Preface

Welcome to SweCog 2018 in Linköping!

This booklet contains the program and short papers for oral and poster presentations at SweCog 2018, this year’s edition of the annual conference of the Swedish Cognitive Science Society. Following the SweCog tradition and its aim to support networking among researchers in cognitive science and related areas, contributions cover a wide spectrum of research.

A trend in recent years, also reflected in this year’s conference program, is an increasing number of contributions that deal with different types of autonomous technologies, such as social robots, virtual agents or automated vehicles, and in particular people’s interaction with such systems. This clearly is a growing research area of high societal relevance, where cognitive science - with its interdisciplinary and human-centered approach - can make significant contributions.

We look forward to two exciting days in Linköping, and we thank the many people who have contributed to the organization of this year’s SweCog conference, in particular of course all authors and reviewers! The organization of SweCog 2018 has been supported by the Faculty of Arts and Sciences, the Department of Culture Communication (IKK), and the Department of Computer Information Science (IDA) at Linköping University, as well as Cambio Healthcare Systems and Visual Sweden.

Tom Ziemke
Mattias Arvola
Nils Dahlbäck
Erik Billing

The reviewers were: Alexander Almér, Anna-Sofia Alklind Taylor, Mattias Arvola, Christian Balkenius, Erik Billing, Nils Dahlbäck, Linus Holm, Erik Lagerstedt, Rob Lowe, Alberto Montebelli, Tarja Susi, Annika Wallin, and Tom Ziemke.
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:00 — 13:00</td>
<td>Registration</td>
</tr>
<tr>
<td>13:00 — 13:15</td>
<td>Welcome</td>
</tr>
</tbody>
</table>
| 13:15 — 14:00 | Invited speaker - Claes von Hofsten (Uppsala)  
Cognitive Science Reflections. |
| 14:00 — 14:30 | Robert Johansson (Linköping) Arbitrarily applicable relational responding as non-axiomatic logical reasoning (7-9) |
| 14:30 — 15:00 | Coffee in Ljusgården                                                |
| 15:00 — 15:30 | Merve Akca (Oslo) Time course of selective attention for instrumental timbres and human voice (p. 10-11) |
| 15:30 — 16:00 | Sara Mahmoud (Skövde) Self-driving cars learn by imagination (p. 12-15) |
| 16:00 — 16:15 | Break                                                                |
| 16:15 — 17:00 | Invited speaker - Anders Arweström Jansson (Uppsala)  
How many people does it take to keep AI employed? — Critical reflections on autonomy and decision making in man and machine. |
| 17:00 — 18:30 | Poster session and coffee in Ljusgården                              |
| 19:00 — | Conference dinner at Universitetsklubben                              |
Programme Friday October 12th
Ada Augusta Lovelace (B-building, Campus Valla, Linköping University)

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>Invited speaker - Ginevra Castellano (Uppsala)</td>
</tr>
<tr>
<td></td>
<td>Building socially engaging interactions with robots.</td>
</tr>
<tr>
<td>09:45</td>
<td>Coffee in Ljusgården</td>
</tr>
<tr>
<td>10:15</td>
<td>Annika Silvervarg (Linköping) What dimensions of virtual agent</td>
</tr>
<tr>
<td></td>
<td>appearance matters? - Exploring preferences of children and adults</td>
</tr>
<tr>
<td></td>
<td>(p. 16-18)</td>
</tr>
<tr>
<td>10:45</td>
<td>Erik Billing (Skövde) Robot-Enhanced Therapy for Children with</td>
</tr>
<tr>
<td></td>
<td>Autism (p. 19-22)</td>
</tr>
<tr>
<td>11:15</td>
<td>Vladislav Maraev (Göteborg) Laughter in interactions with humans</td>
</tr>
<tr>
<td></td>
<td>and machines</td>
</tr>
<tr>
<td>11:45</td>
<td>Visual Sweden</td>
</tr>
<tr>
<td>12:00</td>
<td>Lunch in Ljusgården</td>
</tr>
<tr>
<td>13:00</td>
<td>Sara Stillesjö (Umeå) Dynamic changes in precuneus activity during</td>
</tr>
<tr>
<td></td>
<td>exemplar-based learning: support for a declarative account</td>
</tr>
<tr>
<td>13:30</td>
<td>Andreas Falck (Lund) Contagion of beliefs as an underpinning of</td>
</tr>
<tr>
<td></td>
<td>Theory of Mind (p. 24-26)</td>
</tr>
<tr>
<td>14:00</td>
<td>Break</td>
</tr>
<tr>
<td>14:15</td>
<td>Invited speaker - Pär Nyström (Uppsala)</td>
</tr>
<tr>
<td></td>
<td>Boosting cognitive development using gadgets and games.</td>
</tr>
<tr>
<td>15:00</td>
<td>Conference closing and Coffee in Ljusgården</td>
</tr>
<tr>
<td>15:15</td>
<td>SweCog annual member's meeting</td>
</tr>
</tbody>
</table>
List of poster presentations  
October 11th, Ada Augusta Lovelace (B-building, Campus Valla, Linköping University)

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anton Axelsson</td>
<td>Eye movement behaviour across levels of proficiency in control of simulated train traffic</td>
</tr>
<tr>
<td>Hadi Banaee</td>
<td>Data-driven Conceptual Space as Semantic Representations for Numerical Data</td>
</tr>
<tr>
<td>Erik Schaffernicht</td>
<td></td>
</tr>
<tr>
<td>Amy Loutfi</td>
<td></td>
</tr>
<tr>
<td>Pierre Gander</td>
<td>Memory of fictional information: fictional status and source (p. 27-29)</td>
</tr>
<tr>
<td>Linus Holm</td>
<td>The intrinsic value of solving logical problems</td>
</tr>
<tr>
<td>Vipul Vijayan Nair</td>
<td>Incidental processing of biological motion: effects of orientation, local- motion and global-form features</td>
</tr>
<tr>
<td>Karl Drejing</td>
<td></td>
</tr>
<tr>
<td>Paul Hemeren</td>
<td></td>
</tr>
<tr>
<td>Maximilian Roszko</td>
<td>Investigating Self-Knowledge about Eye Movements using False Feedback in a Forced-Choice Task</td>
</tr>
<tr>
<td>Lars Hall</td>
<td></td>
</tr>
<tr>
<td>Petter Johansson</td>
<td></td>
</tr>
<tr>
<td>Philip Pärnamets</td>
<td></td>
</tr>
<tr>
<td>Sofia Thunberg</td>
<td>RoboCup@Home</td>
</tr>
</tbody>
</table>
Arbitrarily applicable relational responding as non-axiomatic logical reasoning

Robert Johansson1,2, Jonas Ramnerö2, & Arne Jönsson1

1Department of Computer and Information Science, Linköping University, Sweden
2Department of Psychology, Stockholm University, Sweden
robert.johansson@liu.se, jramn@psychology.su.se, arne.jonsson@liu.se

Stimulus equivalence

In the late 1960’s, Murray Sidman was working with language comprehension with severely developmentally disabled individuals. Unexpectedly, he discovered that if subjects were successfully taught to match pictures and printed words to dictated words, and to name pictures, they would without explicit training learn how to match printed words to pictures, match pictures to printed words and to “read” (i.e., name words). From a behavioral psychology point of view, this was very interesting, as it demonstrated a clear example of emitted behavior without a history of reinforcement. This discovery has resulted in over 40 years of research on stimulus equivalence (Sidman, 2009). Stimulus equivalence is a behavioral phenomenon that has only been observed in humans with verbal abilities (with one possible exception, of a California sea lion “Rio”). However, generally no non-human animals seem to have been able to do this (Zettle et al., 2016). The typical way to study stimulus equivalence is with the help of matching-to-sample experiments. In such experiments, participants are exposed to series of arbitrary stimuli (e.g., nonsense symbols) where the task is to match a certain symbol to a given sample stimuli. Such experiment is an example of relational responding. That is, the task for a participant is not to emit a response in relation to a certain stimulus. It is rather to respond to the relation between symbols.

