Preface

Welcome to SweCog 2019!

In an article published in Nature: Human Behavior, Núñez et al. [2019] asks What happened to cognitive science? The authors review bibliometric and socio-institutional aspects of the field and argues that the transition from a multi-disciplinary program to a mature inter-disciplinary coherent field has failed. Looking at the Swedish environment, we can nothing but agree. Many of us identifying ourselves as researchers in cognitive science are working at departments primarily focused at other disciplines, teaching within other objects and publishing in journals and conferences adjacent to the field. The diversity of cognitive science is also present in the number of directions that has has evolved over the years. The embodied approaches that many of us align with are not evolving towards a coherent view, but is today found under numerous labels such as situated cognition, distributed cognition, extended cognition, and enactive cognition. The so called 4E perspectives on the field have now ventured beyond the four, and is today more often referred to as the multi-E framework.

While we agree with Núñez et al. that we remain a multi-disciplinary, multi-perspective, and multi-method group of researchers who may share an interest for the science of the mind, rather than a coherent approach or perspective, we disagree that this entails a failure for the enterprise of cognitive science. We dare to say that the Swedish Cognitive Science Society has embraced the multi-perspectives idea by adopting an inclusive approach in the selection of research and methods presented at our conferences. We hope that SweCog will remain a forum for inclusive discussions, working against discipline conformism and isolation, in a time where both public and scientific debate is increasingly shattered.

Most welcome to SweCog 2019! In this booklet you’ll find abstracts and short papers for all oral and poster presentations at the conference.

The reviewers were:

Rebecca Andreasson, Erik Billing, Hung Chiao Chen, Andreas Falck, Gordana Dodig Crnkovic, Linus Holm, Arne Jönsson, Erik Lagerstedt, Maurice Lamb, Jessica Lindblom, Rob Lowe, Sara Mohamed, Vipul Nair, Kajsa Nalin, Patrick Oden, Maria Papakosma, Kai-Florian Richter, Niclas Sthål, and Felix Thiel.
Conference Programme

Thursday November 7th

12:00 — 13:00  Registration
13:00 — 13:15  Welcome
13:15 — 14:00  Invited speaker - Ronald van den Berg (Stockholm University)
               *The visual system as a detective*
14:00 — 14:20  Coffee
14:20 — 14:40  Gabriela-Alina Sauciuc (p. 8)
14:40 — 15:00  Nils Dahlbäck (p. 5)
15:00 — 15:20  Nicholas Judd (p. 6)
15:20 — 15:50  Break
15:50 — 17:20  Poster session (see p. 11)
17:20 — 17:40  Sara Stillesjö (p. 8)
17:40 — 18:00  Lars Oestreicher (p. 46)
19:00          Conference dinner

Friday November 8th

09:00 — 09:45  Invited speaker - Paolo Medini (Umeå University)
                *Microcircuits behind multisensory integration in the mouse neocortex*
09:45 — 10:15  Coffee
10:15 — 10:35  Sofia Thunberg (p. 51)
10:35 — 10:55  Thomas Hellström (p. 6)
10:55 — 11:15  Timotheus Kampik (p. 37)
11:15 — 11:35  Edvin Listo (p. 43)
11:35 — 13:00  Lunch
13:00 — 13:20  Linnea Karlsson Wirebring (p. 9)
13:20 — 13:40  Vipul Nair (p. 7)
13:40 — 14:00  Andreas Falck (p. 6)
14:00 — 14:15  Break
14:15 — 15:00  Invited speaker - Andreas Theodorou (Umeå University)
                *Responsible AI with Black, White, and Glass Boxes*
15:00 — 15:15  Conference closing and Coffee
15:15 — 15:45  SweCog annual member’s meeting
Verbal presentations

Below you’ll find abstracts for the talks by Nils Dahlbäck, Andreas Falck, Thomas Hellström, Nicholas Judd, Vipul Nair, Gabriela-Alina Sauciuc, Sara Stillesjö, and Linnea Karlsson Wirebring. In addition, the following short papers are presented verbally:

- Edvin Listo Zec and Olof Mogren — Grammatical gender in Swedish is predictable using recurrent neural networks (p. 43)
- Lars Oestreicher — Detecting chunks in EEG-patterns using neural networks supporting children with severe communication impairments (p. 46)
- Samantha Stedtler, Timotheus Kampik, and Helena Lindgren — Artificial agents for fairness in economic games: a demonstration (p. 37)
- Sofia Thunberg and Tom Ziemke — Are People Ready for Social Robots in Public Spaces? (p. 51)

Nils Dahlbäck

Clark and Chalmers’ made in their (1998) paper Extended Mind the provoking claim that mind is not solely something in the head but extends into the world. If a process that, were it done in the head would be called cognitive, is in part done in the world, then the world is part of the cognitive process.

This claim triggered an intense discussion, where a multitude of arguments both for and against has been presented. We do not wish here to present further arguments for or against this thesis. Instead we want to address the problematic conceptual foundations for this discussion.

We claim that the participants base their arguments for or against on quite different explicit or implicit definitions of ”cognition”, which in turn suggests that the different ”marks of cognition” are about different kinds of cognition. This suggests in turn that ”cognition” is a family concept in Wittgenstein’s sense. A consequence of this, we claim, is that no progress can be made in this discussion until the different meanings of the term cognition are clarified. This in turn requires an untangling of the major conceptual dimensions used in the different definitions or views of what cognition is and is not. In this paper we will present a first stab at the task of developing a taxonomy of such dimensions.
Andreas Falck  
*Do others’ apparent beliefs affect our object detection?*

Do we automatically adopt others’ beliefs as if they were our own? A famous experiment by Kovalcs et al. (2010) suggest we do: in their study, adults were faster to process an unexpected event if the event appeared to be expected from another agent’s point of view. However, later replication studies have highlighted alternative explanations for the results, and the available evidence for this purported phenomenon is inconclusive (Phillips et al., 2015).

In the present research, adults were repeatedly required to indicate the location of an object revealed at one of two possible locations, each time after watching one of four different types of animations involving an agent and the object. The animations were such that the object would show up either where the participant believed it to be (i.e. where they had seen it hide last), or where the agent believed it to be (i.e. where the agent saw it hide last). It was found that participants did not spontaneously respond in line with the agent’s apparent belief, but could be tricked into doing so by being instructed to keep track of the agent’s attention. Thus no evidence of automatic belief adoption was revealed, however, contextual factors may create similar effects.

Thomas Hellström  
*Understandable robots*

The ability to estimate mental states and planned actions of others is essential for the human species. As robots become more and more intelligent and autonomous, the need to make also their actions and decision processes understandable increases. Failed understandability impairs interaction quality, and degrades user experience, efficiency, and safety. This need has been acknowledged also for AI in general, recently under the notion of Explainable AI. In this talk I will describe our conducted and ongoing work on understandable robots [Hellström and Bensch, 2018]. Understandability is formally defined, and a general model for mutual understanding based on Theory-of-Mind is described. In the model, Communicative Actions is the general tool by which an agent makes itself understandable. Five questions are identified as particularly important when designing for understandability. In addition to this theoretical work, I describe ongoing work in which we address some of the identified questions:

- Teams of robots verbally describing what they are doing for human bystanders
- The effect of backchanneling as confirmation of understanding in HRI

Nicholas Judd

Genetic factors and socioeconomic (SES) inequalities play a large role in educational attainment, and both have been associated with variations in brain structure and cog-
nition. However, genetics and SES are correlated, and no prior study has assessed the neural effects independently.

Here we used polygenic score for educational attainment (EduYears-PGS) as well as SES, in a longitudinal study of 551 adolescents, to tease apart genetic and environmental effects on brain development and cognition. Subjects received a structural MRI scan at age 14 and 19. At both time-points, they performed three working memory (WM) tasks.

SES and EduYears-PGS were correlated ($r = 0.27$) but both also had independent effects on brain structure and cognition. Specifically, lower SES was related to less total cortical surface area and lower WM. EduYears-PGS was also related to total cortical surface area, but in addition had a regional effect on surface area in the right parietal lobe, a region related to non-verbal cognitive functions, including mathematics, problem solving and WM. SES, but not EduYears-PGS, affected the change in total cortical surface area from age 14 to 19.

This is the first study demonstrating the regional effects of EduYears-PGS and the independent role of SES on cognitive function and brain development. It suggests that the SES effects are substantial, affects global aspects of cortical development and continues to influence brain development during adolescence.

Vipul Nair

This study investigates human perception of action similarities, where we explore patterns or clusters of similarities (similar actions)? Moreover, if there are clusters, what are the salient visual features of the action that humans rely upon? Insights to these questions are helpful to devise computational models to create visual primitives for human motion segmentation and understanding. Such models would be advantageous in understanding a human-event scenario or a human-robot interaction setting, for the model would find the same action regularities salient, as would a human. To that extent, we study how humans judge similarities between different familiar human hand based actions. A total of nineteen commonly seen kitchen based hand actions (e.g., cutting bread, washing dish) are chosen as stimuli. Participants performed two psychophysical experiments, an action similarity judgment task (experiment 1) and action discrimination task (experiment 2). Human judgment data are analyzed to see for human similarity patterns. Additionally, similarity patterns from three different visual computing algorithms for motion understanding (low-level spatial and velocity features), are used to compare against the human judgment patterns, which shows some overlap. We discuss the similarity patterns as a way to model action-perception that builds on action primitives.
Gabriela-Alina Sauciuc

**Social antecedents and consequences of rhythmic behaviours in chimpanzees (Pan troglodytes)**

Recently there has been a growing interest in the rhythmic abilities of nonhuman animals, as a way of tracking the evolutionary origins of human musicality. Currently, there is an effervescence of speculations and debates on the topic, with theories being generally divided between relating the emergence of rhythmic abilities to demands for (1) flexible vocal learning (and thus specific audiomotor neural adaptations) or (2) social bonding and cooperation. Extant evidence from non-human species is, however, insufficient to substantiate any of these theories, and data from our closest genetic relatives - the chimpanzees - is almost totally missing. To address this issue, we tracked the range, contexts and consequences of rhythmic behaviours exhibited by chimpanzees in observational study carried out in Sweden (Furuvik Zoo - N=6, Kolmården Zoo - N=22) and in Spain (MONA Foundation, N=14). Preliminary results show that the great majority of rhythmic behaviours (i.e. 94%) occurred in social contexts (both affiliative and agonistic), being relatively absent in solitary contexts, with the exception of object play. Given their high frequency in social contexts, it is plausible that rhythmic behaviours accomplish a communicative function, which in turn corroborates social theories on the evolutionary origins of rhythmic abilities.

Sara Stillesjö

**Retrieval Practice promotes superior memory retention via higher brain activity in Hippocampus and Caudate independent of Cognitive Ability**

Here, we provide novel neuroimaging evidence of brain regions related to the well-known effects of retrieval practice (“the testing effect”) in a sample of students (N = 86, MAge = 18.17). One week prior to fMRI, students learned a set of word-pairs, half through retrieval practice and half through study. Contrasting items learned through retrieval practice with those acquired through study revealed higher brain activity in a number of brain regions, including the bilateral hippocampus and the caudate. Moreover, brain-behavior correlations were identified such that functional brain activity in caudate-hippocampus was positively correlated to behavioral performance, although stronger for items learned through retrieval practice. A principle component analysis was used to extract a broad cognitive composite index predictive for learning. The index was then used to divide subjects into cognitive ability groups (CA, low/high). Independent of CA, significant testing effects were present. Notably, the higher brain activity in caudate and hippocampus associated with retrieval practice was also evident independent of CA. These results confirm recent suggestions that the superior retention associated with retrieval practice is independent of cognitive ability. Retrieval practice may promote retention by updating and storage of contextual memory representations via higher brain activity in the caudate-hippocampus.
Linnea Karlsson Wirebring

Active compared to passive mathematical learning: durable effects on brain activity one week after constructing your own solutions

Recent theoretical advances in how to best foster durable mathematical learning in school focus on the benefits of active (constructing one’s own solutions) compared to passive (imitating a given solution) modes of learning. FMRI provides unique information about the underlying neurocognitive mechanisms involved in these different learning modes. Previous neuroimaging studies of mental arithmetic demonstrate that brain regions in the parietal cortex, especially angular gyrus, is more activated for well-learned compared to novel arithmetic problems. Angular gyrus is thus seen as a characteristic neural correlate to successful arithmetic learning and a key region of interest when comparing brain activity after active vs passive learning. Here, 71 pupils ($M_{Age} = 18.0$) practiced on 24 mathematical problems: 12 in an active mode and 12 in a passive mode. One week after learning participants performed a mathematical test while undergoing fMRI. In a whole-brain analysis, significantly higher brain activity in left angular gyrus was found for test items previously learnt with the active compared to the passive mode. These results have important implications for the design of mathematical teaching methods, and add to the growing body of evidence favoring active compared to passive learning modes as a way to strengthening memory and performance.
Posters

Please find abstracts for posters presented at SweCog 2019 below. In addition, the following short papers are presented as posters:

- Erik Billing, Lars Hanson, Maurice Lamb, and Dan Högborg — *Digital Human Modelling in Action* (p. 24)
- Johanna Björklund and Marlene Johansson Falck — *How Spatial Relations Structure Linguistic Meaning* (p. 29)
- Marie Dornheim and Elisavet Kaltsouni — *Aphasia Rehabilitation Agent: towards personalized aphasia rehabilitation* (p. 33)
- Mattias Lantz Cronqvist, Rachel R. Philips, Carl-Oscar Jonson, and Erik Prytz — *Preliminary findings of an empirical study into human ability to accurately perform visual estimates of blood loss* (p. 40)
- Priyantha Wijayatunga — *Probability, Paradoxes and Human Thinking* (p. 54).

