the train traffic controller and should be shared among all relevant actors. This metaphoric perspective of the operational train traffic management system is based on the idea that the task of running train traffic is an automatic control engineering task (Andersson, Sandblad, Tschirner & Jansson, 2015). Consequently, the train traffic management system is regarded as a closed loop system, assuming the work of traffic controllers to be a stable and predictable component regardless of the different time frames and the need for different actions. The improvements made are expected to be beneficial for the traffic controllers, but it is still a great deal of uncertainty as to the benefits for the train drivers. Moreover, it is also unclear to what extent the remote control process carried out by the train traffic controllers is consistent with the idea of train traffic management in terms of automatic control and closed control loops. The new work strategy has, so far, not provided a proper solution for realising a robust and efficient train traffic in Sweden. Whether this is due to the fact that the plans have not fully been put into practice, or whether the plans themselves are endangering an appropriate understanding of the conditions for efficient operative train traffic management is not yet possible to decide. In a research project called DIALOG, we challenge the idea of operative train traffic management as automatic control engineering and closed control loops, and instead we propose a perspective based on distributed cognition (Hutchins, 1995a). Generally, train traffic research has focused on either one of the two main roles of traffic controller and train driver. However, this diminishes the importance of a well-functioning collaboration and information sharing activities between these roles. Therefore, we propose that the complex socio-technical system of train traffic must include both traffic controllers and drivers. We argue that the train traffic in Sweden will not reach its full potential unless the two roles are viewed as interdependent parts of the same socio-technical system.

The call for a systems perspective is not new and the relevance of widening the unit of analysis to include not only individual workers but whole socio-technical systems has long been asked for in HF&E research (e.g. Wilson, 2000). When expanding the unit of analysis to include both humans, the multiple types of artefacts they use, and the interactions between all these entities, humans and their actions are understood within their context. In this way, it is easier to grasp and clarify the complexity of a dynamic environment and the effects the social and physical surroundings have on behaviour and performance.

There are several theoretical approaches available for studying work in natural settings and with the socio-technical system as the unit of analysis. However, the theoretical framework of distributed cognition (DCog) has been noted as one of the most pertinent theoretical approaches when it comes to the study of work and interactions between human and technology (Luff, Hindmarsh, & Heath, 2000; Rogers, 2012). DCog focuses on understanding the organisation of complex cognitive systems and proposes that cognition should be studied "in the wild" as it naturally unfolds (Hutchins, 1995a). The underlying principle of DCog is that human cognition is fundamentally distributed within the socio-technical environment and extended to the system level. A main concern is the way information is represented, transformed, and propagated in the performance of tasks in the cognitive system (Hutchins, 1995a). Thus, cognition and knowledge are not viewed as confined within the individual but extended to the system level. DCog has successfully been applied to a wide range of complex domains, such as for example aviation (Hutchins 1995b), and nuclear power plants (Mumaw, Roth, Vicente, & Burns 2000). However, the theoretical framework of DCog, as introduced by Hutchins (1995a), has previously not been applied to the train traffic system.

In the ongoing research project, we would like to explore the train traffic system from a DCog perspective to identify problems, possibilities, and challenges in the Swedish train traffic system. The overall aim of this is to investigate how cognitive processes unfold in the real-world settings of train driving and train traffic control. A secondary aim is to challenge the theoretical framework of DCog by applying it to the highly distributed system of train traffic (to the best of our knowledge, the Swedish train traffic system is more distributed than the previous domains in which DCog has been applied). The work in this project will be done with cognitive ethnography as the basis for data collection. This is not a specific technique or method for analysis but rather a collection of techniques such as interviews, observations, and video recordings (Hollan, Hutchins, & Kirsh, 2000). With an understanding of the functional properties of this socio-technical system, we will build knowledge about the community of practice and reveal how cognitive activities are accomplished in these complex real-world settings of train driving and train traffic control. An application of DCog, and allowing the boundaries of what constitutes cognition to expand into the systems perspective, offers a new toolset for identifying both potential problems as well as opportunities for improvement of the human based work within the train traffic domain.

The ongoing research project embarks on a new perspective of the train traffic domain and will, contrary to prior research on traffic control and train driving, emphasise the collaborative nature of the work and highlight the fact that an understanding of both these roles and their interdependence is essential for realising a successful train traffic. Due to the complexity of the train traffic system, we consider DCog to be a suitable framework that will pave the way for a discussion about the benefits of a systems perspective in Rail Human Factors research. It is our belief that DCog, with its focus on information flow, will offer insights into the challenges related to the collaboration between the highly distributed train drivers and the traffic controllers.

## References

Andersson, A. W., Sandblad, B., Tschirner, S., & Jansson, A. (2015). *Framtida tågtrafikstyrning. Sammanfattande forskningsrapport*. Slutrapport från FOT-projektet. Department of Information Technology, Uppsala University. Retrieved from: http://kajt.org/forskning/rapporter.html

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*(2), 174–196. https://doi.org/10.1145/353485.353487

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, Mass: MIT Press.

Hutchins, E. (1995b). How a Cockpit Remembers Its Speeds. *Cognitive Science, 19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Jansson, A., Olsson, E., & Erlandsson, M. (2006). Bridging the gap between analysis and design: Improving existing driver interfaces with tools from the framework of cognitive work analysis. *Cognition, Technology & Work, 8*(1), 41–49.

Jansson, A., Olsson, E., & Kecklund, L. (2005). Acting or reacting? A cognitive work analysis approach to the train driver task. In J. R. Wilson, B. Norris, T. Clarke, & A. Mills (Eds.), *Rail Human Factors: Supporting the Integrated Railway* (pp. 40–49). Aldershot, UK: Ashgate Publishing Limited.

Kauppi, A., Wikström, J., Sandblad, B., & Andersson, A. (2006). Future train traffic control: control by re-planning. *Cognition, Technology & Work, 8*(1), 50–56.

Luff, P., Hindmarsh, J., & Heath, C. (2000). *Workplace studies- recovering work practice and informing system design*. Cambridge: Cambridge University press.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors: The Journal of the Human Factors and Ergonomics Society 42*(1), 36–55. https://doi.org/10.1518/ 001872000779656651

Rogers, Y. (2012). *HCI Theory: Classical, Modern, and Contemporary*. Morgan & Clay-pool Publishers.

Tschirner, S., Sandblad, B., & Andersson, A. W. (2014). Solutions to the problem of inconsistent plans in railway traffic operation. *Journal of Rail Transport Planning & Management, 4*(4), 87–97. https://doi.org/10.1016/j.jrtpm.2014.10.002

Wilson, J. R. (2000). Fundamentals of ergonomics in theory and practice. *Applied Ergonomics, 31*(6), 557–567. https://doi.org/10.1016/S0003-6870(00)00034-X

# Is media multitasking beneficial for attentional control? Predicting attention shifting abilities from self-reported media multitasking

**Pia Elbe**[1], Daniel Sörman Eriksson[1], Elin Mellqvist[1], Julia Brändström[1], & Jessica K. Ljungberg[1]

[1]*Department of Psychology, Umeå University*

*piel0005@student.umu.se*

Media multitasking is an increasingly prominent topic in affluent societies, and is defined by using multiple sources of electronic media at once, such as sending messages while watching TV. It is not clear if the simultaneous use of several modes of media content has an influence on higher cognitive functions, such as attention (Duff, Yoon, Wang, & Anghelcev, 2014). In this study, attention shifting was the primary focus, because of the idea that switching attention between tasks is necessary for media multitasking. Media multitasking was measured using the Media Multitasking Index (MMI), a questionnaire measuring frequency and intensity of multitasking several sources of virtual media (Ophir, Nass, & Wagner, 2009). The MMI was administered to participants as an online self-report questionnaire via email. The participants who arrived in the lab without having completed the MMI were given a paper version. The MMI measures the frequency of media multitasking as a proportion, and is interpreted by categorizing respondents into high media multitasking or low media multitasking according to questions regarding the use of 11 types of media.

We hypothesized that high media multitaskers would perform significantly better than low media multitaskers on two attentional control tasks. This is in line with the findings of Alzahabi and Becker (2013), because they found the attention switching cost of the number-letter task to negatively correlate with scores on the MMI.

In our study, the number-letter and local-global tasks were chosen, because they measure attention switching (Miyake et al., 2000; Rogers & Monsell, 1995). The number-letter task measures the cost of switching attention by presenting participants with a letter paired with a digit in one of four spatial locations on a screen. The participant must then reply if the digit is odd or even, or if the letter is a consonant or vowel. The response is dependent on the location of the number-letter pair on the screen. For example, the instruction may be to focus on the number if the stimulus is presented in one of the bottom two quadrants. Secondly, the local-global task requires participants to respond with the number of sides a certain shape has. These shapes are present in either a global figure, or in the local figures which the global figure is comprised of. For example, the screen could show many small triangles which are arranged to form a square. The square is the global figure and the triangle is the local figure. Participants must switch their attention from focusing on the local or global figure in order to correctly identify the number of sides.

We included an additional measure, the Cognitive Reflections Test (CRT; Frederick, 2005), as a possible confounding variable, with the purpose to control for intelligence. This task is a short-form intelligence test containing six question that measure participants' ability to reflect on an initial intuitive response, and use logical analysis to come to the correct answer. Participants may be tempted to give a heuristic response, rather than careful examine the seemingly easy problem.

Thirty-six participants between the ages of 20 and 29 completed the MMI, the two behavioral tasks, and the Cognitive Reflections Test (CRT; Frederick, 2005). Once participants completed either the online or paper version of the MMI, they were given both verbal and written instructions for the behavioral tasks. The local-global task and the number-letter task where constructed using the E-Prime software, using the standard procedure of Miyake et al. (2000). In addition to these tasks, participants completed the CRT in the lab.

Contrary to our hypothesis, the results showed that low media multitaskers were significantly better at shifting attention. This replicates the original findings of Ophir, Nass, and Wagner (2009), but it is not in line with the more recent study of Alzahabi and Becker (2013). In the number-letter task, Alzahabi and Becker used a four-button design, while Ophir, Nass, and Wagner used a two-button design. We also used the latter, so similar methods may explain the findings. Since the regression model showed a significant correlation between low media multitasking and attention shifting skills, it could be theorized that frequent media multitasking causes deterioration of attention shifting skills over time. This would mean that there is an element to media multitasking that is not conducive to the development or maintenance of attention shifting abilities. However, we favor the explanation that individuals who perform poorly on behavioral measures of shifting attention are more likely to choose activities where their attention can be more distributed, such as media multitasking. More research is needed to determine possible factors which might predispose a person to become a high media multitasker.

