CANADIAN UNDERGRADUATE JOURNAL OF COGNITIVE SCIENCE

2023 Edition

Canadian Undergraduate Journal of **Cognitive Science**

2023 Edition

The Canadian Undergraduate Journal of Cognitive Science (CUJCS) is an academic journal run by undergraduate and graduate Simon Fraser University students and funded by the Cognitive Science Student Society (CS^3). Founded in 2001, CUJCS has always aimed to provide undergraduate students in cognitive science and its related fields (computer science, linguistics, philosophy, and psychology) an opportunity to publish and showcase their work. We believe that our undergraduate journal helps further academic interest by providing a unique opportunity to experience the publication process and the space for students to make connections and exchange ideas with like-minded peers. Most importantly, it reflects the interdisciplinary work that is the hallmark of cognitive science.

ISBN 978-1-77287-107-4

ISSN 1499-7487

CS³
cognitive science student society
simon fraser university

simon fraser
student society

The CUJCS team gratefully acknowledges that this journal was edited and published on unceded Traditional Coast Salish Lands, including the Sḵwx̱wú7mesh Úxwumixw (Squamish), səl̓ilw̓ətaʔɬ (Tsleil-Waututh), xʷməθkʷəy̓əm (Musqueam), and kʷikʷəƛ̓əm (Kwikwetlem) Nations.

Letter from the Editor

In Richard E. Nisbett's *The geography of thought: how Asians and Westerners think differently...and why,* a common, recurring theme is that human beings are governed by culture. Their perceptions, behaviour, and biases are heavily influenced by the environment around them. Even how they think of themselves are influenced by the people they surround themselves with.

Similarly, in cognitive science, how we see ourselves is determined by the type of vehicle we place ourselves in—language, behaviour, theories of mind. All of this accumulates in cognition, the heart and soul of what studying Cognitive Science means.

It is our pleasure to introduce the 2023 edition of the Canadian Undergraduate Journal of Cognitive Science. This year's volume explores 11 papers, each one presenting numerous exciting areas within Cognitive Science and engaging with contemporary thoughts and ideas. These papers are presented with accompanying pieces of cognitive science-themed artwork and header doodles, all created by undergraduate students from universities across Canada.

Throughout this journal, our student authors look at and discuss cognition through a lens that is unique to them, a lens that encompasses their individuality and the various influences in their life. From analyses of learning to unconscious preferences, from understanding the chimpanzee mind to communicating through emojis, each paper presents a different perspective on cognition.

We hope that as you read through the next 11 papers, each discussing a different angle and story of cognition, you can gain a deeper, and more varied, understanding of what Cognitive Science is.

Hilary J. H. Tsui & Mark Giles, Editors-in-Chief

Table of Contents

Articles

Art Features

Age Influence on the Effect of Emojis on Affective Vocabulary Items: A Priming Approach

Hiu Man Ho

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Abstract

This exploratory study investigated whether younger and older people differ in the extent to which their comprehension of emojis activates associated affective content based on the priming effects of the participants in an affective emoji-word priming experiment. No notable priming effect on accuracy and reaction time (RT) was observed in the emoji-word priming experiment among the two age groups, suggesting that emojis do not significantly influence the results of linguistic processing among people in different generations. However, several limitations in the research method, experiment design, and procedure were identified. Future research on age and emojis with comprehensive approaches is encouraged.

Keywords: **age, emoji, priming, linguistic processing.**

Introduction

Emojis are digital pictograms developed from emoticons and internationalized since 2009 (Giannoulis & Wilde, 2019). They have been widely adopted to compensate for the boundaries of expressing emotions unambiguously through words under the demand for speed and the dearth of nonverbal cues in online text communication (Karpinska et al., 2019). Over 90% of the online population across the globe uses emojis in everyday interactions (Daniel, 2021); however, interpretations of emojis vary across people of different backgrounds, resulting in inconsistent interpretations and sometimes miscommunication (Danesi, 2017).