Mattias Arvola

*Swarms, Teams, Packs, or What? Generative Metaphors for Design of Systems-of-Systems*

We are working on how to design the interaction with and control of autonomous systems-of-systems that build on large-scale integration of multiple individual systems to meet overarching needs. Our first scenario will be the use of drones (unmanned aerial vehicles, UAVs) in forest firefighting. Two critical aspects of such systems-of-systems are that they will show emergent behaviours that cannot be attributed to component systems, and they must be resilient and continue to function despite the failure of single components.

One starting point for our design research effort is to figure out how to conceptualise what kind of beast the system-of-system is. This kind of systems are often conceived using biomimetic metaphors of a swarm or a team. However, we could also think of them as groups, flocks, herds, shoals, packs, crowds, colonies etc. With each of the metaphors comes different control strategies, structures (i.e. hierarchy, division of labour), controllability, and level of autonomy. The conceptual metaphor for the
Posters

Elpida Bampouni

Numerous research studies have been done in order to explore the link between video games and cognitive skills. Visual search skills in particular are utilized by action gamers in realistic game settings on many occasions, in addition to its usage during their daily lives. Yet, results of relevant studies have been inconclusive in regards to whether gamers possess enhanced visual search ability or not. In this study we decided to approach the topic using realistic measurements of visual search in comparison to the classic computerized visual search tasks. Action gamers and non gamers were recruited and presented with a realistic visual search task of our creation. The participants looked at photographs of Umea city and were asked to quickly respond upon seeing a) a person or b) a car in the images. Our results showed that all participants were faster upon viewing a person than a car. They further indicate no differences between action gamers and non gamers in both accuracy and speed.

Hadi Banaee

Can Computers Play Codenames as We Play? Towards Cognitive Aspects of Subjective Games

This study aims to address the problem of modelling subjective games by taking into account the cognitive aspects involved in playing such games with human. The main purpose is to characterise the cognitive features of subjective games, distinguishing them from the objective ones (e.g., Chess and Go game) that are already modelled based on existing AI solutions. Codenames, as well known social board game, is a suitable case study to investigate the cognitive aspects of interest in subjective games. This game consists of two teams with a set of words shown to all the players. One player (spymaster) makes a clue associated with a set of words belonging to her/his team. Then, the clue is shared with teammates (detectives) to let them guess the chosen words. Playing Codenames is highly related to how humans organise and associate the concepts in mind, and share them in a common-sense way. The existing AI approaches are not often designed to address such cognitive challenges, and therefore, are inadequate to model cognitive processes involved in playing subjective games. Rather in this work, we investigate cognitive aspects that are required to be considered in AI techniques in order to enable machines to play such games with humans.
Frida Bertilsson, Tova Stenlund, Carola Wiklund-Hörnqvist, and Bert Jonsson

*Retrieval practice in school: What happens when we merge cognitive science and teaching practice?*

Substantial evidence exists for retrieval practice as an effective learning strategy that leads to superior long-term retention compared to other strategies (i.e., the testing effect). However, less is known about the implications and consequences of merging science and teaching practice. Here we implemented a retrieval practice paradigm in an upper secondary school mathematics course \( n = 59 \). Quizzes were based on the course material (one per chapter in the course book) and were available at the course site, ensuring that the pupils had unlimited access to each quiz during the corresponding chapter. To examine how frequently the pupils used quizzes while studying on their own we applied an A-B-A-design \( (A = \text{quiz use regulated by pupils}; \ B = \text{quizzes also included in classroom teaching}) \). Initial analyses indicate that frequency of quiz use is positively related to performance (as measured by score on a posttest). Importantly, when quizzing was included in class, the performance level on the posttest increased to a minimum of 60%, as compared to 10% when quizzing was not mandatory. This suggests that while retrieval practice is beneficial for learning, one challenge is motivating students to use the technique when it is not included in mandatory classroom teaching.

Oscar Bjurling and Mattias Arvola

*Can’t See the Swarm for the UAVs? Traversing System Levels in Human-Swarm Interactions*

As a system-of-systems, an autonomous swarm of Unmanned Aerial Vehicles (UAVs) must afford human-robot interaction and control in several layers, ranging from the entire swarm, to subswarms, to individual UAVs and their payload. Moreover, in a forest firefighting scenario, circumstances may require an operator to frequently switch between these different system levels. The complexity of this task switching activity is exacerbated by several factors, such as the level of system autonomy, governing command and control structures, available means of control, and, of course, the environmental and task contexts. The goal of our research is to understand the cognitive processes and workloads involved in this task switching activity, and to develop a framework for the design of safe, effective, and intuitive human-swarm interaction interfaces.

Emma Christersson

Classic psychedelic substances, such as lysergic acid diethylamide and the active compound in magic mushrooms, psilocybin, are being studied again in a renaissance of psychedelic research. Psychedelic substances have profound effects on perception, emotion, and cognition, as well as the capacity to induce mystical-type experiences
and ego-dissolution. Recent clinical studies indicate that these substances have positive effects on patient populations and healthy participants, both acutely and long-term. Neuroimaging studies show that psychedelics alter neural integration, by the disintegration of normally stable resting-state networks, and increasing network connectivity between normally anti-correlated networks. From the recent research on psychedelics a novel theory of conscious states has evolved, the entropic brain theory. A comparison between two theoretical accounts of the brain basis of psychedelic-induced altered state of consciousness will be presented, between the novel entropic brain theory and the integrated information theory, a well-established theory of consciousness within cognitive neuroscience.

Pia Elbe

Working memory can be influenced by a host of different factors, including daily habits of computer use (e.g., Cansino et al., 2018). The aim of this study is to test if estimated computer use and perception of computer skills are related to working memory, and to see which other factors may play a role. In the Successful Aging dataset, 188 participants (aged between 50 and 75) answered questions about their computer habits and completed the following working memory tasks: symmetry span, operation span, and rotation span (Foster et al., 2014). In addition, an n-back task (2-N) and a fluid intelligence task (Ravens progressive matrices) were used. Fluid intelligence ($r = 0.83$, $p < 0.01$), education ($r = 0.17$, $p = 0.02$), and age ($r = 0.48$, $p < 0.01$) were significantly correlated with the reported average number of hours spent at the computer. Regression analyses were used to test if education, intelligence, age, and working memory are related to computer habits (adjusted $R^2 = 0.24$, $F(6,149) = 8.68$). From this study we may conclude that computer use does not make a large impact compared to other variables. Computer habits are part of a bigger picture in determining working memory performance.

Daniel Fellman — The role of strategy use in working memory training outcomes

Abstract: Cognitive mechanisms underlying the limited transfer effects of working memory (WM) training remain poorly understood. We tested in detail the Strategy Mediation hypothesis, according to which WM training generates task-specific strategies that facilitate performance on the trained task and its untrained variants. This large-scale pre-registered randomized controlled trial ($n = 258$) used a 4-week adaptive WM training with a single digit n-back task. Strategy use was probed with open-ended strategy reports. We employed a Strategy training group ($n = 73$) receiving external strategy instruction, a Core training group ($n = 118$) practicing without strategy instruction, and Passive controls ($n = 67$). Both training groups showed emerging transfer to untrained n-back task variants already at intermediate test after 3 training sessions, extending to all untrained n-back task variants at posttest after 12 training sessions. The Strategy training group outperformed the Core training group only at the beginning of training,
indicating short-lived strategy manipulation effects. Importantly, in the Core training group, strategy evolvement modulated the gains in the trained and untrained n-back tasks, supporting the Strategy Mediation hypothesis. Our results concur with the view of WM training as cognitive skill learning.

Filip Grill

*Neural correlates of incentive processing: a two-component response involving hippocampus*

Well-functioning incentive processing is of paramount importance for an individual’s survival and wellbeing. The ventral striatum (VS) which is densely innervated by dopamine neurons from the ventral tegmental area (VTA) has been identified as an integral brain region for incentive processing across several species. In rodents, it has been shown that subfields of the hippocampus modulate the availability of the dopamine neurons in the VTA by influencing VS neurons, creating a hippocampus-VS-VTA loop. Human resting-state fMRI has also demonstrated functional coupling between VS and hippocampus. Somewhat paradoxically however, the hippocampus is seldomly considered a brain area implicated in human incentive processing tasks. The present study sought to elucidate the relationship between VS and hippocampus in relation to incentive processing utilizing both task (gambling) and resting state fMRI data from the human connectome project (N = 175, mean age 28.8). We show that a cluster in the VS responding to incentive processing could be further parcellated into two distinct sub-clusters, each corresponding to different cortical networks, one of which heavily involving hippocampus. The hippocampus target identified in this analysis also showed a robust task-dependent response (as indicated by a replication analysis) but it was relatively subtle and only included a small area of the medial hippocampus, which may be the reason that prior studies in smaller samples have not reported hippocampal activation during incentive processing.

Manar Halwini, Eva Lindgren, and Kirk Sullivan

*Towards a model of motivation and engagement for adult professional second language learning*

Using questionnaires and interviews with adult immigrant professional learners of medical Swedish, this study aims to develop a model of that better describes their motivation and engagement in their professional language learning. Starting from Directed Motivational Currents [Dörnyei et al., 2016] we consider how to better integrate Engagement with Language [Svalberg, 2009] into the dynamics of process of learning a second language. Specifically we hope to develop a model that will allow more detailed investigation of relationships between:

1. Facilitative motivation structure and cognitive engagement
2. Positive emotionality and affective engagement
3. Goal-oriented motivation and social engagement

This will provide a better understanding of how learners navigate different situations to develop L2 skills, engage autonomously with the professional language being learnt, and continue interacting with other, including native, speakers of the L2.

Paul Hemeren

The ability to classify human movement into meaningful gestures or segments plays a critical role in creating social interaction between humans and robots. In the research presented here, grasping and social gesture recognition by humans and four machine learning techniques (k-Nearest Neighbor, Locality-Sensitive Hashing Forest, Random Forest and Support Vector Machine) is assessed by using human classification data as a reference for evaluating the classification performance of machine learning techniques for thirty hand/arm gestures. The gestures are rated according to the extent of grasping motion and the extent to which gestures are perceived as social. The results indicate that the machine learning techniques provide a similar classification of the actions according to grasping kinematics and social quality. Furthermore, there is a strong association between gesture kinematics and judgments of grasping and the social quality of the hand/arm gestures. Our results support previous research on intention-from-movement understanding that demonstrates the reliance on kinematic information for perceiving the social aspects and intentions in different grasping actions as well as communicative point-light actions.

Erik Lagerstedt

The main focus in my research is perception and interaction, leading to an interest in categorising artificial agents with respect to what function they are perceived to have for another agent. Such categorisations can from a pragmatic perspective help when evaluating options in design of artificial agents, and from a theoretical perspective provide a structure on which known phenomena and empirical findings can be mapped. In addition, such structure can facilitate comparisons between artificial agents, but also between artificial and natural agents. This is particularly relevant for the field of Human-Robot Interaction, in which different kinds of agents of various degree of automation are considered.

There are several existing categorisations of artificial agents, and they vary in discrimination criteria, context and domain [e.g., Sheridan and Verplank, 1978, Breazeal, 2004, Lagerstedt et al., 2017]. Creating categories that emphasise the perceived function or intention will provide a pragmatic system that could be used by the artificial agents themselves. Self-driving cars, for instance, need to determine what other road-users are present to safely navigate the road system. Classifying the road-users in terms of what they can and probably will do, could facilitate improved predictions and decisions.
Charlotte McNulty

The literature suggests that fluid intelligence correlates with individual differences in cortex and subcortical volume, although these associations are modest. The lack of a strong correlation between brain structure and fluid intelligence may be because available studies have focused on establishing linear associations between measures of fluid intelligence and brain structure. New evidence has suggested that biological correlates of individual differences in cognitive measures associated with fluid intelligence are better described in terms of population subgroups. Latent-profile analysis was used to identify subgroups of fluid and crystallised intelligence scores in a population of healthy young adults \( n = 1179; 22-35 \) years, which were then compared on basic demographics and structural brain imaging variables to assess the hypothesis that associations between fluid intelligence and brain structure are not uniform across the entire population. The analysis identified four subgroups which demonstrated significant group differences in nucleus accumbens and total cortex volume. The findings confirm that cognitive measures associated with intelligence are better described in terms of population subgroups, and that fluid intelligence and brain structure do not scale linearly across the entire population. Future research should consider the potential interactions between neurological and psychiatric disorders, and associated dopamine levels, and intelligence scores.