A formal definition of stimulus equivalence follows. Assume three nonsense symbols, which we for simplicity will refer to as A, B and C (they might be nonsense words, pictures, or something else). Within a given experiment (like the matching-to-sample), participants are taught to select B rather than some other option in the presence of a sample A (i.e., the relation A → B will be established). In the same way C is trained as the correct response in the presence of B (B → C). After these relations have been trained, without training in other relations, participants demonstrate an increased probability of selecting A from a set of options when B is presented as a sample (B → A; symmetry), selecting C when A is displayed (A → C; transitivity), selecting A when C is displayed (C → A; equivalence), and also the trivial case of selecting A when A is displayed (A → A; reflexivity).

Demonstrating symmetry and equivalence are examples of derived relational responding, as these relations are not directly taught but instead derived. Prior to the research by Sidman and colleagues the emergence of these derived stimulus relations was not expected in similar experimental setups. The stimulus equivalence phenomenon opened up for a new way of studying symbolic relations (i.e., how a word “represents” an object in language), and supported the idea that derived stimulus relations were an important component in language and cognition. Importantly though, the idea is not new. The abstract concept of sameness or equivalence has long been regarded as “the very keel and backbone of our thinking” (William James, Principles of Psychology, 1890/1998, p. 459).

Arbitrarily applicable relational responding

In the late 1980’s, the developers of Relational Frame Theory (RFT; Hayes et al., 2001) started to ask questions on what was beyond equivalence, for example: What kind of derived relational responding based on other relations than equivalence are human beings capable of? And if so, would such responding also be reflexive, symmetrical, and transitive?

Consider the following statement: “A is more than B and B is more than C”. Not only are the AB and BC relations specified, we immediately derive the BA, CB, AC, and CA relations. Hence, we are able to answer a
question such as “Is C more than A”? (The answer would be “No”.) Consider an elaborated version of this example. Imagine someone standing in a coffee shop looking at the menu with words such as “Espresso”, “Americano” and “Caffé Au Lait”. If the person looking at the menu has no experience of these brands, he/she might ask “How are these related regarding strength?” An answer might be “the Espresso is stronger than Americano, and the Americano is stronger than the Caffé Au Lait”. The person asking will immediately be able to derive the other relations, for example that Caffé Au Lait is less strong than Espresso. Furthermore, let’s say that the person tastes Americano, we could in behavioral terms say that Americano acquires various stimulus functions, such as taste and smell. Importantly though, what will happen is that the stimulus functions throughout the whole Espresso-Americano-Caffé Au Lait-network will be transformed due to the information given. For example, the person might be able to imagine the strength of taste on Espresso, despite no actual experience. Such “derived experience” could affect future decision making for the customer in the coffee shop.

How is the above related to stimulus equivalence? Clearly, both phenomena is about derived stimulus relations. For stimulus equivalence, reflexivity (A is the same as A) was required. This seems not to be the case for more than/less than (A is not more than A). Regarding symmetry (if A=B then B=A), there seem to be modified versions of this for more than/less than relations (If A>B then B<A). Transitivity seems to behave similarly (if A>B and B>C then A>C). Consider another relation such as opposition, and someone learns “A is the opposite to B, and B is the opposite to C”. This relation is not transitive, as someone will be able to derive that A is actually the same as C. In RFT, these “generalized versions” of symmetry and transitivity are labeled mutual entailment and combinatorial entailment, respectively (if A>B>C then the relation B<A is mutually entailed, and C<A is combinatorially entailed).

Several broad classes of relating have been discovered: Coordination (equivalence, sameness), Opposition, Distinction (“is different from”), Comparison (e.g., more/less or bigger/smaller), Hierarchy (contains/member of), Temporal (before/after), Spatial (Here/There), and Deictical (relations in terms of the perspective of the speaker; a combination of interpersonal relations like I/You with spatial and temporal) (Zettle et al., 2016). Expressed in RFT terms, all of these overall patterns of relating have the properties of mutual entailment, combinatorial entailment and transformation of stimulus function. In RFT, these patterns with these properties are referred to as relational frames.

Now consider this scenario: Someone new to a country learns that three never before seen coins A, B, and C are ordered along a comparative dimension such as “A is worth more than B that is worth more than C”. Having learned this enables the person to operate efficiently in several potential decision making scenarios that involves money in the new country. What’s crucial about this example is the fact that the comparative relations between A, B and C are along an arbitrary dimension of worth, rather than for example size, weight or other directly observable properties. This is an example of when a relational frame of comparison is arbitrarily applied. In RFT terms, it is the contextual cue of “is worth more” that controls the application of the frame of comparison. Had someone instead said “A is equal to B”, we could have expected the frame of coordination to be applied. This illustrates the arbitrary nature of relational frames.

In summary, arbitrarily applicable relational responding (AARR) is defined as abstract response patterns, that has the properties of mutual entailment, combinatorial entailment and transformation of stimulus functions, and that are controlled by contextual cues. From an RFT perspective, cognition is not a mental event that mediates between environment and behavior. It is rather a behavioral event (AARR), and hence, it can be studied and understood within a behavioral psychology framework, using experiments such as the matching-to-sample task. Another way to put it: arbitrarily applicable relational responses are what “minds” are full of, and when we speak of “cognitive” phenomena (such as “thinking”, “planning”, “remembering”, “decision making”) we are referring to complex instances of relational framing that are more or less evident under different environmental conditions (Zettle et al., 2016).
NARS and non-axiomatic logic

It is clear that AARR is a domain-independent process, potentially occurring at many levels, and seems to be involved in many “cognitive” functions. One example of a system that aims to reason about such processes is the general-purpose intelligent system NARS (Non-Axiomatic Reasoning System). NARS is designed to be adaptive and to work with insufficient knowledge and resources. Its various cognitive functions are uniformly carried out by a central reasoning-learning process following a “non-axiomatic” logic (Wang, 2013).

NARS makes use of a formal language, “Narsese”, for its knowledge representation, and this language is defined using a formal grammar (Wang, 2013). The system’s logic is developed from a so-called term logic. Statements in this logic have the form subject-copula-predicate. The smallest element that can be used as one of these components is referred to as a term. In Narsese, the most basic statement is the inheritance statement, with the format “S → P”, where S is the subject term, and P is the predicate term. The “→” is the inheritance copula, which is a reflexive and transitive relation. The intuitive meaning of “S → P” is “S is a special case of P” and “P is a general case of S”. For example, the statement “bird → animal” intuitively means “Bird is a type of animal”. Importantly, such a statement doesn’t say anything about the meaning of the terms itself – it merely states the relationship. Terms can be grouped together in various forms of sets, for example {Cat, Dog, Giraffe}, enumerating instances, or [yellow, tall, four_legged], enumerating attributes. Various arbitrary relations can be represented, for example the relation “Silvia is the mother of Victoria” is in Narsese represented as “((Silvia] × {Victoria}) → mother-of”. A statement such as “Tim knows snow is white” can be represented as a higher-order statement “(Tim) → (know / {snow → [white]})”, where the statement “snow → [white]” is used as a term. Beside the inheritance copula (“→”, “is a type of”), Narsese also includes three other basic copulas: similarity (“↔”, “is similar to”), implication (“⇒”, “if-then”), and equivalence (“⇔”, “if-and-only-if”). The last two copulas are “higher order”, meant to be applied to statements themselves.