There was also a negative correlation between media multitasking and the Cognitive Reflections Test. When included in the linear regression model, scores on the CRT explained enough variance that the number-letter task was no longer significant. These findings are in line with results from other studies in the area (Frederick, 2005), and in a recent study by Schutten, Stokes, and Arnell (2017) similar results were found when using the same set-up as in the present study (using the CRT and the MMI). This could indicate that intelligence is an important aspect of a possible relationship between attention switching and media multitasking. The individuals who are less prone to reflect logically before reaching conclusions, as indicated by the CRT, are also more likely to be the ones who report media multitasking.

How intelligence and attentional control both relate to media multitasking is an interesting question. Since the CRT requires time for focusing on a single problem, what high media multitaskers and low scorers on the CRT may have in common is unwillingness to prioritize attention on a single stimulus. Our results could be related to how much stimulus activation individuals prefer. Individuals who prefer to distribute their attention during behavioral tasks (such as focusing on distractor stimuli) may also prefer to distribute their attention to several sources of media. The negative correlation of the CRT and media multitasking can be explained with the same reasoning. The CRT requires focus on one single problem. High media multitaskers seem to exhibit a breadth-bias, which favors multiple activation in different scenarios. Conversely, low media multitaskers may be more motivated to attend to only one aspect of a task at once.

The limitations of this study are, besides the obvious small sample size, possible order effects because the number-letter task was always administered after the local-global task. This could explain higher variance in the number-letter scores. This experiment should be replicated with a larger sample size and a complete randomized design. Further research could include personality factors which might influence media multitasking, such as sensory processing sensitivity (Aron & Aron, 1997) or introversion-extraversion traits. It would be interesting to see if, or to what extent, personality factors are responsible for the relationships found in this study. For future research, using participants from populations with attentional deficits, or similar, would be of interest (Seo, Kim, & David, 2015).

## References

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. Journal of Personality and Social Psychology, 73(2), 345-368.

Alzahabi, R., & Becker, M. W. (2013). The association between media multitasking, task-switching, and dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1485–1495.

Duff, B. R. L., Yoon, G., Wang, Z. G., & Anghelcev, G. (2014). Doing it all: an exploratory study of predictors of media multitasking. *Journal of Interactive Advertising*, *14.*

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15583–7.

Rogers, R. D., & Monsell, S. (1995). Costs of a Predictable Switch between Simple Cognitive Tasks. *Journal of Experimental Psychology*, *124*(2), 207–231.

Schutten, D., Stokes, K. A., & Arnell, K. M. (2017). I want to media multitask and I want to do it now: Individual differences in media multitasking predict delay of gratification and system-1 thinking. *Cognitive Research: Principles and Implications*, *2*(1), 8.

Seo, M., Kim, J.-H., & David, P. (2015). Always Connected or Always Distracted? ADHD Symptoms and Social Assurance Explain Problematic Use of Mobile Phone and Multicommunicating. *Journal of Computer-Mediated Communication*, *20*(6), 667–681.

# Curiosity and expected information gain in word learning

**Linus Holm[1]**, Gustaf Ådén-Wadenholt[1] & Paul Schrater[2]

[1] *University of Umeå*

[2] *University of Minnesota*

*Linus.Holm@psy.umu.se*

Curiosity is a pervasive phenomenon that drives exploration and invention, and may reflect an intrinsic motivation for learning. Evidence suggest that information is rewarding in and of itself (Bromberg-Martin & Hikosaka, 2009; Katz & Gelbart, 1978; Bevins et al., 2002) and that curiosity may reflect the anticipation for an information- ((Kang et al., 2009; Jepma, Verdonschot, van Steenbergen, Rombouts, & Nieuwenhuis, 2012) or even learning (Berlyne, 1950; Marvin & Shohamy, 2016; Kang et al., 2009) reward. However, it remains unclear what quantity instills curiosity. The purpose of this study was to clarify how curiosity is computed. Already in the 1950's, Berlyne (Berlyne, 1950; Berlyne, 1957)) suggested that curiosity should be investigated using information theory. But to date, empirical studies of information theoretic tests of curiosity have been scarce. From an information theoretic viewpoint, a straight-forward prediction is that the more information an exploratory action is *expected* to yield, the more likely it is that the exploratory action is taken. For instance, an unfamiliar word may trigger your curiosity and compels you to look it up, if you think you will understand the word's definition from the thesaurus, and know how to use a thesaurus. Your prior uncertainty is reduced by looking up the word, and advances your semantic knowledge accordingly. Additionally, the meaning may be surprising, forcing you to adjust your general confidence in word knowledge. Thus, the simple action of checking the word may also calibrate your metacognitive knowledge. Therefore, curiosity may not only reflect the simplistic anticipation for any information reward, but rather a drive for improving predictive models. Importantly, the prediction that uncertainty drives exploration in curiosity goes against current standard curiosity theory (Loewenstein, 1994; Kang et al., 2009), according to which exploration (as manifested curiosity) as well as subjective curiosity exhibits an inverted u-function of uncertainty. To begin to test our prediction, we devised an experiment which imposed a time penalty on the *exploratory action of revealing the correct alternative in a forced choice word synonymy task*. This trade-off between information and time allowed us to quantify information investment in seconds per unit information. Curiosity in this experiment is assumed to be reflected in the (exploratory) action of choosing to wait in order to receive the correct synonymy answer. Individual assessments of 87 participants' prior word knowledge was used to compute expected information gain on correct synonymy reveal. We found strong support for the hypothesis that expected information gain drives exploration and self-rated curiosity; participants were 19.5 times more likely to reveal the answer per bit of expected information. Moreover, learning benefits scaled with information gain as assessed by a surprise memory test one day later. Importantly, the expected information gain predictor we employed assumed participants used an accurate model of uncertainty. Conversely, without this representation of uncertainty, as in model-free reinforcement learning, exploration could not be predicted. Thus, our findings go beyond the simplistic notion that curiosity reflects anticipation for information reward and suggest curiosity reflects a drive for improving predictive models.

## Method

Each of 87 participants completed three daily sessions. Day 1, Participants rated the synonymy on a scrollbar scale of 1 to 100, for each of 800 word-pairs. The words were taken from the swedish sholastic aptitude tests between 2004-2011, and were selected to produce a uniform distribution based on the performance of the original test takers. In session 2, these pairwise synonymy ratings were used to compute expected information gain from revealing the correct synonym out of five alternatives (see equations 1 and 2). Importantly, participants could opt to wait 6s to see the correct synonym on each trial or skip to the next trial (see *Figure 1*) thus imposing a time penalty on synonymy reveal.

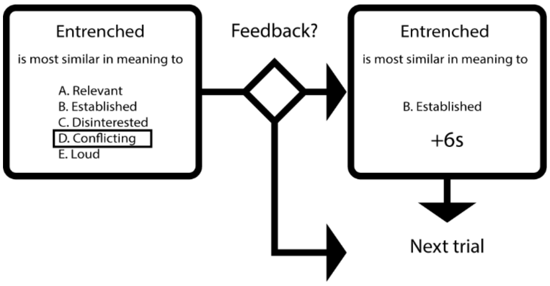



*Figure 1. Experimental design session 2. Target word and five response alternatives were presented for 7s while the subject indicated her guess. The subject then made a reveal decision within 3s. On reveal, the target word and the correct response alternative were displayed for 6s, and then the next trial started. If a reveal was not requested, the next trial started immediately after the reveal decision.*

*Independent variables*. The five synonymy pair-ratings (i.e., one target word and five synonym options) were first normalized. We verfied that the highest uncertainty appeared at ratings near 50 because participant reponse distributions exhibited an inverted u-function peaking at 50. To cover this property in our estimate of each subjects' uncertainty, we weighted the normalized distribution according to:

$$p(\hat{x}_i) = \bar{x} * (1 - argmax(P(X))) + argmax(P(X)) * p(x_i) \qquad \text{[eq. 1]}$$

Where $x_i$ indicates one of the five normalized ratings, $P(X)$ is the normalized rating distribution, $\bar{x}$ is the mean of that distribution, and $\hat{x}_i$ is the estimated probability. We then computed the expected information grain from reveaing the correct synonym in this distribution. Because the posterior entropy on reveal equals zero, the expected information gain is exactly the initial entropy as given by:

$$E(I) = -\sum_{i=1}^{n} p(\hat{x}_i) * log_2(p(\hat{x}_i)) \qquad \text{[eq. 2]}$$

## Results

*Expected information gain drove reveal requests*. Logistic functions were fitted to each participant's reveal choice data with expected information gain as predictor. The logistic parameters were averaged and the corresponding logistic function is displayed in Figure 2. The average parameters were $\beta_0$ = -2.9814, $p < 0.012$, $\beta_{infogain}$ = 2.9685, $p < 0.009$. Thus, reveal choice proportion increased by 19.5 per bit of expected information gain.

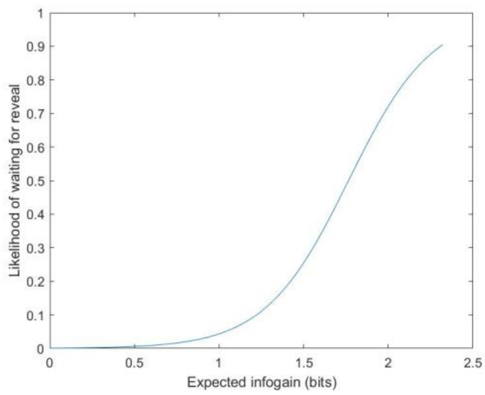



*Figure 2. Proportion of reveal requests as a function of expected information on reveal. Orange line is averaged logistic function across subjects.*

*Learning scaled with expected information gain on the reveal*. At the surprise memory test constituting the third session, participants again took the five alternative synonymy test. Average improvements were computed for revealed and not revealed cases (see Figure 3) as a function of expected information gain. $\beta_{\text{infogain}} = .88$, $p < 0.01$ and $\beta_{\text{infogain}} = 1.37$, $p > 0.22$. given feedback taken and not taken, respectively.