Maria Mystakidou and Ronald van den Berg

No effect of gamification on performance in a visual working memory task

The question whether visual working memory (VWM) capacity is flexible or fixed has challenged cognitive psychologists for more than 15 years. Here we investigate whether people flexibly increase capacity when they are motivated to do so. To manipulate motivation, we "gamified" a standard VWM paradigm (delayed estimation). The experiment started at set size 1. Participants received a score based on the accuracy of their responses and would "level up" to the next set size when they had accumulated enough points. We varied the difficulty of this "game" through the scoring function to obtain "easy", "medium", and "hard" conditions. A fourth group of subjects was tested using the original, non-gamified paradigm (control group). On average, we found no effect of game difficulty on VWM performance, nor a difference between the control group and the three gaming groups. These results suggest that subjects may be unable to modify their VWM capacity through motivation, at least based on the experimental design we used. However, further research should be conducted to rule out alternative explanations for this null finding, with a focus on the interaction between level of motivation and difficulty level.
Kajsa Nalin

Decision making processes are complex and important to understand to avoid unnecessary complications due to design mistakes when creating artefacts or work processes. Research on cognitive processes, such as decision making, has been carried out in several scientific domains, with various cognitive perspectives, and using numerous methods [Edwards, 1954, Kahneman and Tversky, 2000, March, 2010, Patel et al., 2002]. However, the theories usually have an individual perspective, per se not necessarily a problem, though there are occasions when a systems perspective is beneficial.

One situation where new insights followed from using a systems perspective was when the medical decision making process at a surgical emergency ward was studied [Nalin, 2017]. The patients were followed through the entire medical decision making process, with the purpose of examining the gap between descriptions of and the actual process [Nalin, 2006]. In comparison with the description, the study’s results highlighted the actual process’ complexity and multitude of systems components.

Another area of interest is teaching in higher education, and in 2019, the University of Skövde implemented new guidelines for examination. The method described above will be used to conduct an exploratory pilot study, focusing on the examination process of theoretical course content, with the primary goal of creating a better understanding of such processes.

Robin Pedersen

Complex cognitive operations require the brain to flexibly reconfigure the functional coupling between networks in response to environmental demands. More segregated and specialized networks with high local efficiency during resting state are associated with greater cognitive performance. In contrast, greater network integration during goal-oriented tasks with a relatively higher cognitive load are also associated with greater performance. This highlights the importance of dynamic interactions between systems that mediate more integrative or associative operations. However, questions still remain about the relationship between cognitive performance and the degree of flexible adaptation of individual networks at varying levels of environmental demands.

In this ongoing project, we explore the association of cognitive performance and task dependent functional reorganization of individual resting state networks. Functional connectivity analyses and graph theoretical measures are used with rest and task fMRI data from 180 elderly subjects (64-68 years, mean = 66.2, SD = 1.2; 80 women) in conjunction with offline cognitive measures of working memory, episodic memory and perceptual speed. All data was collected as part of the Cognition, Brain, and Aging (COBRA) project. Preliminary results and general discussion are to be presented.
Kai-Florian Richter

As in any interaction process, misunderstandings, ambiguity, and failures to correctly understand the interaction partner are bound to happen in human-robot interaction. We term these failures ‘conflicts’ and are interested in both conflict detection and conflict resolution. In that, we focus on the robot’s perspective. For the robot, conflicts may occur because of errors in its perceptual processes or because of ambiguity stemming from human input. This poster presents a brief system overview, and provides some details on the fuzzy inference system used to encode the robot’s scene understanding, which is used to match identified objects against the objects mentioned by the human interaction partner.

Julia Rosén

Are ethics overlooked in the field of Human-Robot Interaction?

There are, in any scientific research practice, ethical guidelines to adhere to. For example, the Ethical Principles of Psychologists and Code of Conduct by the American Psychological Association [2017], WMA Declaration of Helsinki by the World Medical Association [2018], and Ethics for Researchers by the European Commission [2018], all offer principles on how to conduct research ethically. Although the formulations of guidelines vary, the following aspects are usually included: data protection, privacy, informed consent, deception, and debriefing. However, these aspects are rarely explicitly addressed in publications in the field of Human-Robot Interaction (HRI). Proper ethical conduct is an integral part of scientific research and ought to be included in this field as well. There might be societal implications if participants in HRI studies are deceived regarding the actual capabilities of social robots.

A literature study is planned in order to investigate and analyse how ethical issues are considered in publications from the HRI 2018 conference, e.g., what ratio of publications dealing with human participants mention ethical aspects explicitly. The aim is to contribute to a methodology in HRI where ethical aspects have a significant bearing.

Avinash Singh, Neha Baranwal, Kai-Florian Richter, Thomas Hellström, Suna Bensch

Towards verbalized robot-robot interaction

In this poster, we present ongoing work on a framework for robot-robot interaction that allows for keeping human users in the loop. The robot-robot interaction comprises plan derivation, plan execution and verbalization of robot actions. Our framework uses cooperating distributed grammar systems (CDGSs) which are formal models for distributed problem solving. We use CDGSs for plan derivations and an additional blackboard architecture for plan execution. The robots concurrently decide which robot is to execute the next step(s) in the plan. All robots verbally comment on the current and
planned actions, such that a human bystander can understand what is going on. These verbalizations may range from simple statements, such as “I will move the green mug from A to B”, to more complex sub-plans, such as “Robot 1, please move the red glass from A to B, so that robot 2 can move it to C”. With this approach we contribute to making robot teams understandable for human users. The framework is implemented on three Pepper robots that move objects in a table-top scenario while talking about their actions.

**Kata Szita**

*Narrative Experience in Immersive Moving-Image Entertainment*

The key factors of narrative experience on personal immersive platforms lie in the options to create an irreproducible, personal experience, thus the analysis of immersive moving-image entertainment departs from two lines of thoughts. First, embodiment – viewer behavior as a result of bodily connection with the screen; and secondly, the sensory scope – the systematic view of sensory information originating from content, screening platform, and the user’s body as part of story perception. Being in a close phenomenological connection with an audiovisual stimulus, what one receives is a tailor-made narrative built upon sensory alertness, perceptiveness, cognitive processing; and incoming information is used for making (conscious or unconscious) decisions about whether and how to change the sensory connection with the content. To study these factors, under the term phenonarratology, a model of storytelling, narrative comprehension, and the visual and auditory scope of moving-image representation was developed to explore the challenges of non-canonical storytelling and story perception and to map the cognitive and phenomenological grounds of spectatorship. This paper aims to explore the applicability of the phenonarratology model to a wide range of personal immersive entertainment platforms and present possible methods for the theoretical and empirical study of cognition in terms of interactive and 360-degree movies and media experience in augmented or virtual reality environments, and games.

**Sam Thellman**

When interacting with robots, uncertainties about our own future states and actions are compounded by the fact that we are also often uncertain about what the specific robot knows (and doesn’t know) about objects and events in the context of the interaction [Thellman and Ziemke, 2019]. For example, uncertainty about what a robot knows about the presence and attributes of an object in a shared physical space makes it difficult to predict how the robot will behave in response to that object. It also makes it difficult to communicate using definite expressions like “this candle” to refer to objects in a shared environment. In this talk, we argue that estimating what a robot knows may in many cases be more difficult than estimating the knowledge of another person, due to the fact that robots often have different perceptual and cognitive capabilities from humans. We present empirical evidence from a forthcoming study to support this claim, and briefly discuss how cognitive science research can contribute to resolving
some of the uncertainties that people have about what robots know.

Sofia Thunberg

Beyond the Novelty Effect: Studying Long-Term Interactions with Social Robots in Elderly Care

The older population is growing and society is facing an increase of people with cognitive disabilities, e.g. dementia. Social robots have the potential of providing company, lower stress levels and decrease loneliness. But there is only one robot that has been tested with controlled long-term repeated studies [Wada et al., 2005]. Because of this, there are no evidence that social robots have the same impact beyond the novelty effect, which is the tendency for performance to initially improve when new technology is instituted, and not because of any actual improvement. Therefore, I propose long-term pre-test post-test controlled group design studies with elderly people in care homes. But I also find it important to design and implement robots and other smart-home technique that brings a meaning to the user. My first project would therefore be to conduct a workshop design study with elderly people, both living at home and in elderly care homes, to assess what they want, think and need. Secondly, I plan to conduct a study at an elderly care home facility with the following scenario: A humanoid robot interview the residents about their life, with the purpose to measure the effects for human-robot interaction over time.


Digital Human Modelling in Action

Erik Billing¹, Lars Hanson², Maurice Lamb¹, Dan Högberg²

¹Interaction Lab, University of Skövde
²School of Engineering Science, University of Skövde

erik.billing@his.se

Digital Human Modelling (DHM) is an active research field with the goal of creating detailed models of the human body (Scataglini & Paul, 2019). The field has its roots in anthropometrics and biomechanics and careful measurements of the human body in terms of, e.g., bone, muscle, and fat. Such digital models of human bodies are commonly used as, verification and visualization software, primarily to assess ergonomic aspects of e.g. driver environments and workstations in industry. With the shift towards Industry 4.0 and an increased use of simulation and visualization software in all design phases, DHM has become an important component allowing engineers to evaluate ergonomics of new products and workplaces early in the design phase, long before the first physical prototype is built.

In addition to models focusing on anthropometrics alone, there is also an array of models covering biomechanical and physiological properties (Bubb, 2019). However, modelling of cognitive functions has a vastly different tradition. Cognitive modelling has its roots in psychological descriptions of mental function, including studies of both human and animal cognition (Smith & Kosslyn, 2013). The first computational cognitive models appears in the 40’s and 50’s with the cognitive revolution. Here, we find seminal works by Hebb (1949) and Kleene (1956) that inspired the formation of artificial neural networks underpinning most of today’s AI technologies and Deep Learning. The Magical number seven, plus or minus two by G. A. Miller (1956) was a very influential formalization of the limits of human working memory that still shapes our understanding of memory up until today. Another example is the Human associative memory (Anderson & Bower, 1973) that later developed into the cognitive architecture ACT-R, which is probably still one of the most well-known models of cognitive function around.

Notably, the dominant theories and models in both DHM and cognitive modeling are influenced by René Descartes’ Mind–body dualism. The cognitive models mentioned above are in a very concrete sense “without a body” – the body is essentially described as a container with input and output modalities. Similarly, the tradition of digital human modelling describes the human body’s mechanics without cognition. While DHM does involve simulations of actions and work tasks, modeled behaviors are highly specified by the designer with a focus on biomechanical constraints.

Over the last three decades, embodied and situated approaches to cognition have emphasized the role of the human body and its environment for cognition. We have to a large degree shifted from symbolic descriptions of reasoning to dynamic descriptions of behavior. As a result, cognitive models are increasingly formulated as complex systems of dynamic interdependent biomechanical and cognitive elements (Pfeifer, Bongard, & Grand, 2007). Still, very few of these models get close to the anthropometric detail found in DHM literature. Recently, there are also initiatives from the opposite direction, incorporating cognitive aspects into DHM models. One such example is RAMSIS Cognitive³ allowing modelling of human visual perception, comprising analysis of direct view, minimum visual range, and optical attributes of displays. However, these examples are still sparse and are primarily used to evaluate a limited set of e.g. perceptual properties (Scataglini & Paul, 2019).

Within the project Virtual Ergonomics², gathering researchers from both DHM and cognitive sciences, we see a potential to bridge these domains in a way that is fruitful for both disciplines. Specifically, we aim to develop models that can assess and predict human action, considering anthropometrics, biomechanical and cognitive

---

¹ RAMSIS Cognitive is a plugin for the RAMSIS Software by Human Solutions, www.human-solutions.com
² Virtual Ergonomics is a synergy project lead by the University of Skövde, funded by the Swedish Knowledge Foundation (KK).
aspects of action selection and execution. While action selection can imply a wide range of scenarios from selecting a piece of candy from a candy bin to selecting a company’s business plan for the upcoming fiscal year, we will focus specifically on the analysis of repetitive work tasks, e.g. as part of manual assembly in industry or lifting tasks in health care.

Modelling action

In DHM research, action selection is not part of the model itself, but specified at design time. The exact motions are typically put in terms of optimization over some performance index (Björkenstam et al., 2017). The performance index is derived from the biomechanical constraints of the human body, considering flexibility of joints, anthropometrics, and the human’s environment including e.g. obstacles. Action goals are specified in an absolute frame of reference, without consideration of how humans’ select and produce these actions.

DHM models provide biomechanically valid simulations of preselected actions and action plans but typically do not simulate the cognitive constraints that lead to the selection of actions and action plans. Given the abundance of motor degrees of freedom in the human body, additional constraints introduced by a cognitive system can have a significant impact in how the task is completed (Latash, Levin, Scholz, & Schöner, 2010). For example, in the context of object reaching and grasping, anticipation of end-state comfort can constrain both how and where an individual grasps a task relevant object (Rosenbaum, Van Heugten, & Caldwell, 1996). This phenomenon has also recently been modelled in a DHM setting (Yang & Howard, 2020). While motor variability in human behavior is often treated as noise to be reduced, recent attempts to quantify and model motor variability have demonstrated that some dimensions of variability can drive task learning and success more than others and that motor variability can be used as a measure of both expertise and adaptability (Latash et al., 2010; Scholz & Schöner, 1999).