Furthermore, NARS can reason on events, that are described as statements with temporal attributes. For example, “event E1 happens before E2” is described in NARS as “E1 / E2”. Finally, procedural operations in the system are events realized by the system itself. These operations are typically executable commands or procedures of the system. Formally, an operation is an application of an operator on a list of arguments, written as op(a₁, ..., aₙ) where op is the operator, and a₁, ..., aₙ is a list of arguments. Such an operation is interpreted logically as statement “(× {SELF} {a₁} ... {aₙ}) → op”, where SELF is a special term indicating the system itself, and op is a term that has a procedural interpretation. For example, if we want to describe an event “The system is holding green brick nr 2”, the statement can be expressed as “(× {SELF} {green_brick_2}) hold”.

Goal of this work

The primary aim of this work is to investigate if NARS can do AARR with gradually increasing complexity, and under which conditions this is made possible. During the presentation, we will describe a research plan, starting with stimulus equivalence, and then continuing with more advanced relational responding. How this work potentially could be beneficial for RFT research and for research on intelligent systems will be discussed, as well as potential future applications.

References


Time course of selective attention for instrumental timbres and human voice

Merve Akça

1RITMO Center for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo
merve.akca@imv.uio.no

Selecting goal-relevant information within a brief time period can leave us ‘blind’ or ‘deaf’ to other relevant information while searching for them among the irrelevant ones. Attentional blink (AB; Raymond, Shapiro, & Arnell, 1992) refers to the phenomenon that when two targets are presented in close temporal proximity (between 200 and 500 milliseconds), people often fail to report the second target. In the classical AB paradigm, participants are presented with a stream of stimuli one at a time at a very fast presentation rate and asked to report the two pre-defined target items among the distractors at the end of the stream. Through systematic manipulation of the interval between the two targets, the AB paradigm allows us to measure the dynamics of temporal selective attention.

Recently it has been demonstrated that AB greatly differs between individuals, or groups of individuals, and that these differences in the AB magnitude are dependent on various factors such as stimulus category, duration, and modality (Willems & Martens, 2016). Furthermore, it has been shown that certain stimuli, such as faces (Awh et al., 2004; Landau & Bentin, 2008), which often is referred to as an expertise object for almost all of us, and other objects of expertise (e.g., car for car experts, Blacker & Curby, 2016) can be less susceptible to visual AB effects.

However, to the best of my knowledge, no previous study has explored whether these kinds of expertise effects in the AB magnitude can be observed in the auditory domain. In this presentation, I address temporal dynamics of selective auditory attention for the potential objects of expertise among expert musicians and non-musicians (i.e. cello tones for expert cellists and human voice for both groups, contrasted with organ tones as control). Musical training is known to produce benefits beyond the near transfer of the skill, extending to high level processes (e.g. attention and working memory), and individuals with higher levels of executive working memory functioning and broad attentional focus perform better in the AB paradigm (Willems & Martens, 2016). Therefore, it seems logical to assume that highly experienced musicians may have an advantage in their performances in the AB paradigm in general, and that this advantage might be more pronounced when searching for targets relating to their musical expertise.

Using the auditory attentional blink paradigm, the present study compares AB magnitude among expert cellist and musically naïve participants by varying the category of the second target (human voice, cello, organ) and the interval between the first and second target in the rapid auditory stream. My hypotheses are that 1) the AB magnitude of expert musicians would overall be lower than that of less musically experienced participants 2) both participant groups would show less auditory AB when the second target is human voice, due to our long experience with human voices (as has been shown in the AB studies in the visual modality where the second target was human faces), and that 3) expert cellists would show less auditory AB to the timbres associated with their principle instrument, that is, cello tones, compared to organ tones. The results of this study may have a potential to guide us in our understanding of the selective attention in relation to musical expertise and expertise in general.

Acknowledgement. This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262762.

References


Self-driving cars learn by imagination

Sara Mahmoud1, Henrik Svensson2

1,2: Interaction Lab, School of Informatics, University of Skövde
Sara.mahmoud@his.se

Introduction

This short paper outlines a cognitively inspired “imagination” based approach for interactive machine learning for self-driving cars. The approach is based on learning from experience but with the idea of optimizing the driving from “dreaming” or “imagining” driving on different roads and is part of the general approach developed in the EU Horizon 2020 project Dreams4Cars (Lio et al., 2017). The work presented here builds, partly, on the general idea of cognition as simulation of perception and action (e.g. Hesslow, 2002) and robot models based on that idea (summarized in Svensson, 2013). While the quality of the imagined worlds for cars and humans are quite different, the model presented here takes inspiration from a possible facilitating role of dreams for learning (Svensson, Thill, & Ziemke, 2013). The approach is thought to increase the agent’s performance even by learning from internally generated scenarios not seen by the agent. At the heart of the approach to improving learning in the driving domain is what we call an episodic generator simulation that can generate different driving scenarios to train the agent to drive and learn without manually programming every scenario. During the learning phase in a certain road scenario, the agent learns to behave in these situations but may not perform well in new situations. The episodic generator generates suitable roads that improves the learning performance, for example, by recreating variations of what the agent performed poorly on.

Self-driving cars are interactive agents that need to learn from experience. Reinforcement learning allows the agent to explore the environment, perform actions and learn from these actions. This approach has been used for discrete state environments. However, it has started to flourish in continuous state space as well. Lillicrap et al. (2015) designed a reinforcement model with deep learning that can learn different games with the same configurations and hyper-parameters with continuous control that they patented in 2017. Their model is a deep network of convolution layers followed by filter layers and the output layer. In this work, multiple openAI environments1 were used. OpenAI environments are designed to handle the reinforcement learning needs, such as the state space that the agent observe in each step, the accepted actions that agent can take and the rewards obtained from each action in each state (Brockman et al., 2016). This work shows promising opportunities for deep reinforcement learning that can learn to play video games (Singh, Okun, & Jackson, 2017), however this work is still restricted by the game world image pixels as input. If we want to deploy dreaming learning to the real world, this representation of game image pixel inputs doesn’t fit for real world deployments because of the major differences of the agent’s view of the world. In order to deploy such models for real world agents, the agent needs to get meaningful sensory inputs that can then to some extent match the real world, such as distance readings from objects.

However, imaginations in an off-line world may be useful. For example, Mahadevan (2018) argues that what he calls “imagination science” is the future of AI. Imagination science is according to him different from data science. Data science is based on statistical calculations for finding correlations in the given data, such that machine learning of image and object recognition doesn’t go much beyond labeling items, while imagination science strives to construct machines that not only learns to categorize and label the world but also create novel worlds. In imagination science, the system is able to reason and find causal explanations.

Imagination is important in several human tasks and a possible advantage of such an ability is to come up with a new usage of items. For example, a bed is an item for lying down in the adults’ perspective, while for children it is an item for jumping up and down. Additionally, imagination is a way of problem creation. While machine

1 https://openai.com/
learning has shown promising results in playing Atari games (Silver et al., 2016), it still can’t create a new game (Mahadevan, 2018).