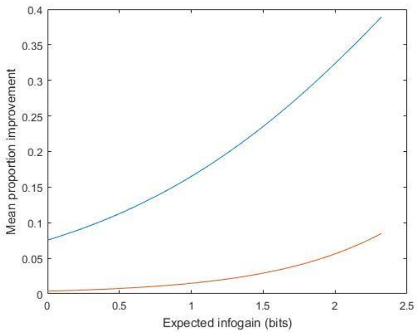



*Figure 3. Blue line represents learning improvement given feedback request, red line learning given no feedback request, both conditioned on entropy ratings. Both lines are fitted logistic functions.*

*Reinforcement learning model*. To test if a model-free learning policy predicted reveal choices, we implemented several versions of an exponential decay-weighted moving average on feedback. Specifically, the prediction error between the moving average and the last observation was used as a predictor of reveal choice. Moreover, to test across-trial learning, we permuted (10k samples per participant) the time series and averaged the logistic parameters of the simulations for each subject. Neither the empirical order nor the permuted order displayed any significant relation between prediction error and reveal choice: empirical order yielded $\beta_0 = -0.10$, $p > .23$, $\beta_1 = -0.49$, $p > .48$, and the permuted orders produced average parameters $\beta_0 = 0.65$, $p > .17$; $\beta_1 = 0.81$, $p > .48$.

## References


Berlyne. (1950). Novelty and curiosity as determinants of exploratory behavior. *British Journal of Psychology*, *41*, 68–80.

Berlyne, D. E. (1957). Conflict and information-theory variables as determinants of human perceptual curiosity. *Journal of Experimental Psychology*, *53*(6), 399–404.

Bevins, R. A., Besheer, J., Palmatier, M. I., Jensen, H. C., Pickett, K. S., & Eurek, S. (2002). Novel-object place conditioning: behavioral and dopaminergic processes in expression of novelty reward. *Behavioural Brain Research*, *129*(1–2), 41–50. https://doi.org/10.1016/S0166-4328(01)00326-6

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119–126. https://doi.org/10.1016/j.neuron.2009.06.009

Jepma, M., Verdonschot, R. G., van Steenbergen, H., Rombouts, S. a. R. B., & Nieuwenhuis, S. (2012). Neural mechanisms underlying the induction and relief of perceptual curiosity. *Frontiers in Behavioral Neuroscience*, *6*(February), 1–9. https://doi.org/10.3389/fnbeh.2012.00005

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

Katz, R. J., & Gelbart, J. (1978). Endogenous opiates and behavioral responses to environmental novelty. *Behavioral Biology*, *24*(3), 338–348. https://doi.org/10.1016/S0091-6773(79)90197-4

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266–272. https://doi.org/10.1037/xge0000140

# Psychotherapists' interest in using the Furhat social robot for clinical training

**Robert Johansson**[1,2], Sam Thellman[1], Gabriel Skantze[3], & Arne Jönsson[1]

[1]*Department of Computer and Information Science, Linköping University, Sweden*

[2]*Eailab AB, Stockholm, Sweden*

[3]*KTH Speech, Music and Hearing, Stockholm, Sweden*

*robert.johansson@liu.se*

## Introduction

For more than two decades it has been possible for surgeons and other medical specialists to practice in a simulated environment. This has enabled a safe way to practice complicated procedures systematically. However, for clinical psychologists and psychotherapists very few options exist for similar training procedures.

Recently, we created and demonstrated a prototype of such a simulated environment (Johansson, Skantze, & Jönsson, 2017). The practice environment uses the Furhat social robot (Moubayed, Beskow, Skantze, & Granström, 2012) and the open-source toolkit IrisTK (Skantze & Al Moubayed, 2012) to implement simulated patients in a dialog system. A goal of our research has been to develop means for training, where it would be possible to practice psychotherapeutic methods on virtual patients in a similar way to that of medical training simulators. While there exist studies on using social robots as part of therapy (e.g., Feil-Seifer & Mataric, 2008), there seem to be hardly any research conducted on using social robots as patients for training purposes. We believe this could open up for a completely new approach to training within psychiatry and clinical psychology.

It is however unknown if psychologists and psychotherapists have an interest in this type of training tool. In this study, we aim to explore this, as well as other dimensions of perceptions of the Furhat robot in this population. Finally, we look at the predictive values of these dimensions towards the clinicians' interest in using the Furhat robot as a training tool.

## Methods

Details of the psychotherapy training environment has been described elsewhere (Johansson et al., 2017). Briefly, the patients are implemented as conversational agents using a statechart-based XML formalism for designing the dialog flow. Hence, the current implementation relies on hand-crafted rules.

A prototype of the psychotherapy training environment was presented to an audience of prospective end users at a yearly psychotherapy workshop held in Stockholm, Sweden. The theme of the workshop was an advanced application of a form of psychotherapy called Intensive Short-term Dynamic Psychotherapy (ISTDP) in a challenging group of patients. A total of about 65 therapists took part of the presentation, a majority of which had several years of clinical experience. First, the general idea of the system and its overall purpose was presented. Then, a video demonstration was given showing a psychotherapist interacting with the Furhat robot as a simulated patient. Figure 1 shows an example of an interaction with the simulator, as illustrated to the audience. This was followed by a Q&A session. After this, a survey was distributed to the psychotherapists.

The survey contained the GODSPEED questionnaire (Bartneck, Kulić, Croft, & Zoghbi, 2009) and two additional questions: *"How interested would you be in using a simulator such as the one described for your own training?"* and *"How valuable do you think a simulator such as the one described could be for [Intensive Short-term Dynamic Psychotherapy] training in general?"* (Respondents gave their answers by marking a position on a numbered five-step Likert-style sequence ranging from *"Not at all"* to *"Very much"*). The GODSPEED questionnaire is a collection of instruments for measuring the dimensions *anthropomorphism*, *animacy*, *likeability*, *perceived intelligence*, and *perceived safety of robots* (Bartneck et al., 2009).

To explore the value of these dimensions in predicting A) the therapists' interest in using the simulator for own training and B) the perceived value of the simulator for the community, two multiple linear regression analyses

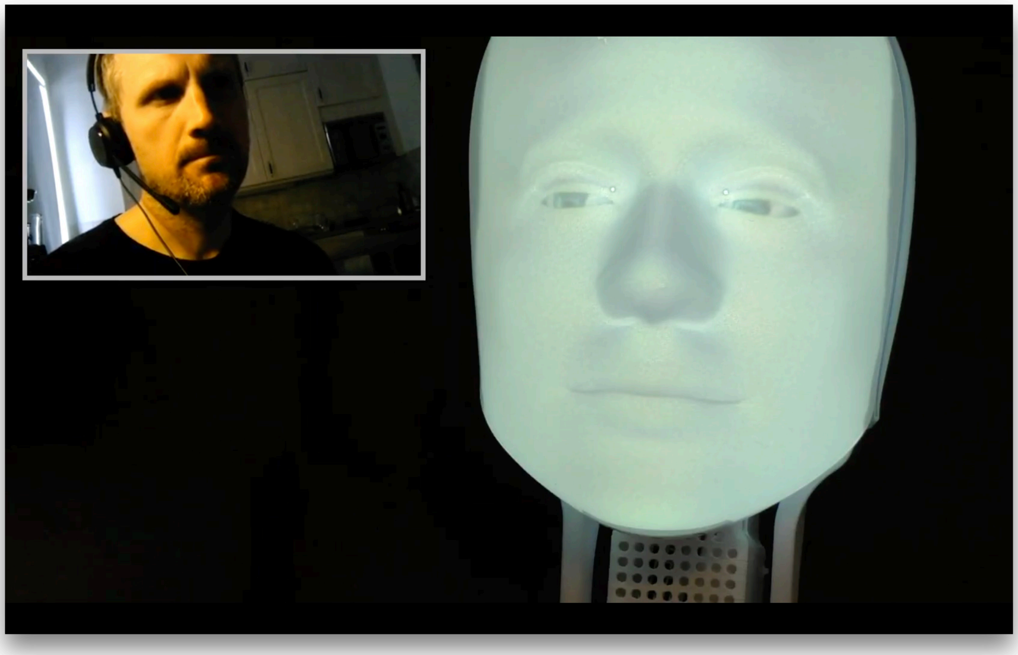


**Figure 1**. An example interaction with the training environment

were conducted with A and B as target variables and individual GODSPEED dimensions as predictors. Backward stepwise selection with a selection criteria of $p < .10$ was used to decide the final set of predictors in the models.

## Results

Thirty-nine therapists completed the survey. The mean interest in using the simulator for own training was 3.95 (*SD* = 1.40, *Md* = 4), and the mean perceived value for the community was 4.08 (*SD* = 1.11, *Md* = 4). Mean values on the GODSPEED scales can be seen in Table 1.

**Table 1. Mean values and standard deviations of ratings on the GODSPEED scales.**

| Anthropomorphism | Animacy | Likeability | Perceived Intelligence | Perceived Safety Beginning | Perceived Safety End |
|---|---|---|---|---|---|
| 2.65 (0.60) | 2.90 (0.60) | 3.63 (0.85) | 3.07 (0.56) | 3.44 (0.71) | 3.71 (0.65) |

The GODSPEED dimensions *animacy* ($p = .002$) and *likeability* ($p = .053$) was selected for inclusion in the model. A significant regression equation was found ($F(2, 36) = 19.463, p < .001$), with an $R^2$ of .520. The model predicting perceived value for the community also included *animacy* ($p = .001$) and *likeability* ($p = .064$), with a significant regression equation $F(2, 36) = 19.441, p < .001$, $R^2 = .519$.

## Discussion

This study found preliminary indications that clinicians' interest in using the Furhat robot as a training tool was high. Furthermore, the GODSPEED dimensions *animacy* and *likeability* seem to have a predictive value, explaining more than half of the variance in the clinicians' interests. These results indicate that a psychotherapy training environment using a social robot seem feasible. Arguably, two strengths of the Furhat robot are its highly realistic facial expressions, and its perceived "friendliness". Hence, the results from this study point in the direction that the Furhat robot is indeed a good choice for this kind of training environment. Further research is warranted in this area.

**Acknowledgements**: The development of the application described in this paper was supported by a grant from Vinnova (2017-00727). We also want to acknowledge professor Allan Abbass, Dalhousie University, Halifax, Canada, for allowing us to demonstrate the system and collect data at his workshop.

## References

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1(1)*, 71–81.

Feil-Seifer, D., & Mataric, M. (2008, June). Robot-assisted therapy for children with autism spectrum disorders. In *Proceedings of the 7th international conference on Interaction design and children* (pp. 49-52). ACM.

Johansson, R., Skantze, G., & Jönsson, A. (2017). A psychotherapy training environment with virtual patients implemented using the Furhat robot platform. In *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings* (Vol. 10498, pp. 184–187). Springer.