In the present work, we will leverage on the modelling tool IPS IMMA – Intelligently Moving Manikins3. IPS IMMA comprise a detailed biomechanical model allowing designers to simulate many types of tasks involving complex working positions (Hanson et al., 2019), allowing ergonomic evaluations directly in the simulated environment. From a cognitive science perspective, tools like IMMA provides a powerful environment where many embodiment effects can be modelled and tested with a physical realism rarely available in cognitive models. The work is still in preparation phase, with focus on two specific directions:

Can we predict when people are reaching or walking? In previous research (Lamb et al., 2017), we studied hysteresis effect in human behavior. Specifically, action selection is not only based on an optimal strategy, but also primed by previous behavior. We plan to further investigate, and model, this phenomenon in a reach-grasp-carry task where participants have the option to either stretch over a table to reach for an object, or walk around the table to achieve a more comfortable reach. The setting bares many similarities to repetitive assembly tasks in industry, commonly modelled with DHM tools like IPS IMMA. Individual’s selection of behavior (reach or walk) certainly depends on many factors, where anthropometrics is one. Using the detailed anthropometrical model fitted to each participant, we will investigate to what degree calculation of comfort in combination with hysteresis is a predictor for participants’ behavior selection.

Is proactive eye-gaze present also in repetitive tasks? It is well known that people display a specific eye-hand coordination pattern during object manipulation, where the eye-gaze is proactive in relation to the hand. Since first demonstrated by Johansson et al. (2001), this phenomenon has been intensively studied in experimental settings, primarily during action observation (e.g., Flanagan & Johansson, 2003). Proactive-eye gaze could constitute a very useful clue for human-robot interaction (Billing, Sciutti, & Sandini, 2019; Sciutti et al., 2012). However, to what degree this proactive eye-hand coordination is present also in everyday, repetitive, conditions is to large degrees unknown. We plan to implement proactive eye-gaze as part of IPS IMMA and in this way model complex working situations. The model behavior can later be compared to observed worker behavior in industry, allowing underlying theories related to proactive eye-gaze to be validated in real working environments.

3 IPS IMMA by Industrial Path Solutions, https://industrialpathsolutions.se/ips-imma/
Acknowledgements

This work has been made possible with the support from The Knowledge Foundation and the associated INFINIT research environment at the University of Skövde in Sweden, in the project Synergi Virtual Ergonomics (KKS Dnr 20180167), and by the participating organizations. This support is gratefully acknowledged.

References


Johanna Björklund* and Marlene Johansson Falck

How Spatial Relations Structure Linguistic Meaning

Work in progress

Keywords: Corpora analysis, Prepositional phrases, Semantics.

The usage patterns of spatial prepositions give us an idea of how natural-language speakers use spatial relations to structure their thinking. They provide information not only of how we structure our experiences of concrete, real world physical objects such as bowls, houses, and stages, but also of how we structure our experiences of more or less abstract concepts such as our concepts of the air around us, love, or a diet. From the usage patterns of the English preposition in we learn that not only real world physical entities such as bowls and houses are construed as bounded entities (things can be in a bowl, or in a house), but this is also true of the somewhat more elusive notion of the air around us (we say that something is in the air), and of the abstract concept of love (we are in love). Accordingly, from the usage patterns of the preposition on we learn that not only stages are construed as supporting surfaces (we say that we are on stage, or on the stage), so are the diets that we are on or the gas that our cars run on.

We have indeed long known that natural language speakers tend to use spatial relations to structure other domains of experience (Lakoff and Johnson, 2003, 1999), but no one has done a systematic analysis of how spatial relations are used to structure meaning. One of the earliest example’s is the work by Reddy (1993) on the conduit metaphor. It showed that speakers of English has a preferred framework for conceptualising communication as a conduit that transfers thoughts and feelings from one person to another, and that within this framework speakers of English insert their thoughts and feelings into words, into phrases, into sentences, into phrases, and into poems, which then contain the inserted meanings (e.g. we express things in words, or in a sonnet). However, despite the multitude of linguistic investigations into metaphorical mappings between experiential domains we still do not know what types of concepts natural-language speakers typically structure by which spatial relation. Although knowledge about structures such as these could be

*Corresponding author: Johanna Björklund, Dept. Computing Science, Umeå University
Marlene Johansson Falck, Dept. Language Studies, Umeå University
Tab. 1: The ten most common collocates that follow three different prepositional sentences involving the preposition *in*. After each collocate, we give the number of times the particular construction was seen in COCA. For example *in a new way* occurs 416 times, and *in any way* 5648 times. We see that *in a new* is used roughly equally many times in a concrete and abstract sense; that *in an old* is primarily used in a concrete way; and that *in any* primarily lends itself to abstract constructions.

<table>
<thead>
<tr>
<th>In a new...</th>
<th>(Occ.)</th>
<th>In an old...</th>
<th>(Occ.)</th>
<th>In any...</th>
<th>(Occ.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>way</td>
<td>416</td>
<td>house</td>
<td>51</td>
<td>way</td>
<td>5648</td>
</tr>
<tr>
<td>book</td>
<td>313</td>
<td>movie</td>
<td>24</td>
<td>case</td>
<td>4835</td>
</tr>
<tr>
<td>direction</td>
<td>242</td>
<td>farmhouse</td>
<td>19</td>
<td>event</td>
<td>1781</td>
</tr>
<tr>
<td>era</td>
<td>222</td>
<td>building</td>
<td>18</td>
<td>other</td>
<td>1677</td>
</tr>
<tr>
<td>light</td>
<td>208</td>
<td>warehouse</td>
<td>16</td>
<td>given</td>
<td>688</td>
</tr>
<tr>
<td>study</td>
<td>106</td>
<td>man</td>
<td>15</td>
<td>kind</td>
<td>505</td>
</tr>
<tr>
<td>place</td>
<td>88</td>
<td>wooden</td>
<td>13</td>
<td>form</td>
<td>437</td>
</tr>
<tr>
<td>world</td>
<td>85</td>
<td>stone</td>
<td>13</td>
<td>direction</td>
<td>417</td>
</tr>
<tr>
<td>movie</td>
<td>57</td>
<td>book</td>
<td>13</td>
<td>particular</td>
<td>274</td>
</tr>
<tr>
<td>country</td>
<td>57</td>
<td>folks</td>
<td>11</td>
<td>sense</td>
<td>257</td>
</tr>
</tbody>
</table>

of interest to any area that deals with knowledge based on human thought patterns, systematic analyses of this type of knowledge have not been made.

In an on-going project, we focus on how speakers of American English use the spatial relations encoded by the prepositions *in* and *on* to construe meaning. We investigate both how they are used to describe real-world physical relations, such as those involved when we are *in* prison, or *on* a stage, and abstract relations such as those involved when we are *in* love, or do something fun *on* weekends.

Our aims are (i) to identify the types of concepts are typically construed by means of these specific spatial relations, (ii) to provide a better understanding for how speakers of American English use spatial relationships to construe meaning, and (iii) to show how we can systematically study how the speakers of a language convey different concepts by means of spatial relations.

The patterns identified might be valuable for robotic world modelling, where an autonomous agent gradually assembles information about its surrounding and may use human speech as one source of input. If the agent encounters an expression of the form *in the third* *<unknown>*., it may draw the conclusion that regardless of what *<unknown>* will turn out to be, it refers to something that is considered to have container-like qualities. Examples are world (which refers to a real world physical entity that we can be in), and year (which refers to a temporal unit that is metaphorically understood as something that we can be in). The agent can thus draw partial conclusions about an unknown term, by being observant of the prepositional phases in which it occurs.
In our initial experiments, we use the web-based interface against the corpus of contemporary American English (COCA) compiled by Davies (2008). COCA consists of approx. 560 million words divided over 220,225 document, published over the years 1990–2017. The corpus is evenly divided between the five genres of spoken, fiction, popular magazines, newspapers, and academic journals, and predominantly collected from native English speakers. The corpus interface allows us to search the corpus for relevant phrases, and to find the most common collocates for our target prepositions. A collocate for a preposition is a word that frequently co-occurs with the preposition. Because the number of spatial prepositions in a language is limited, the ways in which they are typically used to construe meaning can be investigated systematically by going through the most common usage patterns of each preposition one by one. To illustrate the idea, Table 1 shows the ten most common collocates following three different prepositional phrases.

A natural continuation of studying collocates is to compare prepositional phrases by means of word embeddings. A word embedding is, simply put, a function that maps each word in the corpus to a real-valued vector that reflects how likely the word is to occur in different contexts (Harris, 1954). Two words that are used in similar ways have as a rule also similar embeddings. Word embeddings have recently been generalized to phrase embeddings (Yu and Dredze, 2015), and we are interested in finding out what insights may be gained by means of this new analysis tool. Our current understanding is that the vector space models can provide a distance relation between prepositional phrases, that in turn can be analysed by an expert to reveal new relations and usage patterns.

Our previous corpus linguistic analysis of a random set of abstract instances of the prepositions in and on (Johansson Falck, 2017) has shown that these two prepositions are used to construe different types of abstract concepts. Our study confirms the observation by Reddy (1993) that segments of language are construed as containers, but it also shows that several other types of abstract concepts are construed this way (e.g. the manners, fashions, or ways in which we do something, and abstract areas, fields, subjects, matters, and processes that we are in), and that yet other abstract concepts are construed with on (e.g. the direction of our our thoughts on some other topic, or the relation involved when we do some kind of thing on the basis of something else. However, given the limitations of this study, we still do not know what is typical of the system involving these relations. We do not know what types of concepts are typically construed by means of which spatial relation, and we do not know what types of concept are most frequently construed by means of a given spatial relation. Moreover, we do not know what types of concepts that pertain to real world physical entities are typically construed by each spatial relation. This is the focus of the present study.
At the close of the project, follow-ups studies will be needed on other bodies of text to validate our findings. This would also improve our understanding of how the genre influences the choice of preposition. Aside from the main line of work, there are several new and interesting directions for further investigation. One is a comparative study between prepositional patterns in different languages. Such a study would give insight into how metaphors develop in different linguistic frameworks. Moreover, it would be instructive to conduct an experimental study involving brain imaging, event-related potential (ERP), or other kinds of brain activity recordings, to shed light on how we process spatial statements.

**References**


Aphasia Rehabilitation Agent: towards personalized aphasia rehabilitation

Marie Dornheim\textsuperscript{1}, Elisavet Kaltsouni\textsuperscript{2},

\textsuperscript{1}Department of Integrative Medical Biology, Umeå Sweden
\textsuperscript{2} Department of Computing Science, Umeå, Sweden

mado0045@student.umu.se, elisavet.kaltsouni@umu.se

People in industrialized nations are getting older than ever and average life expectancy has increased by 25 years the last century (Butler 1997), while healthcare is facing staff shortage since 1984 (Aiken, 1989). Besides the benefits of increased lifespan, the risk of age-related diseases is also heightened (Lakatta et al., 2003). The trend of the cardiovascular disease stroke has increased in past decades and the prospective stroke risk for adults below 45 years of age is predicted to significantly raise until 2030 (Giang et al., 2018). The eminent stroke calls for more intensive rehabilitation services (Dickey et al., 2010). This in turn leads to the growth of financial costs to healthcare providers since, in fact, 8.5% are responsible for healthcare costs related to stroke (Ellis et al., 2012).

Stroke-induced aphasia can yield deficits in language comprehension and production that are caused by damage to brain regions recruited for language processing (Engelter et al., 2006). Traditionally, judgment of speech production in aphasics serves as a first assessment of aphasic syndrome, discriminated by the ability of fluency. Generally, fluent aphasics can produce connected speech with relative intact sentence structure, while speech production of non-fluent aphasics is halting, effortful, and grammatically impaired. An approximate of 20% of stroke survivors are affected of chronic aphasia, namely the persistence of symptoms after 6 months post-stroke (Dijkerman, Wood, & Hewer, 1996). Intensive speech therapy has been shown to be very effective and have a positive impact on both behavioral and biological outcomes even on individuals with chronic aphasia (Cherney, Paterson, & Raymer, 2011). Neuroplasticity of the adult brain can be positively influenced by intensive training, with different brain networks shown to be engaged in long-term successful treatment compared to immediate success, which could imply strengthening or reorganization of language processing networks (Menke et al., 2009).

\textbf{State-of-the-art.} Technological research has focused on post-stroke motion impairment rehabilitation (Hondori, Khademi, Dodakian, Cramer, & Lopes, 2013; Burdea, et al., 2019). Furthermore, since social isolation is a prevalent issue with aphasia patients, as deficiencies impact their self-esteem, researchers are exploring 3D environments in supporting them in communication and confidence (Galliers, et al., 2011). Ma, Nikolova and Cook (2009) have developed a language assistant in order to enable everyday communication and interaction. So far, no other organized approach for aphasia long-term rehabilitation is documented.