The notion of creativity in Artificial Intelligence is, however not new (Boden, 1998), but it has recently gained interest, perhaps because of the improved capabilities of deep networks. For example (Ha & Schmidhuber, 2018) used Generative Adversarial Networks (GAN) that learns to generate new unseen data from a given data. In particular, they trained a car in openAI car racing environment as well as training a network to generate similar scenarios constituting a kind of off-line “dream world”. The trained car was then trained in a generated situation by the GAN model as a dreaming world. The works showed that the agent performed higher when trained in the generated unseen environment and then injected back to the real environment. However, this technique lacks the availability of physical dimension of the environment. Generating simulated world using similar images doesn’t imply embedding the physics of this world in it. In order to have more effective learning by dreaming, our approach uses a dynamic physical simulation. This simulation should have both abilities of generating new situations as well as the physics simulation. Related work done by the OpenAI team has shown that a robot arm can learn in simulation and then highly perform in the real world (Duan et al., 2017). Duan et al. trained a robot arm in simulation using learning by demonstration to build a block tower. The robot arm picks blocks from different locations and then stacks them horizontally to form a tower. The trained agent was then tested with real blocks and proved to perform similarly to the simulation. Our approach to learning by dreaming is outlined in the following paragraphs. Sutton (1991) proposed the Dyna architecture which includes both reinforcement learning and learning from previous experience. In Dyna architecture, the agent memorizes the reinforcement learning components of state, action, reward and next state in memory then repeat this memory in supervised learning. This approach was introduced in Deep Q Learning as a replay mechanism (Mnih et al., 2013). The replay is a buffer of previous experience that the agent re-practice while it learns new experience in reinforcement learning. Although the Dyna architecture or the replay buffer mechanism can be considered as imagination learning, it restricts the imagination by the previous states and actions that the agent has tried. The imagination technique should also allow the agent to explore unexperienced states. In order to have more effective learning by dreaming, our approach uses a dynamic physical simulation. This simulation should have both abilities of generating new situations as well as the physics simulation. Related work done by the OpenAI team has shown that a robot arm can learn in simulation and then highly perform in the real world (Duan et al., 2017). Duan et al. trained a robot arm in simulation using learning by demonstration to build a block tower. The robot arm picks blocks from different locations and then stacks them horizontally to form a tower. The trained agent was then tested with real blocks and proved to perform similarly to the simulation. Our approach to learning by dreaming is outlined in the following paragraphs.

The system consists of the following parts (as also shown in Figure 1):

- **OpenDS**: a dynamic game tool engine that simulates driving in a road taking into account the physics relying on this. The unique function of OpenDS (opends.dfki.de) is that it is a dynamic simulation which means that it creates a simulated road from a road description without explicitly program the details of the road. The created road is the environment that the agent interacts with to know the current state and send the taken action. The environment sends the current state as a scenario message 20 messages per second. When the environment gets the maneuver message it calculates the new car position and renders the car to the new state. As such the car simulator is a kind of container for the episodic simulation.

- **Co-driver agent**: a learning agent that learns in an interactive way through deep reinforcement learning. The co-driving receives the current state as a reinforcement learning state. Then the agent chooses the best action according to a policy and sends the selected action. Due to the high dimensionality of having continuous actions, the range of actions were discretized into four actions: right, left, accelerate and break. According to the action taken, the co-driver receives a reward that represents how good the action is in the given state. The reward could either be a punishment for a wrong move or a positive reward for fulfilling the needs. For this experiment the main task is lane keeping. For the sake of simplicity, a tabular method is used for the reinforcement learning. In this method, the space of states is discretized into all the possible states with all the possible actions and assign expected reward for each.
- **Middleware connector**: a layer responsible for connecting the learning agent to the car simulator OpenDS. It receives the scenario message from OpenDS and converts it to reinforcement learning states and passes it to the co-driver agent. It then receives the selected action and creates a maneuver message which it passes it to the simulation. This layer is also responsible for determining the rewards that the agent takes for each step. The reward is calculated based on the lane keeping task. In addition, all the data is packed and passed to the episodic generator to keep track of the learning performance.

- **Episodic generator**: this layer is responsible for creating road descriptions as block components allowing OpenDS to render it into a physical simulation. The generator gets the performance data from the middleware connector, analyses it and finds the weak points where the agent struggled to learn. Accordingly, it generates new roads that focuses on the agent’s weaknesses.

![Episodic generator system architecture for self-driving car](image)

From a process prospective the setup works as follows: First, the episodic generator creates road descriptions of 100 roads. These initial rewards are generated randomly from the available possible building blocks of the simulator to explore the possible roads. Next, OpenDS renders these descriptions into the physical road simulation and runs the first road. The middleware connector is responsible for connecting the simulation to the agent. At the beginning, the agent starts with no knowledge about driving, which means that all the initial space state pairs are initialized and the initial rewards are set to values higher than zero to stimulate the agent to explore new states. The training process then starts with the connection between the simulator OpenDS and the agent through the middleware connector. OpenDS sends the scenario message. The agent receives reinforcement learning states from the middleware connector after being manipulated from the scenario message. The agent matches this state with the states in the state-action pair table. The agent then chooses the action that is expected to receive the highest future reward as:

$$Q^*(s, a) = \max_{\pi} Q^\pi(s, a)$$

When the agent receives the reward along with the new state, the agent updates the state-action pair table by:

$$Q_{\text{new}}(s_t, a_t) \leftarrow (1 - \alpha). Q(s_t, a_t) + \alpha. (r_t + \gamma. \max_{a} Q(s_{t+1}, a))$$

Where $s_t$ is the current state, $a_t$ is the action taken in the current state, $\alpha$ is the learning rate, $r_t$ is the reward, $\gamma$ is the discount factor.

When the agent drives out of the road, the agent gets a high penalty (negative reward) and starts the road over. The simulation runs till the agent finishes all the roads successfully. During training, the log data including the road description, the driving performance, number of time the agent losses in a road, the accumulated reward and the distance driven are accumulated through roads. The episode generator analyses this data and uses it to generate new roads. The episode generator finds the roads that the agent struggled to pass (took more trials to finish) and generate similar features of this road. The new roads consist of the road blocks that the agent performed less well in. For example, if the agent poorly passed the sharp curves, the episode generator creates road descriptions that includes sharp roads.

**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**


What dimensions of virtual agent appearance matters? – Exploring preferences of children and adults

Annika Silvervarg¹, Veronica Hägg

¹Dept. of computer and information science, Linköping University, 58331 Linköping, Sweden

How do we design the appearance of a virtual agent to be suitable for all contexts and all type of user characteristics, such as age, gender, level of competence etc.? Well, the easy answer to this is we do not! There are plenty of studies that illustrates how virtual agent appearance influence how users perceive them and interact with them, but that this differ for different user groups and contexts (cf. Baylor, 2009; Lee, 2003; Ring et al., 2014). Thus, the more important question to answer is, what dimensions of agents and users’ characteristics are important to consider when designing an agent for a specific context? Many dimensions can be considered, Straßmann & Krämer (2017) lists the following: i) embodiment vs no embodiment ii) species or humanness iii) realism iv) 2D vs 3D, v) and feature specification (such as, gender, ethnicity, clothing, etc.) The domain and task is also important to consider. For example, Lee (2003) showed that people conform more with virtual characters when the topic and gender of the agent was matched, and Ring et al. (2014) found that a realistic agent was deemed more appropriate for a medical task, while a cartoonish agent was preferred for social conversation.

User preferences can thus vary with domain and task, but there are also differences related to user variables such as gender and age. For gender, similarity attraction have been found in many studies, but also some were the opposite gender was preferred, and it seems to depend on the domain and task (cf. Baylor, 2009; Payne et al., 2013). Age has not been as thoroughly explored, with only a few studies. Straßmann & Krämer (2017) interviewed students and elderly and asked them with describing how a virtual agent that was to assist them with “daily life support” should look. Most of the students (M_age = 24.17), preferred an animal, object or machine-like agent with a low degree of realism, while the elderly (M_age = 67) mostly preferred a human-like agent, and a high degree of realism. Another study investigated children’s (M_age = 10.06) versus adults’ (M_age = 20.33) attribution of mental and emotional abilities to agents with varying degrees of human-likeness, from a blob, cat, cartoonish human to realistic human (Krumhuber et al., 2015). They also asked the participants to rate likability, trustworthiness and intelligence of the agents. Both adults and children rated the blob as the most likable, followed by the cat and realistic human for the adults, and the cat and both humans for the children. Regarding the other traits pertaining to mental and emotional states, the adults rated the humans higher than the other agents, while the children gave fairly equal rating to all agents.