Moubayed, S. A., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. In *Cognitive Behavioural Systems* (pp. 114–130). Springer, Berlin, Heidelberg.

Skantze, G., & Al Moubayed, S. (2012). IrisTK: A Statechart-based Toolkit for Multi-party Face-to-face Interaction. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (pp. 69–76). New York, NY, USA: ACM.

# Using Eye-Tracking to Study the Effect of Haptic Feedback on Visual Focus During Collaborative Object Managing in a Multimodal Virtual Interface

**Jonas Moll**[1], & Emma Frid[2]

[1]*Department of Information Technology, Uppsala University*

[2]*Department of Media Technology and Interaction Design, KTH Royal Institute of Technology*

*jonas.moll@it.uu.se*

Haptic feedback provides feedback to our sense of touch. The effects of haptic feedback on task performance and dual-task capacity in multimodal virtual environments have been extensively studied (Burke et al., 2006). Moreover, the effects of this type of feedback on collaboration and communication in collaborative virtual environments are gaining more and more interest (Moll, 2013; Sallnäs, 2004). However, none of the previous studies on this topic has, to our knowledge, investigated how haptic feedback affect visual attention during collaborative object handling. The pilot study described in this abstract extends the work on collaborative haptic interfaces to also involve visual attention, by employing eye-tracking methodologies. Eye tracking is a technique in which an individual's eye movements are measured in order to detect where this person is looking at a specific time (Pool and Ball, 2006). Common eye tracking metrics include measures of fixations, such as number of fixations and total fixation duration. Fixations are moments when the eyes are relatively stationary due to information processing.

In our previous work (Frid et al, 2017), we presented an exploratory study on the effect of auditory feedback on gaze behavior in conditions with versus without haptic feedback. Analysis of eye tracking metrics indicated large inter-subject variability and the difference between subjects was greater than the difference between feedback conditions. No significant effect of feedback type was observed, but clusters of similar behaviors were identified, and certain participants appeared to be affected by the presented auditory feedback.

Following up on this study, we developed a collaborative interface in which two users had to work together in order to move an object to a defined destination. The graphical interface is similar to the one used in Frid et al. (2017), but the task is different; two users may move an object by pushing it from each side, the object can then be lifted and placed on top of one of two pillars (the interface and setup is shown in Figure 1). The two users performed the experiment in the same room, with two separate displays. The users could talk to each other, but a white screen separated them.

**Figure 1. Experimental setup.**

A pilot experiment was conducted using the above-described collaborative interface. The experiment had a between-group design. Six pairs participated, three pairs had access to haptic feedback (could feel all parts of the workspace, including the object and forces from the other user when moving the object) and three pairs did not experience any haptic feedback. The following hypotheses were tested in the study:

H1: Users will focus significantly longer on the target areas (i.e. the pillars) in the haptic condition than in the nonhaptic condition.

H2: Users will have significantly more fixations and visits (glances) in the haptic condition than in the nonhaptic condition.

Although we found no significant differences for the investigated eye tracking metrics (fixation count, total fixation duration, visit count, and total visit duration) between the haptic and non-haptic groups, interesting tendencies arose, related to the effect of haptic feedback on visual focus.

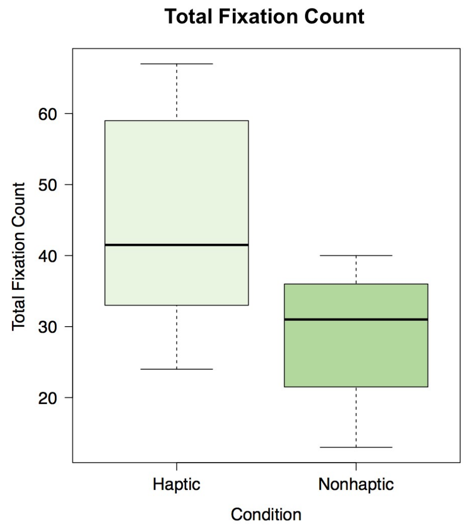



**Figure 2. Box plots of fixation counts for the left and right target (i.e. pillar) (n=4 for nonhaptic case, n=6 for haptic case).**

Boxplots of fixation counts for the left and right target (pillar) can be seen in Figure 2, with lower median and smaller interquartile ranges for the nonhaptic case. A heat map analysis on total fixation duration also indicated that participants focused more on the target areas in the haptic case. Even though the null hypotheses could not be rejected, the results indicated that a larger sample size might result in significant findings. The pilot experiment results suggested that haptic feedback could indeed affect gaze behavior during joint object manipulation in virtual environments. These tendencies encourage us to move forward with future experiments with a similar setup. The next step of the study would involve adding movement sonification to the task in order to evaluate the effect of auditory feedback (and combined auditory and haptic feedback) on gaze behavior, similarly to the previous study performed by Frid et al. (2017).

## References

Burke, J., Prewett, M.S., Gray, A.A., Yang, L., Stilson, F.R.B., Coovert, M.D., Elliot, L.R., and Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. Proceedings of the 8th international conference on Multimodal interfaces (Banff, Canada, November 2006). pp. 108-117.

Frid, E., Bresin, R., Pysander, E. L. S., & Moll, J. (2017). An Exploratory Study On The Effect Of Auditory Feedback On Gaze Behavior In a Virtual Throwing Task With and Without Haptic Feedback. In Sound and Music Computing (SMC) 2017, pp. 242-249.

Moll, J. (2013). The influence of Modality Combinations on Communication in Collaborative Virtual Environments. Doctoral thesis, School of Computer Science and Communication, Royal Institute of Technology, Stockholm.

Poole, A. and Ball, L. J. (2006) Eye Tracking in HCI and Usability Research. Encyclopedia of human computer interaction, vol. 1, pp. 211–219.

Sallnäs, E-L. (2004). The effect of modality on social presence, presence and performance in collaborative virtual environments. Doctoral thesis, School of Computer Science and Communication, Royal Institute of Technology, Stockholm.

# Perceived Intelligence and Protégée Effect in a Techable Agent Software

**Kristian Månsson**[1], Magnus Haake[1], Agneta Gulz[1]

[1]*Lund University Cognitive Science*

*tkman85@gmail.com*

## 1 Introduction

Studies have shown that the "teachable agent" paradigm, i.e. "learning-by-teaching" using teachable virtual agents in educational software, benefits learning by increasing students' sense of responsibility and supporting metacognition (see for instance Schwartz et al. (2009) and Biswas et al. (2005)). The "protégée effect" is a theoretical concept that describes the beneficial factors of the teachable agent paradigm in that the student makes larger learning efforts when the goal is to teach an agent than when the goal is to learn for themselves (Chase, Chin, Oppezzo & Schwartz, 2009). The authors hypothesize that the following three mechanisms lead to the increased effort of learning: a feeling of responsibility towards the teachable agent, an adoption of an incrementalist view of knowledge, and a protection of their ego since it is the agent that is tested for its learning and potentially fails.

Elaborating on the teachable agent paradigm there is a difference between an agent that can learn and an agent that can be taught (Pareto, Schwartz & Svensson, 2009), leading to different approaches for the design of the AI in teachable agent software. This, in turn, spurs an underlying question of how much effort should be put into the development of the underlying artificial intelligence of a teachable agent. It needs to be teachable, but to what extent does it need to learn versus seem to learn. In order to pursue this question, this study aims to investigate the connection between the protégée effect, perceived intelligence of the teachable agent, and students' learning outcomes.

## 2 Experiment

The teachable agent (hence TA) educational software used in this study is called Guardians of History (GoH). GoH is aimed at middle school history education (year 4 to 6 in the Swedish educational system) and is developed by the Educational Technology Group at Lund University and Linköping University. In GoH, the student makes time travels to gather information by exploring scenes and interacting with historical characters from different eras in order to subsequently teach the TA in so-called classroom activities (see Figure 1). For this study, a subset of the possible time travels with associated teaching activities were selected for the experiment. After being taught the TA conducts a test, where it provides answers to questions depending on facts it has learned (see Figure 1). The TA was implemented with two different settings: $TA_R$ (as in recency) and $TA_A$ (as in associative). The recency setting ($TA_R$), corresponds to the original implementation of the TA, where the agent learns (and unlearns) the latest fact that it has been exposed to in each learning activity. That means that for every new learning activity, the agent overwrites all the previous facts learned from previous learning activities. The associative agent ($TA_A$), developed as a part of a university project course (Bäckström, Månsson, Persson, Sakurai & Karåker Sundström, 2017), is implemented with a basic associative model using weighted concept relations representing the agent's certainty of different facts. The weights of the concept associations increase or decrease depending on the results of the learning activities. Furthermore, the $TA_A$ asks for confirmation of learned facts at random intervals.

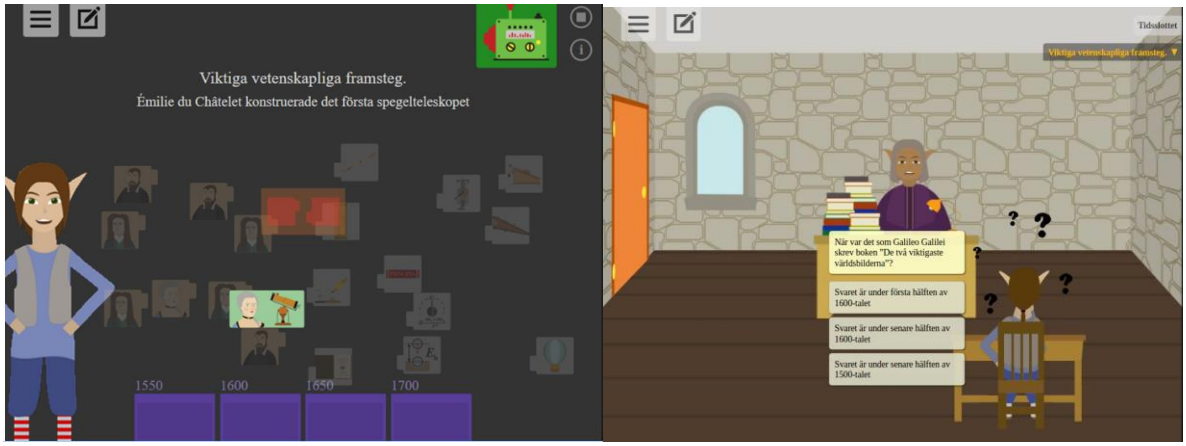



**Figure 1.** Learning activities with the TA to the left. To the right the test setting.

For the experiment, 94 Swedish grade 5 and 6 students from 5 classes from the same Swedish school were recruited. The students were randomly assigned to one of two groups of 47 students each. In order to play down intervention effects, no information on sex or age was explicitly collected for this study – relying on the random assignment of half the students to one group and half to the other in every class to control for sex or age differences. For the analyses, 3 students were excluded from the dataset due to language difficulties. Another 3 students were excluded as they (for unknown reasons) didn't finish the game. An additional 3 students didn't fill out or hand in their questionnaires, resulting in a data set of $N$ = 85 participants with 41 in group $TA_R$ (with 24 in 5th grade and 17 in 6th grade) and 44 in $TA_A$ (with 25 in 5th grade and19 in 6th grade).