In a study by Brown, Worrall, Davidson, and Howe (2010), the personal goals of aphasic individuals during and after therapy were investigated and listed. Subjects expressed frustration over the scheduled therapy sessions and the loss of control, decision making and daily planning their days according to their individual needs and preferences. In this paper, we propose an approach to address the above-mentioned considerations. The current paper’s purpose is to introduce an intelligent technology proposition for language rehabilitation., the Aphasia Rehabilitation Agent (ARA). The suggested system design is intended to help the patient to intensively train and improve their deficits with a portable home-based device that will not demand additional nursing. The patient can work independently and is exposed to individually tailored therapy in the absence of a therapist. ARA is aiming to provide an integrative tool to thoroughly train aphasia patients in their phonological deficiencies.

\textbf{System design scenario and theoretical background}

The ARA platform is suggested to follow a formal framework of a task-oriented integrated system that distinguishes different levels in symptom assessment and subsequent rehabilitation in a mixed reality environment. The system scenario constitutes an application, based on AR technology, through Microsoft HoloLens (AR)-headsets. The subject can fixate, assess, and interact with different multisensory virtual objects superimposed on it, in a mixed reality environment (Figure 1). Stroke patients suffer from brain injuries and thus movement impairments contralateral to the brain damage are often prevalent. The present design scenario is planned to consider such needs, by adapting task implementation in a manner that only one arm will be needed. An alternative is using speech recognition to control the device with voice commands. We believe that this is possible through the use of the AR headset, as the patient has the freedom to manipulate the presented stimuli at their surrounding environment, simulating real-life situations. ARA use is destined to be performed remotely from home, but should
be followed by a therapist, inspecting the individual’s progress every few weeks. The system’s input and output will be visual and auditory, and those aspects will be further specified after development and evaluation.

**Figure 1** Hololens view of a virtual object (lamp), depicting a simulation of the word-to-picture matching task to evaluate semantic access functioning (left and upper right). User rotating the lamp (lower left)

**Assessment and rehabilitation curriculum.** Assessment of the specific impairments in need of training, will be careful and thorough, since errors might seem to be related if only considering speech output, as they can be caused by damage to different modalities. The first step in successfully tracking down errors to their origin is finding out which modalities need to be trained. For this reason, the agent is based on the logogen model (Fig. 2) (based on Morton, 1979, revised by Patterson and Shewell, 1987), an approach that takes different sensual modalities into account (hearing and seeing words & object recognition). The model assumes that stimuli recruited from different modalities introduce variation of sources that are not present for other modalities (Yelland, 1994). Thus, every word has its own feature counter referred to as logogen that collects evidence until the individual threshold is achieved and the word can be recognized. If a stimulus is auditorily presented, say “house” /haʊs/, it needs to reach a certain threshold in the auditory phonological analysis level before this activation reaches the phonological input lexicon and so on. In this regard, the logogen model offers optimal preconditions for the assessment of the nature of errors. We aim to provide input representations from numerous modalities for every logogen to gradually localize the impaired path/level of the cognitive model. Finally, tasks are generated upon the assessments of the nature of errors. We aim to provide input representations from numerous modalities for every logogen to gradually localize the impaired path/level of the cognitive model. Finally, tasks are generated upon the assessments of the nature of errors. Through repetitive stimulation of the impaired connections within specific tasks, those connections within or between modalities can become more stable and eventually leading to the expected outcome: smooth lexical access and appropriate speech output. The prototype will comprise of a manageable size of inventory of stimuli to provide thorough connections to written, acoustic and figurative representations.

**ARA system and tasks.** Assessment and rehabilitation processes are based on the model architecture, through standardized test-sets. An initial stage of the assessment will be to examine the source of errors. Errors are detected by examining each level (errors of semantic/ syntactic/ phonological nature) where a “word” gets its defining feature. At this stage, the user will be introduced to ARA by performing a series of lexical tasks, where their scores will be documented and processed.

The raw scores from the assessment will be converted to variables that the algorithm will utilize in planning the personalized rehabilitation schedule. This will be comprised of tasks designed according to the impaired modality and the source of errors. Multiple agents operate in a cumulative manner, say one agent is in charge of auditory phonological analysis, another for phonological input and a third one for semantical system. That way, if the direct route to the phonological output lexicon is damaged, the agent of the semantical system tunes in and provides tasks to train access of the output lexicon via detour as well as tasks that stimulate and thus strengthen the neural underpinnings of the damaged level.

**Levels of task difficulty** are defined within each step according variables/dimensions, which have been identified to have an impact on the probability that the task will be correctly performed (Shallice, 1988). These dimensions are the following: Word length, frequency, imageability, syllables, lexical categories (e.g. Composita and
Derivation). Once the individuals have scored an 80% within each level, the agent will direct them onto the next one.

Figure 2 Patterson and Shewell’s (1987) model of lexical processing (logogen)

Intelligent System and reasoning. The system’s goal is defined as long-term aphasia rehabilitation that will steadily account for current deficiencies and following progress in adjusting the therapy curriculum, which comprises of the selected tasks and levels of tasks. ARA will follow a multi-agent system architecture, in which rational agents will generate beliefs and set goals by following a bottom-up reasoning framework. This architecture will allow the recognition of external events to be integrated in context information between different agents. The integrated and cooperated knowledge-based systems account for the agents to serve different purposes and are defined by different requirements and end-goals, by storing different input information and collectively achieve effective planning. The reasoning loop of the ARA is depicted in Figure 3.

Figure 3 System architecture

Concluding remarks

ARA is a personalized, self-improving system for post-stroke survivors, considering usability issues that concern this clinical population. The algorithms use the patient’s assessment scores in producing better predictions on the rehabilitation schedule needed. The patient’s progress is providing the system with more information to improve its predictions and task delivery, regarding tasks selection, frequency and level. The aim is to develop an agent that will render rehabilitation easily accessible, affordable and intensive, without burdening the patients with excessive healthcare costs and frequent visits therapy to therapy sessions. As future guidelines, we set a series of goals: participatory design methodology approaches are to be used, in order to involve both clinician and patient views in the interface design and hardware evaluation. It is mandatory to follow an iterative design process, so as to avoid disregarding patient needs. The initial prototype to be evaluated is planned to include basic simple tasks,
drawn from validated aphasia test batteries like PALPA, CAT & WAB-R Western Aphasia Test Battery. At the outset, every level of the logogen model will be equipped with one standardized task and the agents can exclusively operate within modalities. Gradually, more tasks, level of difficulty and interaction of agents between modalities will be added or discarded, depending on evaluation outcomes.

References


Artificial agents for fairness in economic games: a demonstration

Samantha Stedtler, Timotheus Kampik, Helena Lindgren

Umeå University

tkampik@cs.umu.se

Motivation. While autonomous systems have developed considerably in the past, a key challenge that remains to be solved is effective social interaction with human agents. In this context, an important observation is that autonomous systems are typically programmed to optimize a specific metric in a manner that resembles the notion of economic rationality, while humans do not act fully rationally, i.e., they do not comply with formal models of rational economic decision-making (Kahneman, 2003). For example, human agents often exhibit behavior that does not maximize their individual gains but that is compliant with social norms and expectations.

Thus, it seems worth investigating how autonomous agents can be equipped with the ability to act human-like, in a boundedly rational manner, for example to reach action outcomes that can be considered "fair" compromises for all affected agents (human and artificial).

A common way to evaluate human sharing behavior in the field of behavioral economics is the ultimatum game (Güth, Schmittberger, & Schwarze, 1982). In the ultimatum game, the first player (proposer) receives a certain amount of coins (in our case 100 coins), which they can split between themselves and another player (responder). The distribution of the coins is up to the proposer. The responder can either accept or reject the offer. In case they accept, each player gets the amount of coins that was proposed. If the responder rejects the offer, nobody receives any coins. In our variant of the ultimatum game, several game rounds are played: the roles of the proposer and responder are exchanged after each round. According to the rational man paradigm in neoclassical economic theory, the players should accept any amount offered to them that is above zero as they should strive to maximize their own return. However, human agents exhibit a behavior that seems to be nourished by the notion of fairness and tend to reject offers that are similar to the ones a rational agent would make (Thaler, 1988) because they see them as unfair.

Our goal is to create a learning agent that exploits different strategies to facilitate fair human behavior (i.e., the proposal of splits that are close to a 50/50 share, and the acceptance of such proposals) and converges to a maximally effective strategy over time. In this short paper, we describe the agent implementation as a preparation
of an empirical study. The learning agent is represented by a bot against which the participants play the ultimatum game. The hypothesis is that the agent could achieve this effect by emphasizing fairness as a social norm, thus facilitating fair game behavior in the participants through the feedback. Therefore, we implement an agent that autonomously tests different feedback modes (as will be explained below). For this, we use a multi-armed bandit machine (Kuleshov & Precup, 2014) learning algorithm that dynamically converges to the most effective option (feedback mode), enabling the agent to adjust its strategy immediately; i.e., each feedback mode is considered a bandit arm and the goal of the algorithm is to converge to the arm that on average causes the smallest divergence from 50/50 split proposals (and acceptance of such proposals). The approach is different from the traditional A/B testing methods (treatment group vs. control group), which would require a complex manual setup and a larger sample size.

**Demonstration.** During the study, the participants play the ultimatum game against the bot for several rounds. After a short introduction into the procedure of the game, the bot makes the first offer and the test subjects have to react, followed by more rounds with alternating roles. Finally, participants will provide basic demographic data and document their experience with economic games and with the bot. The bot is implemented as a web application (JavaScript-based). We implement four different facial expressions for the avatar that can be used as the non-verbal feedback.

The bot guides the participants through the game. Ultimately it should also react to the actions of the participants using verbal feedback, either by itself or accompanied by the aforementioned facial expressions. The different modes we want to implement in the demo are as follows: i) emotional (the agent shows emotions) vs. argumentative (the agent attempts to provide rational arguments to encourage fair behavior); ii) stingy (the agent highlights "unfair" behavior of the human) vs. fair (the agent highlights the "fair" target behavior) vs. generous (the agent highlights generosity); iii) verbal (the agent adjusts only its text feedback) vs. non-verbal (in addition, the agent provides visual clues in form of facial expressions). Figure 1 shows a screenshot of the bot.

![Figure 1: Human-agent ultimatum game: bot screenshot.](image-url)
A demonstration of the different agent modes is available at https://people.cs.umu.se/tkampik/demos/learning-agents.html.

**Simulation.** To assess the feasibility of employing a multi-armed bandit (Kuleshov & Precup, 2014) for the use case in focus, we model a multi-armed bandits to test the different agent modes and run a simulation of different possible scenarios (i.e., different ranges of disparity between the agent modes). In the simulation, we use a (simple) epsilon-greedy bandit with gradual epsilon decay. Figure 2 shows the graph of one of the simulations.

![Simulation Result](image.png)

Figure 2: Example visualization of a simulation result. Note that the regret on the horizontal axis aggregates the divergence from "fair" 50/50 sharing behavior that is caused by triggering non-optimal feedback modes.

However, more tests are necessary, in particular to assess convergence given greater variance of the human reactions to different agent modes, before the application can be deployed in a human-computer interaction study. In this context, additional multi-armed bandits algorithms should be tested as well (for example: Upper Confidence Bounds (UCB) bandits).

**References**


---

1A snapshot of the full simulation and its documentation is available at https://gist.github.com/TimKam/572d319412f4ac85556af673f3e9a000.
Preliminary findings of an empirical study into human ability to accurately perform visual estimates of blood loss

Mattias Lantz Cronqvist¹, Rachel R. Philips², Carl-Oscar Jonson², Erik Prytz³,⁴

¹Department of Computer and Information Science, Linköping University, Linköping, Sweden
²Department of Psychology, Old Dominion University, Norfolk, Virginia, USA
³Center for Disaster Medicine and Traumatology, and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

mattias.lantz@liu.se

Introduction

This paper will report on a work-in-progress study investigating how people perceive and make estimates about liquid volumes, in the particular in the context of visual estimation of blood loss due to trauma. There is a clinical relevance to these types of estimates, as it is one way to determine the extent of injuries, what interventions are appropriate (Berry, Seitz, & Payne, 2014) and if injured individuals need blood transfusions. It can also be relevant to investigation of crime scenes and accident sites (Frank et al., 2010). Blood loss estimation can be achieved using a variety of techniques including visual estimation of blood loss (VEBL), blood tests, HR/RR evaluations, or with MedTech devices such as Gauss Surgical (Algardien et al., 2009; Didy, Paine, George, & Velasco, 2004; Wilcox et al., 2017). The study reported here focuses specifically on how untrained people (laypersons) perform on a VEBL task.

Previous research shows that people are generally inaccurate in making VEBL. Specifically, a key finding is that small amounts of blood loss tend to be overestimated, and large amounts of blood loss tend to be underestimated (Feng, Michielsen, & Attinger, 2018; Frank et al., 2010; Hancock, Weeks, & Lavender, 2015). This tendency may be exacerbated by characteristics of the situation or environment such as the materials present (Simpson, Kennedy, Chew, & Lubowski, 2001), the presence of other fluids (Simpson et al., 2001), or gender if the environment includes a model (Faul, Aikman, & Sasser, 2016). However, most previous studies suffer from methodological shortcomings and confounds that affect their generalizability. There is, to date, no study that has systematically varied the relevant variables, such as the amount of blood, the appearance of the patient victim, victim gender, bleeding rate, injury appearance, among others. Additionally, precious research has examined VEBL with little attention paid to possible explanations of the empirical results based on established theories of human cognition or perception. Previous research is predominantly clinical in origin and focuses on presenting empirical results (e.g., amount of error) rather than theories explaining why the over- and underestimations occurred.