Method

In the study presented in this paper, we wanted to further explore how children and adults differ in preferences of visual appearance of agents, and learn which dimensions that are distinguishing. Since we wanted to gain insights not only into what agents were preferred, but also why, we settled on an interview study, which also included a short verbal questionnaire. A total of 20 participants took part in the study. Since we used a convenience sampling of voluntary children from a primary school we had 10 children (M_age = 6.90) whereof seven were boys and three were girls. For the adults we had a balanced group regarding gender (M_age = 36.30) recruited at a workplace and a university. Designing the stimuli, i.e. the various virtual agents, there were many possible dimensions that could have been included, but we choose to focus on species (human, animal, robot, elf) and for the humans, realism (2D vs 3D), gender (male, female) and age (child, adult). The participants were shown 12 different depictions of agents (see Fig 1) and were asked to choose a favorite. A follow up question was asked about the reason for the choice, and what personality traits the agent had. The participants were then asked to choose agents four more times with the same follow up questions being posed for each. The choices to be made were i) the least liked of the agents, ii) an agent to teach you how to play an instrument, iii) an agent to help you search information and/or web pages on the Internet, and iv) and agent to hang out with as a friend. After the choices had been made the participants, together with the interviewer, filled out questionnaires for the chosen agents. The questions in the questionnaire where a selection of items from the Godspeed questionnaire (Bartneck et. al, 2009) focusing on aspects of anthropomorphism, animacy, likability and perceived intelligence, with a Likert scale from 1 to 4.
Results

The table in Fig 1 shows the number of times the different agents were chosen by users for different purposes. Preferred and least liked are divided by age and gender. There were both similarities and differences between the age groups in regard of preferred and least liked agent. In sum, the boy agents (2D and 3D), robots and panda were preferred by most children as well as adults. That the children preferred the fantasy figures was in line with earlier studies (Krumbhuber et al. 2015). The 2D humans stood out as the least liked, but while most of the adults disliked the 2D man, several of the boys disliked the 2D woman. With regard to the motivation of the choices, many said that their preferred agent looked kind, and the least liked agents were often described as looking mean or not nice. For example, the 2D man was described by the adults with words such as murderer, dangerous and not trustworthy. This was corroborated by the questionnaire were the chosen 2D and 3D humanlike agents were rated as above average (3-4) on all items: realism, extroversion, niceness, smartness and ability to express emotions, while, the least liked agents received below average scores on niceness and emotions. The preferred fantasy agents were rated a little lower for extroversion and ability to express emotion with around average (2-3). In the interviews, children also mentioned attractiveness (or lack thereof) as a reason for (not) preferring an agent.

During the interviews, the experiment leader noted the type of pronoun used to describe the agents. In cases where pronouns were not used, the participants were asked what gender they ascribed to the character. Interestingly, four of the participants saw the 2D boy or 3D boy as androgynous, and one even saw it as a female. All other human-like agents were perceived as their “intended” gender. The panda was mostly referred to as male (7) but also androgynous (3), the Pepper robot, the humanoid robot and elf were all seen as both male (5/7/3), androgynous (5/3/2) and female (2/7/1). Female participants preferred agents they saw as female and male participants preferred male agents or to a lesser degree androgynous agents. When it comes to the least liked agent, the female participants showed the reversed patterns of choosing a male agent, in many cases the 2D man, or ascribing a male gender to it in cases of the robots or the elf. The male participants only partly followed the same patterns, they often chose a female character such as the 2D woman or 2D girl, and ascribed female gender to the robots and the elf, but some of them also choose the 2D and 3D man. The female participants often referred to the chosen agents as kind or mean, while the male participants more often described them in terms of their looks, ability, and willingness to interact. Thus, the results for gender preferences follows previous patterns of similarity attraction (cf. Baylor 2009, Payne et al. 2013) and that androgynous agents are ascribed gender by users.

What stood out when it came to the question of which agents that were preferred for different context, was that many choose the 3D woman, even though she had previously not been selected as the favorite agent by more than one person. She was described by the adults as reliable, being a good teacher, looking kind, sympathetic and competent. One can speculate that either female agents are common in these areas, for example, Siri and virtual assistants on web pages, or that these are traditional female domains, with teachers and secretaries. According to Social role theory, female agents would likely be more chosen for tasks that are perceived as belonging to their role, which was partly confirmed in a study of virtual agents for checkout in a store (Payne et al., 2013). Most of the other agents chosen for the different domains, like the boys, panda and robots, had been popular choices as
favorites. However, the elf also made a surprising appearance in the social context for women, who preferred a fantasy figure since they had many real life human friends and would like something different in a virtual friend. A very similar argument was made by one of the participants in the study by Straßmann & Krämer [4], who said, “There are enough humans around me. I don’t feel the need to interact with a human in this way in addition”. When the participants rated the chosen agents in the questionnaire niceness was rated above average or high for all domains. Unsurprisingly intelligence was rated above average and high for agents chosen for learning as well as information providing, while emotions were above average for the humans and panda, but not for the robots that were rather rated very low. Surprisingly emotions were not important for the social domain either, but here extroversion, i.e. being talkative, was rated highly.

Discussion and conclusions

There are several limitations of this study that needs to be considered both when interpreting its results and when designing further studies in this area. Since we did interviews and used convenience sampling, we had very few participants and a skewed selection of gender for children with only three girls, while we had seven boys. Since we saw differences in gender preferences we need to consider that the children were represented by 2/3 boys when we compare them to the adults. Regarding the design of the study, the participants only got to choose one agent for each question, so we do not know much about the agents not chosen, or what would have been a second or third choice. In a future study, it might be interesting to let the participants rank the agents instead. The participants only got to view headshots of the agents, so they neither saw their bodies, nor got to interact with them. This might result in different choices. If participants can interact with an agent, it is also possible to collect behavioral data instead of, or as a complement to, self-reported data.

Even though there are several limitations to this study, we think that for the purpose of informing future studies both regarding possible hypothesis and methods to be used, we have learned some important lessons. The stimuli in this study was partly chosen from freely available pictures online, which resulted in the 2D man and woman differing from the others. It highlighted that when constructing stimuli, it is important to be aware of the multitude of possible dimensions, and to make conscious choices of which to keep constant and which to manipulate. For example, attractiveness and the expression of emotions such as anger and the approachableness were mentioned. It is also important with manipulation checks, in this case, it was very clear for the gender of the agents, where some participants ascribed a different gender than the experimenter had intended. The task and domain must also be considered since we saw that asking for the users’ favorite agent versus asking for the one best suited for a specific task or context yielded different responses. Depending on the domain, one need also ponder which user variables that are important and can interact with the dimensions of the agents’ appearance and behavior one want to study. For example, in a learning context, it may be prudent to take into account students’ proficiency or self-efficacy, while in more social settings personality traits such as extroversion/introversion is more important.