The Protégée-effect (PE) was measured using 5 Likert-scale items operationalized by Kirkegaard (2016) for studies with GoH. Perceived intelligence (PI) was measured by 6 items on a semantic difference scale, adapted and translated from Bartneck, Kuli & Croft (2009) who used it for measuring perceived intelligence of a robot. A knowledge test of 10 multiple choice questions based on the content of the game was used to assess the knowledge gained.

## 3 Results

An independent samples $t$-test was conducted to compare perceived intelligence (PI) in the $TA_R$ and $TA_A$ conditions. Results showed no significant difference in PI between the $TA_R$ ($M$ = 16.4, $SD$ = 4.5) and the $TA_A$ ($M$ = 17.8, $SD$ = 4.3) conditions ($t(83) = -1.32$, $p$ = .19), i.e. there was no significant difference between the $TA_R$ group and the $TA_A$ group with regard to perceived intelligence as measured by the questionnaire items.

A calculation of the Pearson product-moment correlation coefficient to assess the relationship between PI and PE suggested a large effect (Cohen, 1988) positive correlation ($r$ = .64, $p$ < .001). The strong correlation indicates that the students' self-reported perceived intelligence of their TA and their self-assessed evaluation of a protégée effect goes hand in hand.

A Matt-Whitney's U test was used to compare the score on the knowledge test between the $TA_R$ group ($M$ = 0.6; *Range* = 1-10) and the $TA_A$ group ($M$ = 0.65; *Range*: 0-9). No significant difference between the groups ($W$ = 843, $p$ = .60) could be found, i.e. neither the $TA_R$ group nor the $TA_A$ group performed better as measured by the knowledge test.

Using a Matt-Whitney's U test, no significant correlation between PE and performance score in the knowledge test could be established. No additional interaction effects could be found. No other significant correlations except for the ones stated were found.

## 4 Discussion

The students did not perceive $TA_A$ as more intelligent. The lack of any significant difference between the groups regarding the perceived intelligence and protégée-effect might point to other factors - such as narrative or the explicitly stated role for the student as the teacher - to elicit the sought-after protégée effect.

The strong positive correlation between perceived intelligence and protégée effect could be interpreted as either that one strongly influences the other or they strengthen each other reciprocally. Another possible explanation is that student's general positive attitude towards the game is what is measured. The strong correlation between how the student either actively ascribe or passively perceive the TA as a thinking and learning agent and the elicited protégée effect, points to one having a strong influence on the other, or both are a manifestation of an underlying phenomena. This correlation should be of interest for researchers as well as designers of teachable agent software.

A lack of correlation between the protégée effect and the score in the knowledge test is somewhat surprising. As the protégée effect is constructed as a theoretical attempt to explain the positive learning outcomes from teachable agents, it might be too coarse to be measured in this way i.e. the result might be sensitive to false negatives. Another possible explanation is that the measurement does not actually reflect the protégée effect; the measurement might rather reflect, or be strongly influenced, by a student's general positive - or negative - attitude towards the software.

A longer study where the students would have more time to interact with the TA might provide further insight towards how the students' perception of the TA varies over time. Validation of the measurements is also of great importance, as they are newly adapted for this study.

## References

Bartneck, C., Kuli, D., & Croft, E. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International journal of social robotics, 1*(1), 71–81.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt University. (2005). Learning By Teaching: a New Agent Paradigm for Educational Software. *Applied Artificial Intelligence, 19* (3–4), 363–392.

Bäckström, J., Månsson, K., Persson, M, Sakurai, E. & Sundström, K. (2017). Designing and Implementing an Associative Learning Model for a Teachable Agent. In Balkenius, C., Gulz, A., Haake, M. & Wallergård, M. (Ed.), *Intelligent, socially oriented technology III: Projects by teams of master level students in cognitive science and engineering.* Lund University

Chase, C., Chin, D. B., Oppezzo, M., Schwartz D.L. (2009) *Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning*, Stanford University School of Education.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Oxford, England: Routledge

Kirkegaard, C., (2016). *Adding Challenge to a Teachable Agent in a Virtual Learning Environment*, Diss., Linköping University

Pareto, L., Schwartz, D. L., & Svensson, L. (2009). Learning by Guiding a Teachable Agent to Play an Educational Game. *Frontiers in Artificial Intelligence and Applications, 200* (1), 662–664.

Schwartz, D.L., Chase, C., Chin, D., Oppezzo, M., Kwong, H., Okita, S., Roscoe, R., Hoyeong, J. Wagster, J. 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*, 340-358

# An exploration into applying predictive processing as framework on critical thinking

**Anders Persson**

*Department of Information Technology, Uppsala University*

*anders.persson@it.uu.se*

Defining critical thinking (CT) has turned out to be a difficult endeavour and many scholars differ in their conceptualization of the activity, even if many have overlapping elements. For example, the often cited Delphi Report defined CT to be: "purposeful, self-regulatory judgement which result in interpretation, analysis, evaluation, and inference, as well as explanation of evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione, 1990). Another popular spokesman for methodologies for teaching CT defines it as "skilful, responsible thinking that facilitates good judgement because it (a) relies upon criteria, (b) is self-correcting, and (c) is sensitive to context" (Lipman, 2003). Recognizing the diversity of definitions, Moore (2013) conducted an interview survey of various field experts, in philosophy, history and literature. The most highlighted aspects of CT in these contexts were, CT as: (1) judgement, (2) scepticism, (3) simple originality, (4) sensitive reading, (5) rationality, and slightly less prelevant, but also as (6) an ethical, activist stance, and (7) as self-reflexivity.

One problem with this diversity is that methodologies for how to teach, and evaluate, CT also differ among conceptualizations, and may be difficult to compare. The level of analysis is, as above, often on high, abstract levels, and it could perhaps be possible to understand these activities on cognitively lower levels of analysis. This would not have to exclude higher and more abstract definitions like the once mentioned, but there could lay a hope in getting a deeper understanding of the process behind it all. The past couple of decades this has been done by looking at neurological structures, like that of Dual-Process theories (DPT) with Greene et al (2004) in the forefront of this field. To some extent this has been insightful in understanding activities like moral reasoning, but at the same time it does not say *a lot* about what is going on in the process of reasoning; what is going on in what is often called the slow and effortful System 2.

Another cognitive perspective that has gained momentum the past few years, that has a hope in being able to say something about this, is predictive processing (PP). It can be described as a theory that depicts the main goal of the neurological process to try to predict *what will come next.* More specifically, as a way of efficient computation a PP brain always has a working model of the world to always, so to speak, try to know what will happen around the organism (Clark, 2016; Hohwy, 2013). What is picked up with the senses, is not raw data to construct this representation directly, rather the information picked up is only affecting the model if it differs from the previously held model. Hence, the only thing picked up on is *errors*, and since the surroundings often are fairly stable with no drastic changes, the computational power needed should be comparably less than having to constantly construct a full representation. The brain is depicted to always have a prediction, a generative, working model that is projected from "high" neurological levels onto lower levels until it reaches the levels closest to senses, and further on to the senses and the world around it. If the signals picked up do not match the predictions of the model, there will be errors transmitted upwards and backwards through the levels, or horizontally across levels (Hohwy, 2013). To handle these errors, the model needs to be changed, by associating to other predictions within the neurons, and make better estimations.

A key concept that Clark (2016) introduces is what he calls prediction-weighing. As the name suggests this would be the process of weighing the likelihood of different predictions/possible models against each other. The way to understand how these predictions, expectations and estimations are formed is probably best *through experience*. That is, the more you experience, the more you will get predictions of. And more consistent experience will become stronger predictions (intuitions if you like), and would then likely win over less likely predictions in the so called prediction-weighing. This is described as a sort of "hypothesis testing" mechanism by Hohwy (2013); to find the best possible explanation to what is in front of the organism. Do note that this is not supposed to the judgements of a homunculus of the brain, but rather basic neuronal processes. So what about those higher cognitive activities like reasoning: both Hohwy (2013) and Clark (2016) does not have *very* much to say about it, but the former is optimistic in that it can be understood with the PP framework.

Instead of a view with two functionally distinct neurological systems, like DPT, PP could possibly be able to accommodate similar modes on different levels; of using fast and heuristic based models that is often correct, like system 1, as well as inner mental simulations of models, which would be in similarity to system 2 (Clark 2016, pp. 245-254). In other words, the suggestion is that “lower” processes as perception and action, as well as “higher”, and more abstract conceptualizations such as CT, in broad terms could be understood with the same kind of underlying processes of predictions and errors.

To explore how this could be understood and more specifically applied to CT, it would probably be necessary with a systematic analysis. The initial investigation presented here is limited in its scope, and will only go into slightly more detail into a few aspect of CT, to try to visualize and sketch what PP, and Clark (2016) seems to suggest. To start with I will concentrate on one of the aspects mentioned by Moore (2013); rationality (5), which can be related to scepticism (2).

Rationality, often being associated to philosophy, is often understood as something like: believing in things for certain explicit and specific reasons. It is often also associated to logics, as well as cognitive biases, thinking traps and (thinking/logical) fallacies. For example: *Ad Hominem* argumentation is an attack on the source of an idea, rather than the idea itself; like, criticizing someone’s character, rather than what he is proposing. This fallacy, and others like it, forms certain patterns. Being versed in logical schemes like these is rarely seen as sufficient, to become a critical thinker, but it often *is* viewed as necessary, to be able to pick up on faulty conclusions. CT education in philosophical spheres is often centred around learning such “reasoning tools”; which could be seen as patterns of structured information. Having an arsenal of “reasoning tools”, could be one way to understand how predictions internally can be simulated and producing errors, not necessarily onto the world through the senses, but a kind of offline-thinking, horizontally on high-level, as Clark (2016) suggests.