The current study is an attempt to address these previous shortcomings. First, the reported study sought to analyze VEBL using a structured approach controlling for confounds and systematically varying variables of interest. This is done, for instance, by keeping the environment the same between the different comparisons and comparing several different blood volumes and blood flow rates more so than previous research. From an empirical perspective, we hypothesized that blood volume should affect the VEBL in the same way as has been shown previously. That is, small blood losses are expected to result in overestimates and larger blood losses in underestimates. To reiterate, the data reported here are preliminary and drawn from a currently on-going study and we will not attempt to generalize the results using inferential statistics. Rather, the results presented are intended as a preliminary validation of our methodological setup, and serve as a starting point to explore the potential of various theories of human cognition or perception to explain the results.

Method

This initial pilot study consists of data from 16 students over the age of 18 (Males = 2, Females = 14, M age = 22.88, SD age = 5.22) from Old Dominion University in the United States. Participants signed up for the study using the SONA research participation system and received class credit for their participation. The study was approved by the university IRB and all APA ethical guidelines were followed.

Each participant had to make assessments of injury severity and estimate blood volumes by looking at short video clips of 5 seconds. Each video clip consisted of a male or female actor sitting on the floor bleeding from his/her leg (see Figure 1). The amount of blood present (0, 50, 100, 150, 200, 300, 400, 500, 700, 900, 1100, 1500, or 1900 ml) and the flow of blood (80, 200, 400 ml/min) varied between videos. To provide realistic bleeding conditions, with regards to color and viscosity, porcine blood was used. Both actors had their legs in the same angle on the floor and their leg length was comparable.
To collect data, a System for Acquiring Blood Loss Estimates (SABLE) was developed. Upon entry to the lab, each participant signed an informed consent and was provided oral instructions. After being logged in to SABLE, each participant completed the study as diagrammed below in Figure 2. The participant first received study information and instructions. When the start button was pressed a video was shown for 5 seconds. During this time period the participant was asked to decide if the bleeding was life-threatening or not using the keyboard. After that, more estimations, evaluations and decisions about the previously shown video were collected, but with no time limit (this data is not reported currently). This process was repeated until all 78 videos had been shown in random order. After the video portion of the experiment, each participant completed a demographics questionnaire in which they provided gender, age, and information about previous experiences related to the study. When the experiment was completed on SABLE, participants were thanked for their participation and provided a copy of the debrief and contact information for the researcher to take with them.

**Preliminary results**

The visually estimated blood loss compared to the actual blood loss can be seen in Figure 3. The actual blood loss (i.e., blood on the floor) is shown on the horizontal axis and participant estimates on the vertical axis. As hypothesized, small volumes, were overestimated. However, at around 200 ml of actual blood loss, the participants started to underestimate the amount of blood present. The amount of estimation error increased as the volume of blood increased.
Discussion

The empirical findings of previous research appear to be replicated in this limited subset of 16 participants, out of a planned total sample size of 60. Small volumes were overestimated, and large volumes underestimated. These preliminary findings suggest that the experimental setup is sound, and that the resulting final data set will allow fine-grained analyses on the magnitude of error given volume of blood loss and potentially also rate of blood flow and patient gender. Future analyses on the entire data set will also include subgroup analyses, e.g. by gender or previous professional experience. Further, with the current experimental design each participant sees each blood volume a total of eight times, with different flow rates and gender. This gives averaged estimates for each blood volume, in contrast to the single estimates used in the majority of previous studies, which should further increase the validity of the results.

An issue in the past research is the lack of theory-driven experiments. Currently, no study or series of studies have attempted to use theories of human cognition or perception to guide exploration or to explain their findings. We claim that this is a shortcoming that limits the usefulness of past research, and that the focus on empirical findings only has resulted in multiple replications of obvious effects, such as the over- and underestimation effect shown here, without necessarily providing a systematic exploration of the relevant problem space. Thus, we are currently in the process of exploring suitable theories that could provide useful insights and testable hypotheses related to factors relevant to VEBL. It is our hope that the SweCog community will contribute with valuable suggestions for theories that could be potentially usable toward this end.

References


The grammatical gender of Swedish nouns is a mystery. While there are few rules that can indicate the gender with some certainty, it does in general not depend on either meaning or the structure of the word. In this work we demonstrate the surprising fact that grammatical gender for Swedish nouns can be predicted with high accuracy using a recurrent neural network (RNN) working on the raw character sequence of the word, without using any contextual information.

There are two different grammatical genders for Swedish nouns, common (utrum) and neuter (neutrum) (Rebbe, Gullberg & Ivan, 1954). Swedish used to have more than two genders; however the masculine and the feminine later merged into the common gender and modern Swedish makes no difference between them. The grammatical gender affects the indefinite article (en boll -- ett bord; English: a ball -- a table), as well as the definite article and the definite suffix. The Institute for Language and Folklore, a Swedish government agency, states that there is no unequivocal way of determining the grammatical gender of a word. Further the Swedish Academy states that the grammatical gender in Swedish is a lexical property which is to a large extent not dependent on neither meaning nor the structure of the noun (Teleman & Hellberg, 1999). There are exceptions: living things are generally utrum (the more common class), and there are a few specific suffixes that are reasonably good predictors.

However, determining the gender of a word is generally considered a major difficulty for non-native Swedish speakers learning the language, and it is an open question how predictable it is.

In this paper, we investigate the predictability of grammatical gender for Swedish nouns without any context words. We train a recurrent neural network (RNN) model to predict the property given only the raw character sequence as input. The model quite easily learns the necessary patterns and achieves a 95% accuracy on the test set, which should be compared to the 71% majority class baseline.

In related work, (Nastase & Popescu, 2009) trained a kernel ridge regression (KRR) model to predict the gender of nouns in German and Romanian to a high precision. It is however well known that Romanian has phonological patterns that indicate gender, while German is more similar to Swedish. Previous work has been done for predicting the grammatical gender of Swedish nouns using neural networks (Basirat & Tang, 2018). However, here the authors used word embeddings with context information from a corpus and achieve a high classification accuracy of 93.7% on their test data. Our work differs from theirs since we do not depend on any such corpus; our models act solely on raw character sequences.

A recurrent neural network (RNN) is a class of artificial neural networks that can model a sequence of arbitrary length, using weight sharing between each position in the sequence. Unrolling the RNN results in a feed-forward neural network. In the basic RNN variant, the transition function at time step t is a linear transformation of the hidden state at time step t-1 and the input, followed by a point-wise non-linearity.

RNNs have problems with learning global dependencies in a sequence (Hochreiter, 1998; Bengio et al., 1994), and thus Long short-term memory (LSTM) architectures were invented (Hochreiter & Schmidhuber, 1997). An LSTM is a gated variant of the RNN where the cell at each time step contains an internal memory vector and three gates controlling what parts of the internal memory will be kept (the forget gate), what parts of the input that will be stored in the internal memory will be kept (the input gate), as well as what will be included in the output (the output gate). Another architecture was later proposed containing only two different gates, called the gated recurrent unit (GRU) (Cho et al., 2014).
We train, evaluate and compare three different models: a feed-forward network, an RNN-LSTM network and an RNN-GRU network and show that it is possible to solve the problem of grammatical gender prediction using these models. The implemented networks consist of one hidden layer with 64 units, with a sigmoid activation function in the output layer. We train the models using binary cross-entropy loss. The dataset used to train and evaluate our models consists of 88,480 Swedish nouns acquired from SALDO (Borin, 2013), labelled with grammatical gender. In Swedish, about two thirds of nouns are utrum and one third neutrum. In our dataset, 71% of the words belong to the utrum class, and 29% to the neutrum. Further, the Swedish Academy notes that there are some common suffixes that strongly correlate with the grammatical gender. Words ending with ing, tion, het and ist are mostly associated with utrum and words ending with eri, skop and gram are associated with neutrum.

A batch size of 32 was used in our experiment for all models. Furthermore, all models were trained until no change was observed in the validation loss for 50 epochs. The models are evaluated using prediction accuracy, precision, recall and F1-score. The dataset was randomly split up into 60% training, 20% validation and 20% test. A second test set (13% of the dataset) was created from the first test set, where we removed all common suffixes that correlate highly with a gender. We can thus evaluate how well a model has learned to predict the grammatical gender of a word without suffixes to its disposal.

![Figure 1: The proposed character based RNN model.](image1)

![Figure 2: t-SNE visualisation of the output from the LSTM layer on test data.](image2)
In table 1 the results of the three models are presented on the test data set. We see that all models beat the 71% majority class baseline. The best performing model was the LSTM-RNN with a test classification accuracy of 95.15%. Further, this model achieved a 0.93 (0.89) and 0.96 (0.94) precision score, a 0.90 (0.85) and 0.97 (0.96) recall score and a 0.92 (0.87) and 0.97 (0.95) F1-score on the neutrum and utrum classes, respectively. The parenthesis values are calculated for the test set with removed words. In figure 2, a t-SNE visualisation of the representations of the LSTM is visualised and it is clear that the boundary in the middle between the classes is where most model mispredictions happen.

<table>
<thead>
<tr>
<th>Model</th>
<th>Test set</th>
<th>Test set w/ removed words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feed-forward</td>
<td>0.8548</td>
<td>0.8577</td>
</tr>
<tr>
<td>GRU</td>
<td>0.9492</td>
<td>0.9255</td>
</tr>
<tr>
<td>LSTM</td>
<td><strong>0.9515</strong></td>
<td><strong>0.9324</strong></td>
</tr>
</tbody>
</table>

*Table 1: Test accuracy of the models.*

Our results show that a simple feed-forward neural network achieves around 85% test accuracy, above the naive majority class baseline of 70%. A simple RNN-based model is able to predict grammatical gender in Swedish nouns (a lexical property that is considered difficult and does not adhere to rules other than in a few special cases) with a high test accuracy of 95%. Further, removing common suffixes from the test set does not impair the results markedly. Our results show that we can learn the necessary patterns only by looking at the sequence of characters in a given word, without using any context words.

**References**


Lars Oestreicher

1Visual information and interaction unit, Department of Information Technology, Uppsala university
LarsOe@it.uu.se

Children with severe impairments affecting both motor skills and the ability to communicate are exceptionally dependent on people in their environment for their everyday life and survival. However, when there are no communication channels available, their situation becomes even more exposed and vulnerable. The project described here is aimed towards this target group of children with a locked-in syndrome (Bruno & Nizzi, 2016; Smith & Delargy, 2005). We use EEG-reading devices in combination with neural networks in an attempt to create a simple and affordable communication system, that will allow these children to communicate at least the basic needs and ideas to their environment. For this pilot study we have, however, only used informants that are non-impaired. If we succeed in the current approach, we hope to be able to take this further and include people from the target user group.

We hope that we shall be able to develop a communication means between a person with a “locked-in” condition and people in his or her environment (Mohanchandra, Saha, & Lingaraju, 2015). The hope is that by recording an individual’s EEG-patterns for a given set of pictures, it should be possible to use the CNN model in real time to see which of the pictures the person is currently focusing on. The initial attempts have used simple pictures, which in a later implementation can be replaced by, e.g., pictograms that show the intention of the person using it (wants to eat, has pain, wants to sleep, etc.) Ideally this type of communication could even function as a general communication method in those cases where the person has a fully functional cognitive state, but have lost the normal means of communication due, for example, to higher level paralysis, or a stroke.

There are many projects working with Brain-Computer Interfaces (BCI), and there is also a very rapid development within this area, with results that were not even possible to imagine just a few years ago. According to Jeunet et al. (2016) there are currently two main approaches to Brain-Computer interfaces (BCI): the operant conditioning approach and the machine learning approach. The former is based on a biofeedback mechanism, where the user’s brain needs to learn how to send the correct signals, based on the feedback of the system; the latter approach requires the computer to detect patterns in the signals that can then be converted into commands or other types of interaction. Apart from this there is also research concerned with neuroimaging, i.e., investigating the representation of concepts in the brain, where there is work indicating the existence of semantic fields in the brain, mostly using fMRI (see, e.g., Huth, De Heer, Griffiths, Theunissen, & Gallant, 2016).

Systems based on capturing nerve signals are today already used, for example, to enable people to regain use of artificial limbs after an amputation, or for the actual maneuvering of technical devices, such as wheelchair chairs, using only the signals from the nervous system as controllers. There are also examples from military applications where a “pilot” can guide a plane along a flying path without using any physical controls (though still only in a specially designed flight simulator). Clearly, systems using a BCI show promising results in many areas. The project described in this article has focused on the communication aspect of the use of BCI-systems, which has since some time now shown to be possible with BCI systems (Kübler & Mattia, 2016). Within this area there is now a large amount of research, not least into synthesized speech, either as controlled by BCI systems (Soman & Murthy, 2015), or speech induced directly from the brain patterns (Cooney, Folli, & Coyle, 2018). The latter approach, while interesting, uses invasive electrodes which require surgical operations to be applied. This, of course, gives a higher accuracy than scalp-based EEG-systems, but also restricts the possibilities to make it into a commonly used system. In this project we have settled for using only scalp-based EEG-reading equipment, currently with dry electrodes rather than wet. We are currently in the process of transferring the software to a research system with 64 channels using active electrodes.