References

Straßmann C., Krämer N.C.(2017) A Categorization of Virtual Agent Appearances and a Qualitative Study on Age-Related User Preferences. In *Proc. IVA 2017* (pp. 403-422), Berlin: Springer
Robot-Enhanced Therapy for Children with Autism

Erik Billing1 and Tom Ziemke1,2

1Interaction Lab, University of Skövde
2Department of Computer & Information Science, Linköping University

erik.billing@his.se / tom.ziemke@liu.se

Autism Spectrum Disorder (ASD) is a developmental disorder affecting communication and behavior (NIH, 2018). ASD is typically characterized by abnormalities in social interactions and communication, most often appearing within the first two years of life. ASD is referred to as a spectrum as it comprises a wide range of difficulties for the individual, ranging from relatively mild problems as ADHD to severe deficits in communication and behavior requiring constant supervision out through life.

People diagnosed with ASD often receive medication that may reduce symptoms, e.g., a reduction of aggression, hyperactivity, or anxiety. While medication can be effective in terms of reduced symptoms, there is no medical treatment for ASD as of today. The only known treatment with persistent effects involves cognitive-behavioral therapy.

We here briefly summarize some of the outcomes of the European research project DREAM – Development of Robot-Enhanced therapy for children with Autism spectrum disorders1 (Esteban et al., 2017). The project has been running since spring 2014 and is planned to finish fall 2018, engaging a consortium of seven partners. DREAM targets a specific form of behavioral therapy for children diagnosed with ASD, called Robot Enhanced Therapy (RET).

RET is a variation of the traditional autism therapy, called Applied Behavior Analysis (ABA). ABA targets fundamental social skills that can improve the child’s ability to interact and reduce repetitive behavior. One challenge is that many children diagnosed with ASD perceive social situations as highly distressing, making the child refrain from social interaction and communication, further reducing the child’s chances to learn and improve social skills. In these situations a social robot can provide an environment where the child gets a chance to learn and explore social behavior on a lower level of complexity, less associated with anxiety (Sartorato, Przybylowski, & Sarko, 2017). One key advantage of using robots as social stimuli is that simplicity and predictability of interaction can be maintained, while at the same time providing an engaging embodied social interaction for the child (Huijnen, Lexis, Jansens, & de Witte, 2016).

A wide range of robots have been evaluated in the context of autism therapy, for example Labo-1, Keepon, Pleo, Probo, NAO, KASPAR, FACE. Although several of these robots have shown promising results, most of the clinical findings are exploratory (Diehl, Schmitt, Villano, & Crowell, 2011) and build on a Wizard of Oz (WoZ) setup (Huijnen et al., 2016; Thill, Pop, Belpaeme, Ziemke, & Vanderborght, 2012), where the robot is remotely controlled by a human operator. While WoZ configurations can be an effective evaluation method, making the robot respond as if it can interpret the actions of the child, it is impractical for long-term studies and does not constitute a realistic scenario for clinical use due to its high human workload. The opposite extreme, involving a fully autonomous robot, may be equally problematic. An autonomous robot makes decisions and adapts its actions to the current situation without human intervention. This is still very difficult to achieve from a technical point of view, and may also raise ethical concerns (Richardson et al., 2018).

One of the goal of DREAM is to develop an application for RET using supervised autonomy (Senft, Baxter, Kennedy, Lemaignan, & Belpaeme, 2017). In this context, the robot executes a pre-scripted intervention protocol while autonomously adapting to the child’s behavior. The intervention is monitored by the therapist, who can intervene in situations where the robot and/or the child are not following the described intervention protocol.

---

1 DREAM is funded under the 7th frame programme of EU, grant #611391. https://www.dream2020.eu.
Clinical evaluation

From a clinical perspective, DREAM comprise a rigorous evaluation of RET as an alternative to Standard Human Treatment (SHT). Both RET and STH follow a ABA protocol. Specifically, the evaluation target joint attention, imitation and turn-taking skills for children with severe autism. These skills are believed to constitute key components – or building blocks – necessary for improved interaction and communication abilities in general. The clinical evaluation, a randomized controlled trial which is still on-going, is conducted using an experimental setting were children are randomly divided into two groups: one using RET and a control condition using SHT. In both conditions, the child is seated at a customized intervention table and recorded using a camera-based sensing system (Cai et al., 2018) able to record body motion, eye-gaze, face expressions etc. In the RET condition, a SoftBank NAO robot constitutes the interaction partner while a therapist assigned a supervising role is seated at the side of the intervention table (Fig. 1). In the control condition, two therapists are taking part, one as the interaction partner and the other as a supervisor/mediator.

65 children in the age range 3 to 6 years have participated in the evaluation. Up until today, 50 children have completed the protocol of eight interventions over six weeks. Each intervention involves five to seven interaction sessions targeting joint attention, imitation or turn-taking. In addition, each participant goes through a diagnosis protocol, using Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) as a measurement tool to quantify the child’s degree of autism. The diagnosis is employed before the first intervention and after the child has completed all interventions. The differences in ADOS score constitute the treatment effect. Only children with an initial ADOS score above 12 and no other neurodevelopmental disorders were included in the study.

Technical contributions

In parallel to the clinical application, DREAM involves research pushing the state of the art in several research areas. We employed component-based software engineering (CBSE) as a way to tackle the need for flexibility for developers while still maintaining a strict interface allowing integration of all system components developed within the project (Vernon, Billing, Hemeren, Thill, & Ziemke, 2015). While the most common motivation for CSBE is component re-use, it is here also employed as in integration platform.

The robot application is implemented using the middleware YARP (Metta, Fitzpatrick, & Natale, 2013), integrating a total of 16 software components. Each component implements its own executable that can be co-located on a single machine or distributed over a cluster of computers. The complete application was initially implemented using placeholder components, constituting a complete executable system running at a very early stage of development. Through the developmental process, placeholder components were replaced by partial or complete implementations, until a complete system was achieved.

In order to guarantee the desirable robot behaviour also when the system could not correctly classify the child’s performance, Supervised Progressively Autonomous Robot Competencies (SPARC) was employed (Senft, Baxter, Kennedy, & Belpaeme, 2015; Senft, Lemaignan, Baxter, & Belpaeme, 2016). With this approach, the system will display suggested action a few seconds prior to execution, giving the therapist the chance to intervene and change action if desirable. As such, the system can progress from a WoZ-like setting requiring heavy human workload to supervised autonomy with little human intervention. Another key component is the integrated sensing system, combining recognition of body motion, gestures, head orientation, eye-gaze and voice (Cai et al., 2018; Zhou, Cai, Li, & Liu, 2017).
Discussion

The clinical evaluation is still on-going and its final results are still not available. A preliminary analysis based on partial data does not reveal significant differences between the two evaluated conditions (RET vs SHT). However, both conditions show effects in terms of reduced ASD for all primary outcomes (turn-taking, joint-attention and imitation skills). If these patterns carry over also to the results from the complete evaluation, one interpretation could be that RET may not primarily be feasible for clinical use, but does provide a safe and engaging environment where the child could train social skills also outside the clinical environment. With this perspective, RET may be an excellent alternative in situations where it is difficult to get access to traditional therapy and it may also give children access to more therapy than otherwise would be possible.