To spell this out: a person that is exposed to a statement that subsequently is dismissed, by himself or other, could get an alternative prediction of it being a likely pattern of: “Ad Hominem”. This could lead to an initial prediction-weighing process of how likely this seems to be, and the Ad-Hominem model wins, the subject could enter into addition mental simulations of prediction weighting to create an alternative model that supports the stimulus of the original statement. This could be one, initial, interpretation of what proponents of PP could mean. This could also be related to scepticism (2), which can be described as not accepting ideas on face value, but rather to recognize them, and examine them; as being something like the proposed prediction-weighing, and looking for alternative explanations, in terms of alternative models and predictions, with inner simulations of predictive models.

This is just a brief sketch of some aspects, and there seems to be possibilities to also analyse other aspects of CT with PP. Some proposed explorative question to ask, at this stage, is: In general, can critical thinking as an abstract conceptualization convincingly be understood with such basic processes on neural levels as PP? And subsequently, are essential elements of CT lost when doing so? Finally, the inner, mental simulations, seems to get a central role for Clark’s (2016) theory, for how to understand higher levels of thinking such as reasoning, yet it seems to be an open question to how this would go about, and if that really is what is happening. The aim, and hope, with this presentation, will be to try creating fruitful discussions and feedback on this proposal.

## References

Clark, A., 2016. *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford University Press, Oxford ; New York.

Facione, P.A., 1990. C*ritical thinking: A statement of expert consensus for purposes of educational assessment and instruction*, The Delphi Report: Research findings and recommendations prepared for the committee on pre-college philosophy. The California Academic Press, Millbrae, CA.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. *The neural bases of cognitive conflict and control in moral judgment*. Neuron 44, 389–400.

Hohwy, J., 2013. *The Predictive Mind*. OUP Oxford.

Lipman, M., 2003. *Thinking in education*, 2nd ed. ed. Cambridge University Press, New York.

Moore, T., 2013. *Critical thinking: seven definitions in search of a concept*. Stud. High. Educ. 38, 506–522.

# Cognitive Challenges in eSports

**Jana Rambusch**[1], Anna-Sofia Alklind Taylor[1], and Tarja Susi[1]

[1]*School of Informatics, University of Skövde*

*jana.rambusch@his.se*

eSports is a rapidly growing phenomenon characterized by its own culture and community, comprising players and spectators, as well as other stakeholders such as game development companies, media, and investors. eSports commonly refers to *competitive video gameplay*, and "the input of players and teams as well as the output of the eSports system are mediated by human-computer interfaces" (Hamari & Sjöblom, 2017, p. 211). In this research we have mainly focused on spectator perspectives in eSports. However, the understanding of spectatorship involves taking into account the interrelationship between players and spectators as well. Playing and spectating competitive computer games entail specific cognitive skill sets to cope with the challenges that each activity poses (Spence & Feng, 2010). For instance, most compeive games require quick reflexes and outstanding visuomotor coordination together with strategic thinking and, especially when playing in teams, communication skills. A spectator needs to be able to recognize these skills in order to identify and appreciate player expertise and exceptional moments. Furthermore, most competitive games are considered difficult to follow for novices, especially if they do not have first-hand experiences with the game being watched, or are unfamiliar with the game genre (Cheung & Huang, 2011). This poses a number of cognitive challenges which can undermine high-quality spectating experiences.

A pre-study was carried out in cooperation with two Swedish game development companies and a Swedish adult educational association. The game used in the pre-study was an alpha version of BATTLERITE, a multiplayer online battle arena (MOBA) game developed by Stunlock Studios; it is a fast and action-packed game, with focus on competitive player vs player (PvP) combat in short, intense matches (see Figure 1). By using an early version of the game, we had the opportunity to study spectators with limited knowledge about this particular game, making the results not only useful for further research on the topic of spectatorship in eSports, but also for the developers to improve the game from a spectator perspective. A drawback for the participants however, with their limited knowledge about the game, was that they did not always understand every detail of the game. Yet, they were quick to pick up the gist of it and since we focused on unexperienced spectators, advantages outnumbered disadvantages.

We chose a *qualitative explorative approach*, with the aim to identify factors that contribute to or prevent high-quality spectator experiences. Here, we focus on the cognitive aspects of these factors. The methods chosen for data collection were observations and focus-group interviews. The setting for the study was a home-like environment provided by one of the game development companies, where groups of spectators watched a pre-recorded tournament (ca. 15 minutes) with three matches played 2v2 in BATTLERITE, shown on a large TV screen. The participants (1 woman, 27 men, aged 19-28) were divided into four groups of spectators, with 6-8 people in each group. The experience of playing and/or watching eSports varied from none to playing and watching every week. The collected data consist of video recordings (60 minutes), recorded interviews (130 minutes), and field notes. The analysis revealed first and foremost challenges related to attention and visual perception, social cognition, emotions, and learning.

**Attention and visual perception:** Although the arena in BATTLERITE is small and mostly fits into one screen, the participants sometimes had trouble following the battle. The interviews revealed that this is a prevalent issue in most eSports. Spectators want to be able to quickly switch between following a specific player or team (detailed level) and being able to see all game events at once (overview). Lack of overview was one of the most common complaints among the participants, both in terms of BATTLERITE and in general. Overview entails not only being able to view the whole map or arena, but also having access to statistics that inform the spectators on how the players are doing and how the game is progressing. This is especially useful for experienced spectators, but with careful design it can also facilitate understanding for less experienced viewers.

The study also found that being able to move the in-game camera is useful when switching between different levels of granularity. Most commonly, the spectator has little control over the camera view, since viewing is restricted to the players' point of view. However, some game developers are adding a 'spectator mode' to their games. In that way an independent camera can be controlled by the spectator or caster, as done, for instance, by Stunlock Studios.

**Figure 1:** BATTLERITE *is a fast and action-packed MOBA game with focus on competitive PvP combat.* *© 2016 Stunlock Studios – reprinted with permission.*

This has great potential for improving the spectator experience, but there is also a risk that critical game events are missed due to an ill-considered placement of the camera. Thus, providing overview and statistics is a critical feature to help spectators focus their attention to important game events.

Less experienced participants also found that the fast pace made the game difficult to follow. There seems to be a breaking point at which a game contains enough action to make the game continuously interesting and exciting, but still being simple and slow-paced enough for a spectator to be able to attend to all important stimuli at once (cf., Spence & Feng, 2010).

**Social cognition:** During the interviews, participants mentioned the importance of watching matches together with others. For instance, many were drawn into watching eSports through their friends. Watching together provides a helping hand in understanding gameplay activities, as some spectators usually are more experienced than others. Furthermore, the social activity of spectating also creates social practices that everyone needs to adhere to in order to become an accepted member of the group. It is, for instance, okay to talk during a match as long as nothing important happens in the game. Also, when spectating by oneself, there is an important social aspect; commentators provide valuable information about and insights into game events that would otherwise go unnoticed. The game in our study did not have a commentator, but importance of good commentators was discussed during the focus group interviews.

**Emotions:** It becomes increasingly important to be able to highlight game events and objects to make the game 'viewable', while at the same time creating suspense by keeping certain things hidden from view. A concrete example from BATTLERITE is the special ability of invisibility, that makes a character invisible to the other players for a short period of time. Unfortunately, the character is also invisible to the spectators, who miss an important opportunity of suspenseful interaction, like a movie viewer who can follow a murder scene from both the murderer's and victim's point of view. Thus, exposing hidden objects to the spectator can add an extra layer to the entertainment value of the spectator experience (cf., Reeves, Benford, O'Malley, & Fraser, 2005).

Another important challenge for spectators is to connect emotionally with players and their affective reactions, especially when concealed from view. Some of the participants talked about their interest in being able to see those affective reactions since it is relevant for their own spectating experience.

**Learning:** To cater to less experienced spectators, the participants requested tutorials for new viewers, as well as allowing pauses for replays and comments. Highlighting items or actions could also be helpful (Hoobler, Humphreys, & Agrawala, 2004). Some of our participants also pointed out that they wanted more narrative around the matches, such as knowing the players' or characters' back-story or the stakes of the game. This could, for instance, be added as an optional part of the game, or be provided by commentators.

Our results show that spectatorship is a complex issue that goes beyond the mere watching of a game, involving many cognitive abilities. This gives rise to a number of design challenges for game developers and broadcasters. For instance, most spectators are players themselves since watching and understanding eSports competitions require thorough knowledge and experience of the game being played. One design challenge is therefore to make eSports accessible and attractive to more diverse or new audiences.

As the results are based on a small-scale pre-study, further research is needed in order to gain a deeper and more detailed understanding of cognitive aspects of spectatorship in eSports.

## Acknowledgements

We would like to thank everyone participating in this pre-study as well as Stunlock Studios for all the practical help. This research was funded by Västra Götalands-regionen (VGR), Sweden, and the University of Skövde, Sweden.

## References

Cheung, G. & Huang, J. (2011). Starcraft from the stands: understanding the game spectator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 763–772). Vancouver, Canada: ACM. doi:10.1145/1978942.1979053

Hamari, J. & Sjöblom, M. (2017). What is eSports and why do people watch it? *Internet Research*, *27*(2), 211–232. doi:10.1108/IntR-04-2016-0085

Hoobler, N., Humphreys, G., & Agrawala, M. (2004). Visualizing competitive behaviors in multi-user virtual environments. In *Proceedings of conference on Visualization '04* (pp. 163–170). Austin, TX: IEEE Computer Society. doi:10.1109/VIS.2004.120

Reeves, S., Benford, S., O'Malley, C., & Fraser, M. (2005). Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 741–750). Portland, OR: ACM. doi:10.1145/1054972.1055074

Spence, I. & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92–104. doi:10.1037/a0019491

Stunlock Studios. (2016). Battlerite. Stunlock Studios AB, Sweden.

# The Importance of Natural Hand Interaction in Virtual Reality: Will Memorization Ability Increase with Higher Sense of Ownership in VR?

**Julia Rosén**[1], Kai Hübner[2], Christian Balkenius[1]

[1]*Lund University Cognitive Science*

[2]*Gleechi AB*

*Kog15jro@student.lu.se*

It is easy to underestimate the role our hands play in our lives. After all, they do exactly what they are meant to do while taking up little cognitive capacity. A hand that is working correctly is a hand that will not require attention. Indeed, the hand is our most important tool in many regards and fundamental in helping the brain think through various ways, such as writing down a shopping list or counting with your hands (Kirsh, 1995). As the emerging field of virtual reality (VR) is getting more popular in entertainment, education, industry and research, the need for naturally acting synthetic bodies that can interact with the environment is being explored.