Most people in the target user group of this project has been in this locked-up situation since birth, which means that it is generally unclear in which cognitive state they are. This also means that we may not infer that they can use communication devices based on natural language expressions. Most of these children can at best understand
the meaning of pictorial language, using simple pictures as expressions for activities. This means that we need to bypass the use of standard linguistic expressions, and focus on the non-language based communication assets as a starting point. After that, we may try to introduce some language aspect into the system. Still the focus is on children with very small linguistic capacity. This is especially important for children who have been in a locked-in state since early age.

Currently, we are using a prototype setup, using locally developed software for the recording of the EEG-data while at the same time showing visual stimuli to the informants. The EEG hardware used is a dry-electrode equipment from the openBCI foundation\(^1\), with a 16-channel recording capacity (Cyton + Daisy board) and a recording frequency of 125Hz. The presentation/recording software has been developed in Python, and the data is analyzed using a Convolutional Neural Network (CNN), also developed in Python, using Tensorflow using the Keras wrapper. The EEG-recordings are made on a Mac PowerBook, and the neural network is run under Linux on an HP machine, Intel Core i7-7800X CPU and a GeForce RTX 2080 Ti graphics board as GPU.

The recordings are made with the informant sitting in front of a display screen, where visual stimuli are displayed for 3 seconds per image. Simultaneously, the raw EEG-data from the sixteen electrodes distributed over the entire skull is recorded and continuously tagged with the picture number and some additional data. Each session lasts for about 15 minutes, and the device records the data with a frequency of approximately 125Hz. Each recording is labeled with the recording date and time, and the selected configuration of the recording system. With this setup we have now recorded and analyzed EEG-recordings with a small number of volunteer students aged between 19 and 32 years old. Most of the recordings have been single session recordings, but for some subjects we have collected recordings at several different occasions, in order to see whether the patterns detected are consistent over time.

The images displayed as visual stimuli are simple and easily recognizable, e.g., a bicycle, a strawberry, a football, etc. Additionally, we have used one set of stimuli consisting of the alphabet in order to see whether they were providing enough differences to be used in this manner. Finally, at the end of the first set of trials we added the corresponding word written in the native language of the informant.

The analysis has been made in three different ways: one initial visual inspection of the data transferred into heat maps; one analysis of the data from the simple setup with only pictures; and one analysis where a picture and the corresponding word have been placed in the same prediction category. The results from these three analysis steps have been surprisingly strong as will be detailed below.

The visual inspection of heatmaps generated from the raw data rows turned out to be different for the different pictures, which was expected. However, for the same picture, the heatmaps were almost identical throughout the whole session, i.e., there was a clear visual difference between the data for different pictures (see figure 1). This initial visual analysis indicated that there was something in the patterns to be found through further analysis.

The second phase of the analysis involved using a convolutional neural network (CNN), which was based on one of the most common models used for image recognition with the MNIST data. The raw data was reshaped into 4x4 matrices, where the matrix positions (schematically) represent the actual position of the electrodes on the

---

\(1\) This is an open source research class EEG-equipment, which is not intended for medical purposes. For details about this equipment, see http://www.openBCI.org/
head. This made the model construction a bit difficult, since the “image” was so small. However, once this problem was solved the results were “too good to be true”. The first models reached around 80% accuracy after 2000 epochs. Essentially, this meant that for 8 out of 10 raw data samples, the system would predict which picture a person would look at in a real time usage situation.

Since this seemed too good to be true we arranged a “debunking seminar” where people were instructed to look at the software, the procedures and the data in order to try to find something wrong with the methods. Fortunately, we did not find anything that was wrong with the setup. Building on a gradual refinement of these models we have now reached a best accuracy level of 92% (albeit after 20 000 epochs) and the current model does reproduce these results for most of the raw data collections we run through it. As can be seen in figure 2 (right) the training accuracy is actually lower than the validation accuracy, which is probably due to a very high dropout values in the model. The confusion matrix (left) shows which images that were confused with each other. The two pictures most often confused here are a picture of a hammer and a picture of a tennis racket. It is quite reasonable to see why these concept could be confused from a semantic, conceptual perspective.

The final analysis was of the combination of pictures and words. When the data from the sessions with both pictures and words was analyzed through the same, refined model, we still reached an accuracy level of slightly less than 80%. This was totally unexpected, but when repeated it showed approximately the same level of accuracy. We have, at the writing of the article, not managed to increase this number in the same way as with the pictures only-session. This might have to do with the fact that we did not increase the length of the recording session accordingly, despite having twice as many stimuli.

However, with the repeatability factor. For the current runs we have used training, validation and test data stemming from the same recording session, i.e., data gathered from the same informant, nor between data from two different informants. The former of these difficulties we think originates in the problem of getting the rigid EEG-cap to fit properly on the skull of a person. The second might be less of a difficulty in itself, but is most likely rather a property of the pattern we are trying to detect. Initially, the idea was to investigate whether it was possible to identify the picture that the person was looking at from the EEG patterns. Now, the analysis of the later recordings, where we use both the picture and the corresponding word, goes against this, since they present completely different visual stimuli. Instead, this part of the experiment instead indicates that we can identify the concept that is triggered by the stimuli, i.e., the chunk that is formed as a neural pattern for a certain piece of knowledge. In a study using fMRI, Huth et al (2016) managed to map semantic fields onto different areas of the brain, and this supports our current findings quite strongly. However, it also
implies that it might be possible to enhance the results further by using the semantic fields more extensively when it comes to the selection of stimuli and the placement of electrodes.

This is also consistent with the problem of finding interpersonal predictions from the patterns, since the knowledge chunks\(^2\) have been formed from a person’s experiences and background knowledge (although the study by Kübler and Mattia (ibid.) found that the distribution of the semantic fields was remarkably similar over the test subjects). This means that there is a very low likelihood that two people should display the same brain activity for a certain knowledge chunk. However, this can also be seen as positive, since it will decrease the risk of a “Big Brother sees you” situation. To conclude, we think that the results from this small study strongly indicates that it will be possible to develop a communication device based on these principles.

When it comes to the applicability of these results to the intended target group, it is still too early to draw any definite conclusions from our results. There are earlier studies that suggest that non-invasive EEG will not be sufficient for this kind of communication (Chaudhary, Xia, Silvoni, Cohen, & Birbaumer, 2017) but they still had some success using functional near infrared spectroscopy. However, if we could manage to make the predictions reliable over several sessions, we might have a possible way to enable people in a locked-in condition (and people with severe communication problems in general) to reach out of the isolated condition where they now reside also in a mobile setting. It might also be of help to people with ocular muscular problems, who have large problems using eye-tracking systems, which are otherwise very good communication devices.

References


\(^2\) We adhere to psychological concept of a “chunk” as a singular self-contained piece of knowledge, which has a meaning of its own (cf. Anderson, 2009)
Are People Ready for Social Robots in Public Spaces?

Sofia Thunberg\(^1\), Tom Ziemke\(^2\)

\(^1\)Department of Computer & Information Science, Linköping University
\(^2\)Department of Computer & Information Science, Linköping University

[sofia.thunberg, tom.ziemke]@liu.se

Introduction

The quickly growing passenger volume in train traffic poses a serious challenge for all train carriers and station operators. With the increasing number of passengers arriving and departing at a train station, the probability of delays and missed connection trains grows accordingly. Sweden is also facing a large number of first-time passengers, such as tourists, immigrants and people changing their travelling routines with the environment in mind, people with little knowledge of foreign languages, or those who need any kind of special attention. At the same time, the day to day customer service has moved out on the internet. Therefore, the customers that enter the SJ support shop at the Stockholm Central station are usually not the webpage and app users, and they want to talk to someone instead.

Recent developments in robotics are potentially changing the nature of service (Doorn et. al., 2017), and research in human-robot interaction has previously shown that humanoid robots could possibly work in public spaces, e.g. airports (Triebel et. al., 2016). With this background, an ethnographic study was conducted with the humanoid robot Pepper. The research question was: What expectations do people have regarding Pepper’s interactive abilities?

Method

The study took place in the customer support shop at the central station in Stockholm. As can be seen in Figure 1, the robot had an SJ logo on its display and an SJ employee scarf around its neck. Beside the robot, a roll-up about the study was placed (in English: “SJ in cooperation with Linköping University conduct a study in social robotics. Welcome to try the robot!”). The participatory observation was conducted during five occasions, six hours each, with a total of 30 hours during October 2018. There were hundreds of people interacting with the robot and more moving inside the store during this time. During the sessions, the observer (the first author) sat on the side of the robot against the wall. She also interacted with the participants at times and carried out interviews with open-ended questions about the robot and interaction, e.g. "What do you think about the robot?" and "What do you want the robot to be able to do?". The notes from the observations were open coded, which means that one writes keywords in the margin to mark what is important. In the next step the keywords were put together to identify themes. After creating these themes, the authors did a focused coding, by going through the material again using more general codes and only those that belongs to one of the themes. All quotes from the observation are here freely translated from Swedish to English.

SJ has a chatbot application for the Google Assistant, where customers can ask questions about trains, platforms and also get to know fun facts about SJ. This application is developed in Google DialogFlow and was also used for the Peppper robot. But some of the standard questions in the web application would be confusing if the robot answered, e.g. if a person asked “Who are you?” and the robot answered “I am SJ”. To correct this, another natural language process was developed by Dyno Robotics (a Linköping based robotics company), also in DialogFlow, that overrides some of the questions with a more suitable answer. For example, “I am the robot Pepper” for the above question. In the second chatbot, some behaviours were added, e.g. that the robot could perform a small dance and laugh at jokes. The first author of this study had the role of asking for and adding abilities to the robot, and drawing from abilities previously developed for the robot.

In the general system the robot moved its arms in a random order while talking in order to simulate natural talking behavior imitating human gestures, and it followed faces. The robot was listening at all times (besides when it...
spoke) and dialog was initiated if a person said something that matched what the chatbot knew. The robot did not initiate interactions by itself. When a person talked to the robot it took a few seconds for the system to analyze what was said and how to answer. Because of the parallel chatbots, this process took about twice as much time as it should. During this time the robot showed different modes in the lower part of its eyes. When the robot was listening, they turned blue and when the robot analysed what had been said the eyes were pink. This was implemented to indicate to the user that the robot is doing something (instead of, for example, a progress bar). When nothing happened or when the robot spoke, the eyes were neutral. The system was programmed in Python and C++, and the software stack was based on the Robot Operating System (Quigley, 2009).

Results

The results from the observation are here divided into four main themes, Unaccustomed, Concern and Fear, Emotional connection and Entertainment, which will be explained and exemplified below.

Unaccustomed The most common thing that happened during the study was that a person came in to the store, approached the robot, pointed on the tablet on the robot’s chest and went away. On the tablet there was an SJ logo and nothing happened when it was pressed. We would like to call this the pointer finger syndrome. On several occasions, it seemed like an obsession. For example, a woman had just interacted with the robot and had a conversation with the observer. While they were talking she tried once again to touch the tablet and nothing happened – "it's twitching a little [in the finger]" she said and explained that she had a hard time not pressing it. There were also a lot of people who thought that the robot was a queue ticket machine and that the tablet should be used to get a ticket, or that one should twist the arm a little to get a ticket. Most people did actually not get that it was a robot, just a broken machine. Even if they looked at the roll up on the side the message was not understood, and they did not seem to reflect on why the queue ticket machine had head and arms. Even when the robot was dancing and music was coming from it people went up to it and tried to get a ticket from it. This behaviour increased during stressful times, when the train was late or during commute hours. One explanation could be the physical form of the Pepper robot and that the placement of the tablet on the chest affords pressing it. Also, to some degree, it could be explained by the fact that people do not really reflect on their surroundings and are focused on getting what they want out of the situation they are in. But it can also occur because of the element of surprise, i.e. few people expect to see a robot in the customer support shop.

Moreover, in Swedish society, there so far are very few voice assistants. These are not common in the home environment or at work places. During the study, Google released the first voice assistant in Swedish and apart from that, the only one that is somewhat used is Siri on Apple products. This had an effect on the interaction since people had a hard time understanding what to do with the robot. Many people would just go up to the robot, stare at it and say "it doesn't do anything". Then the observer, or someone from SJ, would say "you can try talking to it", and this fact made people very perplexed. Then, if they tried saying something and got proof that the robot could hear them and reply, they quickly adapted and kept up the conversation. There were also problems with how they acted around and toward a robot. Some people were addressing it as a "machine" and thought "that you cannot talk to a machine". When they were talking to the robot, some had a hard time understanding how such a system works, i.e.that you have to say something comprehensive and then wait for a reaction before you can say something more. When people talked about non-related things, or asked questions with incomplete sentences, the robot had a hard time answering satisfactorily. For example, when a traveller said "I need to go to town to take the train to Eskilstuna" and the robot answers "I did not understand you now, can I help you with a question about the train?", the person got agitated and told the robot "bye, you don't understand anything". The person had not asked a question about the train, but still got upset about the “malfunction” of the robot. This could be a sign of digital exclusion (Umecon, 2016). Umecon defines digital exclusion as a lack of internet, equipment, knowledge, motivation and ability, coupled with digital tools. From recent numbers, 1.1 million people in Sweden live in digital exclusion (Internetstiftelsen, 2018). But we want to broaden this definition. A digital exclusion is not only a total lack of connection to digital tools, it is also a lack of knowledge how digital tools work and how to interact with them. It is a kind of gap between cause and effect. Further, one older man noted the colours in the eyes of the robot and said to his younger friend "one can see that the robot is blushing while it is thinking, there is probably much of these kinds of things that we don't know yet but we will learn". This also indicates that people are not used to talk to machines, but that they believe they will adapt later on when robots are more common.