References (bold, 11pt)


Contagion of beliefs as an underpinning of Theory of Mind

Andreas Falck

1Department of Psychology, Lund University

2Institut Jean Nicod, Département d’études cognitives, ENS, EHESS, PSL University, CNRS, Paris France

andreas.falck@psy.lu.se

Beliefs are, in a basic sense, contagious. This is because, as we follow other’s attentional direction automatically and without reflection (see Birmingham and Kingstone, 2009), we tend to focus on the same information as our peers, without necessarily reflecting on this fact (Falck et al. 2014; Heyes, 2014). I argue that contagion is a powerful means of explaining how we come to share beliefs, regardless of whether we reflect on the fact that beliefs are being shared. Contagion-based sharing is cognitively lean, in the sense that we do not typically have to think about the fact that information is shared. This does not mean, however, that the shared information itself is not a target of elaborate cognitive processing. Rather, sharing through contagion leaves cognitive resources free for processing the shared information, rather than being consumed in order to enable the sharing itself. This explains why sharing information is generally efficient, even though other people are involved who intermittently attract attention. Here I will discuss this phenomenon in the context of so-called false belief tasks (FBTs henceforth), which are tasks probing children’s and adults’ ability to understand situation in which another person has a false belief (Wimmer & Perner, 1983). The seminal task featured a story about Max and his piece of chocolate. Max first put his chocolate in a drawer, subsequently Max’s mom moved it in his absence, so that he came to have a false belief about its location. When Max later wants his chocolate again, where will he look for it? These tasks nowadays come in two main versions. In verbal FBTs the child is probed for their understanding by an experimenter’s question (“Where will max go now to get his chocolate?”), while in non-verbal FBTs children or adults are observed for cues of understanding in their non-verbal behaviour. Passing the task implies indicating that Max will go where he left it, rather than where the chocolate really is. Children typically start passing verbal FBTs around 4 years, non-verbal ones earlier. FBTs of both kinds are often argued to reveal the existence of a ‘theory’ of other people’s minds (ToM henceforth). By appealing to the sharing of beliefs by contagion I will question this close association of FBT performance and ToM ability for a range of tasks.

Contagion of beliefs alleviates the need for meta-representation

Several accounts of ToM suggest that meta-representation of other people’s beliefs are at the heart of human social ability, possibly innate (e.g. Leslie 1994; Baillargeon, Scott & He 2010; Carruthers, 2013; Westra 2017). On this account we track others’ perspectives, i.e. what they see and hear, and attribute to them beliefs according to what we have seen them experience. Consequently, we represent them as representing the world (hence meta-representation), a notion at the core of the ToM concept (Premack & Woodruff, 1978). Here I wish to raise a puzzle for this account, based on a prediction that the theory seems to imply. If we maintain parallel representations of what other people believe at all times we would expect this mechanism to be equally active regardless of whether our social peers’ apparent beliefs differ from our own. Consequently, situations where others have false beliefs should require similar amounts of processing to those where others’ and own beliefs agree. However, this is not what the available evidence suggests. Children have great struggle with verbal FBTs at 3 years of age, and pass carefully controlled non-verbal FBTs around 13-15 months (Surian et al. 2007; Onishi & Baillargeon, 2005). Toddlers share information in a systematic way earlier, around 9 months (Tomasello 1995). In a similar vein, both toddlers and adults sometimes overestimate the amount of shared information in a situation, and act as if they assumed their peers to have access to information that they visibly do not (Moll et al. 2011; Keysar et al. 2003). Acknowledging this problem, it has been proposed that young children (and perhaps adults under cognitive load; see Schneider et al., 2012) attribute their own beliefs to others by default, and that cognitive resources are needed to inhibit this default attribution (e.g. Leslie et al. 2005). Arguing against this view, Westra (2017) points out that after controlling for other developmental variables such as age and verbal IQ the correlation between tests of ToM and tests of inhibition becomes quite low. Rubio-Fernandez (2015) investigated the purported default attribution of true beliefs in adults, and found that answering questions about
an agent’s false belief about an object’s location lead to facilitation rather than inhibition of the true location of the object. It seems that situations in which others have false beliefs are harder to cope with for both adults and children, and that this is not due to attributing true beliefs by default. A more parsimonious hypothesis is that beliefs are typically only attributed to others when the situation demands this, either because the joint activity breaks down or as one is explicitly asked to do so (as in the verbal FBT). Under this model, attributing beliefs at all is only needed when beliefs diverge in the first place. The question of understanding false beliefs translates to a question of understanding situations in which a belief is not (any longer) shared. In the following I will outline how this alternative model might guide the interpretation of two types of experiments, non-verbal and verbal FBTs.

Non-verbal FBTs and contextual specification

An obvious challenge to my account is the existence of ‘fast and frugal’ abilities to process others’ false beliefs in adults and infants (Apperly & Butterfill, 2009), as is paradigmatically investigated in non-verbal FBTs, such as the seminal study by Onishi and Baillargeon (2005). In this study, infants (15 months old) watched an experimenter put a toy in a box. Then the toy was moved to another box in the experimenter’s absence, and infants subsequently looked longer (indicating surprise) when the experimenter searched the actual rather location rather than the one where she believed the toy to be. This suggests that infants in fact expect others to act according to their beliefs rather than according to reality. My account explains infants’ success as follows. Every non-verbal false-belief task, as a design consequence, needs to start out with a shared belief (which under the ToM theory would be the same as attributing a ‘true’ belief to the other person). When the infant in Onishi and Baillargeon’s task watches the toy move in the adult’s absence, information pertaining to the toy ceases to be shared between the infant and the experimenter. Understanding false beliefs in these tasks would then be a matter of activating previously shared information associated with the ‘owner’ of the false belief (Falck et al. 2014). When the experimenter returns in the end of the experiment, they will still be associated with the old information about the toy’s location in the observing infant’s mind. Given that the agent is salient in the scene, they might act as a contextual cue to the infant, forcing a re-instatement of the old situation in which the agent’s belief was true (and the information about the toy’s location was shared). This simple and untested suggestion allows for a parsimonious and non-metarepresentational account of adults’ and infants’ performances on non-verbal FBTs, which predicts a close correlation between infants’ performance on non-verbal FBTs and their ability to form contextualised memories. It is parsimonious in the sense that it avoids postulating a special-purpose module for belief representation, instead relying on memory mechanisms that are multi-purpose and have independent empirical grounding. Moreover, it avoids postulating a large number of stimulus-response mappings to account for adaptive behavior in different contexts, a known problem with other low-level accounts that rely on the infants’ attributing behavioural dispositions to the agent (Tomassello & Call, 2006; Sober, 2016). My account predicts that the ability to contextualise memories of events in social contexts emerges some time before 15 months of age, something that might be investigated with young infants in habituation studies.

Verbal FBTs and pragmatic consequences of sharing

Contagion is also present in interactive situations, such as episodes of mutual joint attention, when both parties are aware that they are sharing information. This is because even though such awareness exists, it does not imply that every way in which the interacting parties control each other’s attention is cognized upon (Wilby, 2010). This arguably leaves cognitive resources free for other tasks. Thus, sharing through contagion is present not only in interactive FBTs such as the helping task (Buttelmann et al., 2009), but also in verbal FBTs, as these are interactive too – the child is interacting with the experimenter asking the question (Wilby, 2011; Helming et al. 2014). The difference here is that in the verbal FBTs, the infants share information with an experimenter who is not also the agent whose belief becomes false, adding complexity to the task in a way that is often overlooked. This can partly explain why verbal FBTs are so hard for children under the age of four. Rubio-Fernández and Geurts (2013) showed that three-year-olds pass verbal FBTs if they are allowed to play with the protagonist

---

1 The arguments presented here was previously discussed in a slightly different frame in the introductory chapters to my PhD thesis (Falck, 2016).
(who was a puppet) and enact the puppet’s search for the desired object. Similarly, Psouni et al. (2018) showed that children between three and four years of age perform better on a verbal FBT if they attend the material together with the experimenter as co-observers rather than as narrator and listener. Whereas Rubio-Fernández and Geurts removed the experimenter as an agent and let the child share perspectives with the protagonist only, Psouni et al. instead made the experimenter share the child’s perspective on the story. Both procedures create less reason for the child to reflect on the experimenter’s beliefs and motives. Helming et al. (2014) and Westra (2017) have targeted the pragmatic implications of findings like Rubio-Fernández and Geurts’. They suggest that the child misinterpret the experimenters question to be about either finding the “correct” answer of a trick question or about helping the protagonist to find the correct location. Whereas this is a plausible possibility, my account suggests that already the difference in perspective between the experimenter and child – that of a narrator that knows the story and that of a listener – forces the child to invest cognitive resources in understanding the experimenter’s question. Whereas when the perspective on the story is shared, as in Psouni et al.‘s study, cognitive resources are free to process the protagonist’s belief. This can explain why three-year olds pass the task when they are not asked to provide an answer, and why providing an answer about what a protagonist will do is simpler when the person asking the question share more common ground with the child.