The benefits of having natural hand interaction in VR are numerous but one goal is to create a sense of ownership in the virtual body. Sense of ownership can be defined as the special perceptual status and the bodily sensations one experiences over one's body and the overall feeling of a body part belonging to a given person (Tsakiris & Carpenter, 2010). This is especially important in VR since the idea is that users will be able to learn more. This could prove useful in training applications, such as surgical training or mine personnel practice (Gallagher et al., 2005; Van Wyk & De Villiers 2009). Since almost all industries are run and operated by humans who apply their hands for actions such as tool- or machine use, hands in VR is an interesting research topic to investigate. It is not only for educational purposes one would want natural hand interaction. It would also be of interest in entertainment to decrease annoyance for players when the hand interaction does not work in an intuitive way. The standard for hand interaction in VR today is subpar, as VR has not yet been adapted to become its own medium properly yet (Slater and Sanchez-Vives, 2016). They only act in very specific pre-programmed ways, or if they are able to interact with the environment, less natural with very few grasp options.

In this study, we test whether a hand interaction in VR that acts more naturally will cause more sense of ownership than a less advanced hand interaction, comparing a natural hand interaction (*VirtualGrasp*) with a less natural hand interaction (*StandardGrasp*). The study explores sense of ownership and cognitive load for these two types. If participants have less sense of ownership over their hands, will they be more distracted in the memorization of a task? Will a hand interaction that acts in a less natural way distract or frustrate the user from remembering a certain memory task? And, will a more natural hand interaction cause a stronger sense of ownership and consequently, increase the level of presence? If true, a more natural hand interaction would be preferable in VR since this will help the overall experience, including the learning process.

## Experiment

After a brief introduction, participants ($N$=20) used the HTC Vive headset and controllers (http://www.vive.com) to see and interact with a virtual reality environment. The scene shows a room with two tables on opposite sides, one of them with a block pattern, the other with a tower of blocks (Figure 1). Each block could give two points: correct position and correct color. Patterns had 54 possible errors and towers had 40

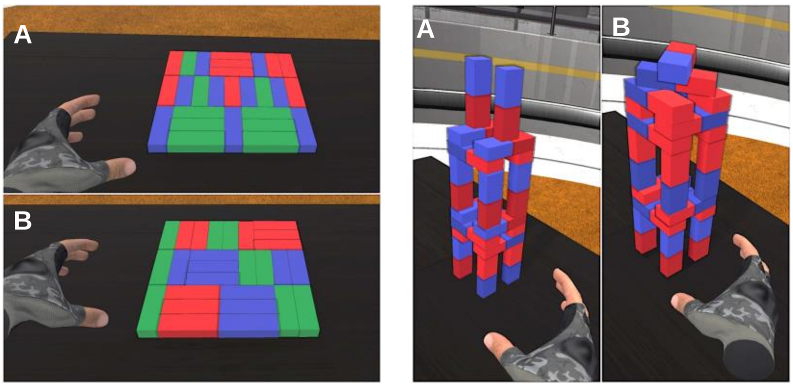



***Figure 1.*** *Pattern and tower tasks for VirtualGrasp (A) and StandardGrasp (B).*

possible errors. Participants were asked to look at the pre-existing pattern and to recreate the pattern from piles of blocks next to it. They could choose to place the blocks anywhere on the table. After recreating the pattern once, the scene was restarted, and participants were asked to recreate the pattern once more. On the second recreation, the amount of time it took to do so was recorded.

Participants recreated each task in VR twice to remove any learning process that could occur while learning how to maneuver in VR that might affect the results. After completing the pattern the second time they removed their VR gear and sat down by a table with the colored blocks in the same color as in VR, and, placed in piles in the same way as in VR. The participants were then asked to recreate the pattern with jenga blocks in real life from memory without any hints on structure or placement. This was done once and timed. When they were done, a picture was taken of the pattern (to calculate the amount of errors made), restacking them into piles again, and then the participants were asked to step into the VR environment once more but this time facing the table with the tower. Thus, the memory tasks the participants were asked to do were recreate the pattern and tower in real life after practicing in VR. After doing the pattern and the tower, the participant was asked to answer a questionnaire about the hand interaction and their presence in VR. From this, they moved on to the other hand grasp interaction. This task was similar to the first hand interaction, but with a new pattern and a new tower (Figure 1).

All participants did both hand interaction types, *VirtualGrasp* and *StandardGrasp*. With *VirtualGrasp*, there are 24 possible grasps on each block, while *StandardGrasp* only grasp an object in one way, creating a more restricted interaction. *StandardGrasp* reflects the kind of hand interaction that is typical of VR today. One half of the participants started with *VirtualGrasp* hand interaction, and the other half started with the *StandardGrasp* hand interaction. All participants remained standing throughout the VR part of the test, and sat down for recreating the pattern and tower as well as when answering the questionnaires.

The presence questionnaires were given after the first task and then again after the second task. There were a total of 20 questions from the Presence Questionnaire, created by Witmer and Singer (1998). The questionnaire regards the presence of the participant using VR and thus deemed appropriate for this study (example: "How natural did your interaction with the environment seem?"). Five of the original questions were modified to focus on the hand interaction used here since there were no appropriate questions for this type of VR interaction (example: "How smoothly did the hands interact with you and the environment"). Participants were asked to answer questions regarding presence on a 1 ("Not at all") to 7 ("Completely") scale score. The questionnaire has an internal consistency measure of reliability of 0.81 (Witmer & Singer, 1998).

## Results

In this study, the aim was to investigate whether more natural hand interaction in VR would increase the sense of ownership over the hands, require less cognitive capacity, and thus score higher on memory tasks. A two-way ANOVA was carried out. Variables under investigation were the type of hand interaction (*VirtualGrasp* and *StandardGrasp*), the type of task (pattern and tower), the time spent on the task, and number of errors on each task. A simple t-test was used to measure sense of ownership and how natural the interaction with the hands seemed.

The results also show that there was a statistically significant interaction between sense of ownership and type of hand interaction. Furthermore, the results show that there were statistically significant relationships ($p<0.05$) between amount of errors and type of hand interaction, and errors and task, as well as errors between interaction of type of hand interaction and type of task. The amount of errors for the patterns was 17.75 of 54 possible errors (33%) for *VirtualGrasp* and 27.7 of 54 possible errors (51%) for *StandardGrasp*. There were similar errors on each hand interaction type on the tower: 12.45 out of 40 possible errors (31%) for *VirtualGrasp* and 10.65 out of 40 possible errors (27%) for *StandardGrasp* (Figure 2A). The results also show that there was a statistically significant relationship between sense of ownership and type of hand interaction. Participants scored an average of 5.5 for *VirtualGrasp* and an average of 3.75 for the *StandardGrasp* on a 7-point scale when asked about their sense of ownership (Figure 2B).

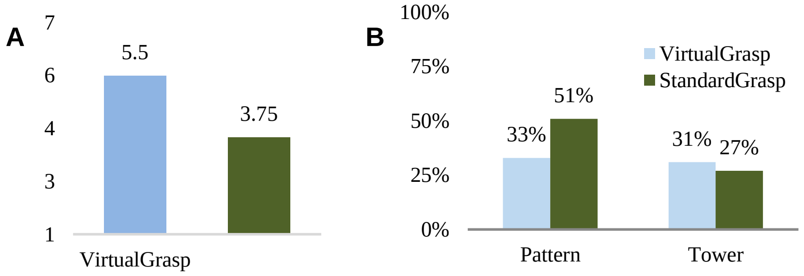



***Figure 2. A.*** *Errors for each task.* ***B.*** *Did you feel a sense of ownership over the hands?*

Moreover, the overall experience and presence in virtual reality was enhanced with *VirtualGrasp* compared to *StandardGrasp*. On questions such as "How natural did your interaction with the environment seem?" participants scored on a 7-point scale 5.55 for *VirtualGrasp* and 3.9 for StandardGrasp. When asked how well the participant could focus on the assigned task, *VirtualGrasp* got a score of 5.7 and *StandardGrasp* got a score of 4.5. Moreover, when asked whether they learned new techniques to improve their performance, *VirtualGrasp* got a score of 6.1 and *StandardGrasp* got a score of 5.3.

## Discussion

The results show that the amount of errors was higher for the less natural hand interaction, *StandardGrasp*. One explanation for this is that the participants, while using *StandardGrasp*, were too distracted by making the hands move in a way that would work, and (as explained by Konstantinou et al., 2014) strains the cognitive capacity. Other than staying focused on making the hands move correctly, it was observed during the tests that participants also showed frustration when *StandardGrasp* would not cooperate, which also fills up the cognitive capacity. We can thus see that it is easier to memorize tasks when a natural hand interaction is used since less time will be spent on trying to control the hands, which would, for example, be very useful in training applications such as the surgical training presented by Gallagher et al. (2005) or other hand intensive tasks.

In conclusion, our results demonstrate that using a more natural hand interaction with several grasp options in VR is not only preferable to users, but also better with regards to memorization abilities. These results might be due to the higher sense of ownership and presence one experiences with more natural hand interaction which ultimately lead to less distraction and thus less constraints for the limited cognitive capacity. That the more natural hand interaction caused higher sense of ownership goes along well with previous studies on this topic (Botvinick & Cohen 1998; Tsakiris & Carpenter 2010; Yuan and Steed 2010; Kilteni et al. 2012; Martini et al. 2015). We can also see the benefits of using a natural hand interaction when designing training applications where the usages of hands are an integral part of the process, like surgical training (Gallagher et al., 2005). Moreover, since there is strong evidence that we think with our hands, it is all the more important to be aware of this when creating hand-oriented applications in VR. In other words, based on our study, natural hand interactions in VR do matter and can affect outcomes in VR.

## References

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature, 391*(6669), 756.

Gallagher A.G., Ritter E.M., Champion H., Higgins G., Fried, M.P., Moses G., … Satava, R.M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery*, *241*(2), 364- 72.

Gleechi. (n.d.). Retrieved from https://www.gleechi.com.

HTC Vive. (n.d.). Retrieved from https://www.vive.com.

Kilteni, K., Normand, J.-M., Sanchez-Vives, M. V., & Slater, M. (2012). Extending body space in immersive virtual reality: a very long arm illusion. *PLoS ONE*, *7*(7), e40867

Kirsh, D. (1995). Complementary strategies: why we use our hands when we think. *Seventeenth Annual Conference of the Cognitive Science Society*, 161-175.

Martini, M., Kilteni, K., Maselli, A., & Sanchez-Vives, M. V. (2015). The body fades away: investigating the effects of transparency of an embodied virtual body on pain threshold and body ownership. *Scientific Reports*, *5*(13948), 1-8.