Concern and Fear People were a bit hesitant to approach the robot and expressed a concern about how the robot is made and how it works, and especially "who is controlling it?". For example, a young woman stood in the door to the customer support shop and asked the observer what the robot can do, and she was told that she could try to ask the robot a question about the train. She then looked scared and said "No, ew, I don't dare to do that. Is somebody controlling it? Promise me that you are not controlling it!" she said and the observer explained a bit how programming works. She then continued saying "No shit, I do not dare go up to it, but it is great to have robots
here. Then you can ask it questions.” and she left. So, she saw an area of use for the robot, but she was scared how the robot worked, if somebody was controlling it, and maybe worried her personal data would be stored. One reason for this could be that people are getting more aware of internet security problems, connected with integrity. For example, people expressed that the robot could be hacked and that it would not be an option to have a robot at home because it would be able to record you. People also had a fear that the robot could get physical with you. For example, follow you out of the store: “What if the robot starts following us now” and “Don't talk to it, it might start following us”.

**Emotional Connection** Some people expressed an emotional connection to the robot. For example, saying “This poor thing looks like a vacuum cleaner”, “We are standing here and harassing this poor thing”, “This poor thing has to stand here and work all day” and “Look, the poor thing really wants to answer the questions (but is failing)”. It is already well established that people tend to anthropomorphise robots and treat them as social entities (Bartneck, Verbunt, Mubin, & Al Mahmud, 2007; Short, Hart, Vu, & Scassellati, 2010). This means that people attribute them with both social rights and responsibilities. The reason for people addressing the robot as a "poor thing" could be multiple. One interpretation is that the robot is rather short (120 centimetres) and has a child-like voice. Some people mentioned that the robot looked and sounded like a child, i.e. the association to children and child labour may be close at hand.

**Entertainment** Some people thought that the robot's main use was for entertainment purposes. There were many people coming in, thinking it was a funny thing, and enjoying when it danced or told a joke. It also seemed like a successful family attraction where many parents and their children tried the robot, danced with the robot and it seemed like the children could be fully occupied having fun with the robot for hours.

**Conclusion**

This ethnographic study has evaluated the human-robot interaction between travellers and the Pepper robot in SJ's customer support shop at the central station in Stockholm. The results show that people are not yet accustomed to talking to robots, and a reason for this might be that voice assistant tools are not common yet. People seem to expect that the robot does not talk, that it gives out queue tickets, and that one should interact with it by using the tablet on the robot’s chest. The (somewhat unsurprising) conclusion of the study is that the robot model has a limited practical value at this point, and that it is not ready to have a working position. But it has a potential high value for marketing purposes and entertainment, so while other applications are developed and evaluated, the robot might be useful for entertaining people in public areas.

**Acknowledgement**

This study was conducted in cooperation with SJ.

**References**


Probability, Paradoxes and Human Thinking

Priyantha Wijayatunga

1Department of Statistics, Umeå University, Umeå, Sweden

priyantha.wijayatunga@umu.se

Abstract

Probability, related calculations and its interpretations are sometimes hard for people to grasp. This may be due to unreasonable or counterintuitive situations that they find in them. Here I take few probability and statistical paradoxes and discuss how people sometimes find them unreasonable, counterintuitive, etc. Often the problems and confusions are solved when the probabilities are interpreted correctly.

Introduction

One thing that many of us get sometimes wrong is probability. It is said that the probability is hard because sometimes people find unreasonable or counterintuitive answers in probability related problems. However, the theory of the probability along with statistics is very important for us to learn about most of the patterns found in our environment. Science is benefited much from the probability and statistics, as they say, “the correct use of statistics is not just good for science but essential.” So, it is important for us to understand why probability calculations on real world contexts sometimes go wrong, especially when the people think in wrong directions. Here, I go through few examples where people sometimes go wrong or think strangely or at least find probability-based calculations counterintuitive. These examples show how people grasp certain situations when they need to do inferences on them. First, I discuss the famous Simpson’s paradox which is regarded as an indecisive situation by many. Then I discuss another case of confusion called Monty Hall problem. And finally, I give some comments on so-called Allais paradox. These examples show that human mind is not a probabilistic machine and often is ignorant about the mathematical structure of the problem or the context.

Probability Paradox and Problems

One famous situation is so-called Simpson’s paradox (Wijayatunga 2014 and references therein) that is an indecisive situation for many. Here is one example of it. For headache, a doctor found that a certain medicine was used by the people. He observed that when it was taken by 10 males 7 of them got cured (70% of success) whereas for 30 females only 9 of them got cured (30% success). And, of 30 males who did not take it 18 got cured (60% of success) whereas, of 10 females who did not take it, only 2 got cured (20% of success). So, the doctor noticed that the medicine was good headache for both males and females separately. However, when he looked at them without taking their gender into consideration the success rate for those who took the medicine was 40% (16 out of 40) and that of non-takers was 50% (20 out of 40). So, now he had to conclude that the medicine was not effective for the headache for the people. Therefore, he was puzzled!

Why was he confused? When the rates are mentioned as percentages or fractions, people often believe that average of such figures is always larger than that of others when former figures are larger (in this case 70% and 30%) than their respective counterparts of the latter (in this case 60% and 20%). What people miss here is that the average is a weighted one that is determined by the counts that give rise to those percentages/fractions. In this case, weighted average of 70% and 30% is 40% and that of 60% and 20% is 50% where two weight-sets are different. That is, the larger pair (former two) has a smaller weighted average than that of the smaller pair (latter two). Since percentages or fraction are standardized figures, people tend to believe that their “aggregations” that correspond to respective pooled groups should always follow the pattern of their component groups. For groups,
when standardized figures (percentages/fractions) are assigned, information of the group sizes are lost in people’s mind when they try to assign aggregated standardized figures to the pooled groups.

In fact, the reason of having the weighted averages in this way here is that the males tend not to take the medicine whereas the females tend to take it for headache even though generally the males have higher curing rate than that of the females. So, sex is a confounding factor in the causal relationship of taking the medicine and getting cured. Therefore, the inference that, for people the medicine is not effective for headache, that is a causal query is a confounded one. The correct inference is the sex-wise one, if sex is the only confounding factor. This is due to the fact that the causal relationship between sex and taking the medicine in the forward direction.

As mentioned in Robert 2016, people think that getting all heads or all tails (consecutively five times) is less likely than getting, for instance, tails-tails-heads-tails-heads (if you toss a fair coin consecutively 5 times)! Why do people think so, even if these events have equal probabilities that are obtained under the assumption of independence of tosses of a fair coin? Does the human mind use some other information apart from independence of outcomes of tosses of a fair coin in this case? The author of the above reference says that human mind is wrong based on the independence assumption. However, if it is a fair coin, it is more likely to give outcomes that show its fairness every instance it is being tossed for certain number of times, say, five times in this case. So, with five tosses it is more likely that heads and tails are shown together rather than either just heads or tails. But it is only when order of outcomes is ignored. Human mind is mindful a mixture of heads and tails when a fair coin is tossed, therefore it is quick to decide as above. However, it is not mindful about the order of the outcomes that is of interest here, i.e., when the order of outcomes is considered any distinct sequence of outcomes of same length has same chance of happening. Note that there are only two things here; primarily independence tosses of a fair coin and secondly, the order of the outcomes is important. However, when deciding the chance/probability the first assumption overshadows the second in the human mind, thus decided wrongly.

In the same article mentioned above, it is discussed so-called Monty Hall problem (or paradox) that is said to fool even trained minds. Suppose you are on a game show. And you are given choice of three doors to select one, where behind one of the doors there is a car and behind the other two there are two goats, one goat each. You select one of the three doors (say, Door 1). The game host then reveals one non-selected door (say, Door 3) which does not contain the car. At this point, you choose whether to stick with your original choice (i.e. Door 1), or switch to the remaining door (i.e. Door 2). What are the probabilities that you will win the car if you stick, versus if you switch? Most people believe, upon first hearing this problem, that the car is equally likely to be behind either of the two unopened doors, so the probability of winning is 1/2 regardless of whether you stick or switch. However, in fact the probabilities of winning are 1/3 if you stick, and 2/3 if you switch.

So, the confusion is between 1/2 and 2/3. Now, let us see how these two numbers come. Let C denotes the event of selection of the door that has the car, G1 and G2 denote the selections of the doors that have the goat 1 and the goat 2, respectively, and S denotes the event of switching (S’ denotes sticking). Let the ordered sequence CG1 denotes the event of selection of the door with the car first and then the door with the goat 1, and so on. Then, when switching is done, the probability of selecting the door with the goat 1 first, and then selecting the door with the car is

\[ P(G_1|C,S) = P(G_1|S)P(C|G_1,S) = P(G_1)P(C|G_1,S) = (1/3)(1) = 1/3. \]

And similarly, we have the other probability

\[ P(G_2|C,S) = P(G_2|S)P(C|G_2,S) = P(G_2)P(C|G_2,S) = (1/3)(1) = 1/3. \]

Therefore, the probability of winning the car, if switch, is

\[ P(C|S') = P(G_1|C,S) + P(G_2|C,S) = 2/3. \]

And simply, if stick, selecting the car is

\[ P(C|S) = 1/3. \]

Now, if switching is done randomly that is implicitly assumed here, i.e., \( P(S) = P(S') = 1/2 \), then we get the probability of winning the car on average is

\[ P(C) = P(S)P(C|S) + P(S')P(C|S') = (1/2)(2/3) + (1/2)(1/3) = 1/2, \]

i.e., the probability of finding the car is 1/2, on average (switched or not). But as we have seen, the probability is 2/3 if you switch and 1/3 if you do not switch. Note that the two probabilities, 1/2 and 2/3 refer two different events. The confusion comes up when these two well-defined probabilities are interpreted as if they mean the same! The value 1/2 corresponds to an unconditional event of winning the car (i.e., \( C \)) on average (switching or sticking) and the value 2/3 corresponds to a conditional event of winning the car if switch (i.e., \( C|S' \)). Similarly, 1/3...
corresponds to a conditional event of winning the car if stick (i.e., C|S'). So, Monty Hall problem is on identifying what those probabilities mean.

Saen et al. (2018) present a systematic review of literature on the problem published between January 2000 and February 2018 addressing why humans systematically fail to react optimally to the problem or fail to understand it. Similar research appears elsewhere. But to my best of knowledge none of the articles claim that the problem the human mind has is differentiating between the conditional event and the marginal event, that is shown above. Note that the first step of mathematical way to solve the problem that human mind faces is to find interpretations to two answers 1/2 and 2/3 (alternatively, 1/2 and 1/3). And also, it is important to note that, the value 1/2 is obtained as the average probability of winning the car due to the fact that switching is done randomly and two distinct probabilities of winning the car if switch (2/3) and if stick (1/3) sum up to 1, that is not happening necessarily in many other similar contexts.

Another context of confusion is with decision-making under uncertainty that is not unfamiliar to humans in their everyday life activities. But some of the well-formulated decision-making theories are contradicted by human, for e.g. so-called the principle of maximum expected utility (Neumann and Morgenstern 1953). The principle says that a decision maker chooses between risky or uncertain prospects according to their expected utility values where the expected utility value of an uncertain prospect is defined as the weighted sum of utility values of all possible outcomes of it such that the weights are their respective probability values. As an example, and of course a counter example, consider the so-called Allais paradox (Belavkin, 2006); suppose you have two lotteries L and M where L is such that you can win $3000 with probability of 1/3 or win nothing with probability 2/3 and M is such that you have a sure win of $1000. And for simplicity assume that the utility value is the same as the winning amount. Then they both have same expected utility of $1000, so, no lottery should be preferred over the other according to the principle. However most of the subjects prefer to play the lottery M. Why do people mostly violate the principle? Perhaps, they look for events with maximum probabilities when it comes to wins and do not care about expected gain since they play the game only once. They ignore the fact that in the long run both lotteries are expected to give the same winning amounts.

And now, instead of winning money if you have losses (but same amounts), then most of the people will like to play the game L. In this case, similar to the other game, people look for events with smaller probabilities when it comes to losses for them. And people have this selection when they think that they play the game once but not a long sequence to times.

Above game is somewhat extreme, therefore let us replace the lottery M as follows. Now the lottery M is such that you win $5000 with probability 1/5 and win nothing with probability 4/5. Again, both M and L have same expected utilities and it is interesting to find which is preferred by the people. Given both have same expected utilities, people may prefer to play the lottery L over the lottery M due to its larger probability of a win!

References