Concluding remarks

Here I have suggested that contagion of beliefs is a basic, non-meta-representational way of converging on shared knowledge. It can act as a substitute for ToM if paired with mechanisms for contextualising memories, and as a building block for more advanced ToM abilities. It can also explain the peculiar performance differences in a range of variants on the verbal FBT.

References


Information comes in many forms from many sources in daily life, such as newspapers, television, the internet, and personal interactions. People need to sort the information and keep track of which is true and believable, on the one hand, and false and fictional, on the other. Fictional information can be defined as information that an individual perceives and remembers, but which the person believes should not be evaluated against the real world (that the truth value of propositions should not be evaluated against the real world, that people and places do not exist for real, and that events did not actually take place) (Gander, 2005). One fundamental question is how this distinction is implemented in the cognitive system. The nature of the mental representation and processing of fictional information is still unexplained (Potts & Peterson, 1985; Potts, St. John, & Kirson, 1989; Gerrig & Prentice, 1991; Prentice & Gerrig, 1999; Marsh, Meade, & Roediger, 2003; Marsh, Butler, & Umanath, 2012).

The source monitoring framework (SMF) (Johnson, Hashtroudi, & Lindsay, 1993) attempts to explain how people encode, store, and retrieve information in memory from various sources. The SMF has been applied to a number of phenomena, such as witness testimony, persuasion, amnesia, and memory and aging. The SMF has also been proposed as a theory to explain how people handle fictional information (Johnson et al., 1993), but empirical research on this issue is lacking. One problem with the SMF is that one and the same source often contains both fictional and factual information (such as television, or a friend). Another problem, which is still an open question, is whether people need to remember the source in order to make the distinction between remembered fictional and factual information.

The aim of the current study is to investigate whether fictional status attribution is a case of source attribution, or whether it is a distinct process. Is the source monitoring framework able to explain fictional status attribution of information? How similar or distinct are the processes of source attribution compared to fictional status attribution? Can people forget the source but still remember whether something is fictional or not?

The present study sets up three hypotheses. Hypothesis 1: People are more correct in their memory of the fictional status of information than they are concerning the source of that information. Even the presence of a difference would indicate that the processes of attribution of source and fictional status are distinct. The hypothesis includes a direction of the effect, however, because the source attribution process involves a multitude of processes (Johnson et al., 1993) it is considered more complex and more prone to errors. The SMF, in contrast, predicts that performance should be the same since fictional status is dependent on source. Hypothesis 2: People do not use similar processes when judging if something is fiction or not as when making source attributions. The case of distinct processes would manifest itself by the fact that neither status and source attributions nor response times in the two conditions are correlated. On the other hand, if judgment of fictional status is dependent on source then the measures of accuracy and response times in the two conditions would correlate. SMF predicts correlation between fictional status judgment performance and source attribution performance. Hypothesis 3: Response times are faster when attributing fictional status than when attributing source. Source attributions depend on several processes and is therefore predicted to take longer. In contrast, the SMF predicts similar response times since fictional status is seen as source information.

**Method**

*Participants.* Thirty-four participants were recruited for the experiment (17 males and 17 females).

*Design.* The experiment was a within-subjects design with attribution type (fictional status, source) as the independent variable and memory recall accuracy and response time as dependent variables.
Material. The stimuli consisted of statements about unusual animals, using the animal's name and a description (e.g., The Takahe is a flightless bird indigenous to New Zealand). The statements were pilot-tested on five other participants prior to the experiment in order to verify that they were unfamiliar (consequently, one statement was replaced because the animal was familiar to one participant). The material differed only on whether it was labelled fiction or fact. The presented stimuli were constructed with two dimensions: source and status. The source was a visual presentation in which the statements were presented. The two sources were made to be distinct from each other and differed in the font colour, background colour, shape and surrounding border colour and type. Status was either fictional or factual, labelled by the words "FICTION:" or "FACT:" above the statement. Responses were made by pressing one of two keys on the computer keyboard. Statements’ fictional status and presentation source were counter-balanced and presented in a randomised order for each participant. Statement order, question order of source and fictional status, and order of alternatives within the questions were randomised for each participant in the test phase.

Procedure. The entire experiment was computerised. Participants were seated in front of a laptop computer. On an initial screen, information about the experiment and contact information about the researcher was presented. After filling in their informed consent to participate in the experiment, and their gender and age, the screen showed instructions for the task. The participants were told that they would see statements about animals, some fictional and some factual, shown above the statement, and that they would be required to remember what they see. After pressing a key, the presentation phase started. Each statement was presented centred on the screen on a source background for 10 seconds followed by a black screen for 2.5 seconds. After all the statements had been presented, the test phase started. Participants were asked to recall source and fictional status for all the statements, in a cued recall memory test. The two sources were presented on the screen and the participant was asked to select one by pressing one of two keys on the computer keyboard. Participants responded to fictional status in the same way. The questions were self-paced. After the test phase, participants were thanked for their participation. The entire procedure lasted around five to seven minutes.

Results

While data analysis is still incomplete, preliminary results show that fictional status attribution appears to have a higher score than source attribution (both seem to be different from chance, that is, if the participants were guessing). In contrast, response times do not seem to differ. Finally, preliminary analysis of correlation between the two measures seem to show no association.

Discussion

In the task of remembering information about unusual animals, preliminary results indicate that that people are more correct in their memory of the fictional status of information than they are concerning the source of that information (which would support Hypothesis 1). Further, people seem to use distinct processes while doing so since the measures were not correlated (which would support Hypothesis 2). Response times do not seem to differ for the two tasks (which would refute Hypothesis 3).

The study set out to test the source monitoring framework (SMF) (Johnson et al., 1993) as a theory of how people process fictional information in memory. Preliminary results seem to go against the SMF. Fictional status attribution does not seem to involve processing source information. Instead, attribution of the fictional status of information seems to be a separate process. Consistent with these findings, it is proposed to consist of following an associative connection in memory (established during encoding) between the information and fiction as a category. This proposal needs to be evaluated in future studies.

If hypothesis 3 (concerning response time differences) is not supported, this could be because of several reasons. Although the response time means preliminary appears in the direction of the hypothesis, it was observed that the standard deviation was high which could lead to a lack of statistical significance. Another reason could be simply that the two processes take similar amounts of time; so that even if they are distinct processes, it may not be reflected in differing response times.
Some open questions remain about the design of the experiment. First, are the statements used in the experiment representative of fictional information? Perhaps a larger context is more typical of fictional information, in terms of length (such as a paragraph or several paragraphs) or structure (such as a narrative structure). Second, what is the validity of the encoding situation? Is it feasible to let people see a list of statements, and later ask questions about fictional status and source? Maybe insufficient resources are spent when encoding the sentences if the participants don’t know exactly on what aspects they will be tested later? Finally, it could be argued that the source attribution in the present study depends on the specific source information used: shape and colour. Perhaps other source information (such as spatial location, textual or auditory modes, or information coming from different persons) would make the source attribution easier. More distinct and elaborate sources could be tested in future studies.

References