Slater, M., & Sanchez-Vives, M. (2016). Enhancing our lives with immersive virtual reality. Frontiers in Robotics and AI, 3(12), 74.

Tsakiris, M., & Carpenter, L. (2010). Hands only illusion: multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects. *Experimental Brain Research*, *204*(3), 343-352.

Van Wyk, E., & De Villiers, R. (2009). Virtual reality training applications for the mining industry. *Proceedings of the 6th International Conference on Computer Graphics Virtual Reality Visualisation and Interaction in Africa AFRIGRAPH* 09, *1*(212), 53-63.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225-240.

Yuan, Y., & Steed, A. (2010). Is the rubber hand illusion induced by immersive virtual reality? *IEEE Virtual Reality Conference 2010*, 95-102.

# Haptic communicative functions and their effects on communication in collaborative multimodal virtual environments

Jonas Moll[1], **Eva-Lotta Sallnäs Pysander**[2]

[1]*Department of Information Technology, Uppsala University*

[2]*Department of Media Technology and Interaction Design, KTH Royal Institute of Technology*

*evalotta@csc.kth.se*

Technology mediated haptic communication has been studied for decades. In the beginning, special hardware devices, like e.g. the HandJive (Fogg et al., 1998) and the Shaker (Strong & Gaver, 1996), were designed for very specific tasks and narrow purposes. In the beginning of the 21st century this situation changed when collaborative functions for haptic communication started being used in multimodal virtual environments for e.g. joint object manipulation (Sallnäs et al., 2003; Kjölberg and Sallnäs, 2002; Sallnäs et al., 2000) and as a complement to a shared text editor designed by Oakley et al. (2001). The haptic devices used in all these studies were the nowadays widely known Phantom devices. The use of these functions in virtual environments enabled the study of when and why users choose to use haptic functions to communicate. In the case of the shared text editor the functions could e.g. be used to move to the other user's position or drag the other user to one's own position in the shared interface, which was much larger than the screen size.

Two specific functions for haptic communication we have been studying in our research on collaboration in multimodal virtual environments (see e.g. Moll (2013), Sallnäs and Zhai (2003) and Sallnäs et al. (2000)), are a function for holding on to the same virtual object and a function for holding on to each other's avatars. The first function is realized through virtual rubber-bands which are created between a virtual object and the respective avatars when the users grasps and lift an object by pushing a button on the Phantom device handles. As long as both users are holding on to the object, they feel each other's pushing and pulling forces through the rubber bands. The second function is realized by a magnetic force between the two avatars, created when the Phantom button is pushed at the same time as the avatars are in close proximity. This function can be compared to virtually holding hands.

Both these functions have been shown to affect collaboration and most of all the communication between a sighted and a visually impaired user, during collaborative problem solving, in a number of different ways (Moll and Sallnäs, 2013; Moll et al., 2012). For collaborative problem solving, between visually impaired and sighted users, that involves moving objects to work efficiently and in a way that involves both users a common ground about the shared work space and the task solving process is vital. We have showed that haptic communicative functions can indeed be used to develop the necessary common ground and include both the visually impaired and the sighted user in all parts of the task solving process. The most important effect is that these functions make it possible for one user, often the sighted one, to guide the other to different places in the virtual environment. This guiding action has been shown to substitute for verbal communication in that complex verbal directions are not needed to communicate about navigation - all relevant information is provided through the haptic channel, making navigation a lot more efficient. For example, while using communicative haptic functions to navigate and move objects together, deictic references like "here" and "this one" were commonly used and almost completely replaced the need for series of navigation directions like "up, up, no too much, go down again…". We have also shown that this, in turn, makes the dialogue more goal focused, since all participating pairs focused their discussions on objects and task solving instead of navigation.

We argue that for effective collaboration and communication to take place in virtual environments by means of haptic feedback the haptic functions need to be designed as to allow for reciprocal exchange of information. That is, both users need continuous feedback from each other during e.g. a guiding process or joint object handling. This can be seen in contrast to e.g. the "forced" haptic functions used by Oakley et al. (2001), mentioned above, by which one user is dragged to a location without being able to communicate anything haptically, to the other user or in any means affect the movement.

## References

Fogg, B.J., Cutler, L.D., Arnold, P., and Eisbach, C. (1998). HandJive: a device for interpersonal haptic entertainment. *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'98)* (Los Angeles, California, April 1998). pp. 57-64.

Kjölberg, J., and Sallnäs, E-L. (2002). Supporting Object Handling and Hand Over Tasks in Haptic Collaborative Virtual Environments. *Proceedings of Eurohaptics 2002* (Edinburgh, UK, July 2002).

Oakley, I., Brewster, S., and Gray, P. (2001). Can you feel the force? An investigation of haptic collaboration in shared editors. *Proceedings of Eurohaptics 2001* (Birmingham, UK, July 2001). pp. 54-59.

Strong, R., and Gaver, W.W. (1996). Feather, scent, and shaker: supporting simple intimacy. Proceedings of CSCW'96 (Boston, Massachusetts, November 1996). pp. 29-30.

Moll, J. (2013). *The influence of Modality Combinations on Communication in Collaborative Virtual Environments*. Doctoral thesis, School of Computer Science and Communication, Royal Institute of Technology, Stockholm.

Moll, J., Huang, Y., and Sallnäs, E-L. (2010). Audio makes a difference in haptic collaborative virtual environments. *Interacting with Computers (22; 6)*. pp. 544-555.

Moll, J., and Sallnäs, E-L. (2013). A haptic tool for group work on geometrical concepts engaging blind and sighted pupils. *ACM Transactions on accessible computing (4; 4).* Article 14.

Sallnäs, E-L., Rassmus-Gröhn, K., and Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *Transactions on Computer-Human Interaction (7; 4)*. pp. 461-476.

Sallnäs, E-L., and Zhai, S. (2003). Collaboration meets Fitts' law: passing virtual objects with and without haptic force feedback. *Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction (Interact 2003)* (Zurich, Switzerland, September 2003). pp. 97-104.

# Improving internal models of performance motivates information seeking actions

**Gustaf Å. Wadenholt**[1], Linus Holm[1], Paul Schrater[2]

[1] *Umeå universitet*

[2] *University of Minnesota*

*gustaf.wadenholt@umu.se*

Humans, and other animals, often explore their environment even when there is no expectation for immediate rewards. Such behavior has been shown to improve learning and activate reward structures in the brain (Kang et al., 2009; Gruber, Gelman & Ranganath, 2014). Therefore it is important to understand how and why such information seeking behavior occurs. Present theory of intrinsically motivated information seeking maintains that the expectance of information gain motivates exploration (Loewenstein, 1994; Gottlieb, Oudeyer, Lopes & Baranes, 2013; Kidd & Hayden, 2015). However, few quantitative attempts have tested this claim. Furthermore, the claim appears simplistic; information is everywhere but everything does not instill the motivation for seeking it out. Rather, the motivation for exploration may reside in an aim of improving predictive models. Exploration then may be expected to produce learning on several knowledge variables at once. To begin to test this notion, we tested exploration in an asteroid avoidance computer game. Participants (n = 43) could monitor their performance on each trial by performing an information seeking action, subject to a time penalty. This trade-off between time and information gain is a novel method for quantifying the incentive value of information. We found support for a relationship between expected information gain and the frequency of information seeking actions, at the individual level. Moreover information seeking actions were related to performance improvements in the beginning of the test. Finally, a model-free method implemented as a moving average on performance outcomes (eq. 1) could not account for the information seeking behavior of our participants.

## Method

43 participants completed 800 trials of a two alternative forced choice task with the objective of making a steering decision in the opposite direction of an incoming asteroid. Ten levels of difficulty was created by changing the angle of the asteroid trajectory, with trajectories closer to the center being more difficult. Following the steering decision, participants could choose to wait one second in order to receive feedback on their steering decision (see *Figure 1*). This time penalty increased total trial time by 50% (from two to three seconds).

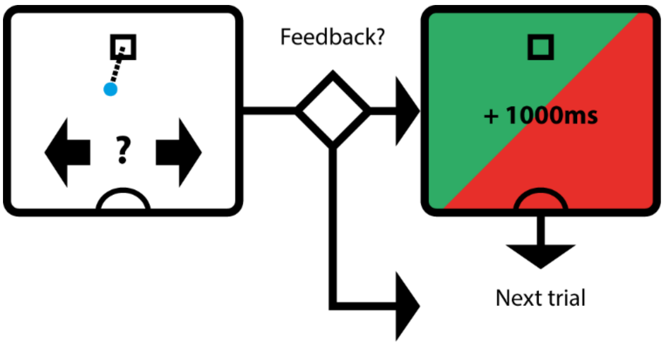



*Figure 1*. Trial example. An asteroid (blue dot) appears at the top of the screen and travels a short distance towards the cockpit (semicircle) at the bottom of the screen with a trajectory going slightly to the left or right side of the screen. The goal is to make a steering decision in the opposite direction of the asteroid's trajectory. Following this steering decision, feedback is available at a cost of waiting for one second. If feedback is not requested the next trial starts immediately.

## Results

Linear regression analysis showed a significant ($p < .05$, $r^2 = .99$) relationship between information seeking and expected information gain at the group level, and for 22 out of 43 participants when analyzed individually.

To find any effects of learning we divided trials into bins of 100 and computed ANOVAs on mean performance and information seeking. A change in performance was detected, $F(7) = 4.85$, $p < .001$, but not in information seeking $F(7) = 2.25$, $p = .09$ (both Greenhouse Geisser corrected).

Finally, we assessed the possibility of participants using a model-free exploration method, such as simply tracking recent performance. This was implemented as a moving average on performance (eq. 1) and regressed on information seeking actions.

Eq. 1

$$MA = \sum_{i=n}^{n-w} C_i e \frac{-i}{\tau}$$

Where $w$ is the integration window and was set to 4*τ. τ is the decay constant and was set to 10. $C$ is the performance outcome in $w$ for instances when feedback was selected, and average performance at that difficulty level when feedback was not selected.

Overall, this model-free method explained a much lower amount of variance (mean $r = .04$) than expected information gain (mean $r = .47$).

## Conclusions

We propose a novel method for quantifying the relationship between information and exploration. Our results suggest that participants used a complex internal model of performance, with several knowledge variables, to guide learning through exploration.

## References

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585-593.

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486-496.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963-973.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449-460.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, *116*(1), 75.

www.swecog.se