With the increase of the global ageing population, variations of emoji usage and interpretations among people of different ages have drawn scholars' attention in the past decade. Several corpus-based studies have described age differences in emoji usage in terms of frequency, preference, and valence interpretation, which defines as an individual's subjective attitude toward an entity, either positive or negative (American Psychological Association, n.d.). Younger users tend to include more emojis in texts than older users (López-Santamaría et al., 2019) and prefer using emojis representing emotions, as opposed to older people who favour emojis portraying people and objects. This was shown in the research of Sun (2021) that younger people tended to employ "smile" emojis with positive connotations to express negative meanings, while older people are more likely to use them to demonstrate positive feelings. Some survey-based research has also observed similar patterns (e.g. Herring & Dainas, 2020). Nevertheless, contrary to these results, a number of other investigations using similar approaches have failed to observe notable age differences in emoji use and understanding (e.g. Gallud et al., 2018; Jaeger et al., 2018; Weiß et al., 2020).

Age influence on emoji usage and interpretations remains indefinite based on the irreconcilable results of the previous studies with traditional corpus- and survey-based methods since they involve numerous variations such as applications, contexts, and digital devices. It implies a need for a different approach to emoji research regarding age, which can remove the impact of possible variations in daily communication. Priming is one of the examples. It is a tool conveying the processing of how a stimulus (prime) impacts one's reaction to a subsequent stimulus (target) in terms of speed based on their relatedness (Janiszewski & Wyer, 2014). If a respondent reacts faster to a target preceded by a related prime than the one after an unrelated prime, the related prime activates the target. Recently, Yang et al. (2021) published the first emoji study with a priming approach. An affective priming experiment was implemented on young people in China aged 19 to 26. With evidence from Event-Related Potentials (ERPs), the study shows that emojis contain affective content (the involvement of emotion and feelings) (American Psychological Association, n.d.), which thus helps facilitate the processing of subsequent affective words with a similar valence. However, age differences were not taken into account in the research.

This research takes the priming approach of Yang et al. (2021) as a reference in order to see whether younger and older people differ in the extent to which their comprehension of emojis activates associated affective content. Similar to Yang et al. (2021), participants read target words (positive or negative affective words) preceded by related or unrelated emoji primes; as this is an exploratory study, I measured their reaction times rather than event-related brain potentials. If participants respond faster to words preceded by a related emoji than words preceded by an unrelated emoji, this would be evidence that participants accessed the affective content of the emoji and that it aided in lexical access. Furthermore, if the size of this reaction time difference (the priming effect) differs between younger and older speakers, it would be evidence of an age-based difference in the comprehension of emojis. In addition to the emoji-word priming task, participants also completed a traditional word-word priming task, in which they read words preceded by related or unrelated words; this is a control task to ensure priming, in general, is observable in the sample.

Methods

Participants

Twenty Cantonese native speakers in Hong Kong recruited from The Hong Kong Polytechnic University campus and personal connections participated in the priming experiment without remuneration. As Gallud et al. (2018) suggested, smartphone owners are six times more likely to use emojis than people without a smartphone. According to the report of the Census and Statistics Department (2020), the two youngest and the two oldest age groups in which the percentage of persons who had at least one smartphone exceeded 95% in 2019 are aged 15–24, aged 25–34, and aged 45–54, aged 55–64 respectively. Participants were categorized into two age groups, i.e. "younger" people aged 18 to 30 and "older" people aged 50 to 64. The younger group incorporated 11 participants (6 women and 5 men), and the older group included 9 participants (5 women and 4 men). All participants were reported as frequent emoji users who have not had any visual impairment nor reading disorders and did not respond to the pre-test beforehand in the pre-experiment questionnaire.

Materials

Emoji Version

Smileys and hand gesture emojis in yellow skin tone, the default colour of “people emojis” among the six skin tones (Evans, 2017), in the WhatsApp version released in August of 2021 (version 2.21.16.20), were adopted for two reasons: emojis representing faces and human beings are the most commonly used emojis on the Internet (Miller et al., 2016), and WhatsApp Messenger is the most prevalent social networking and instant messaging app in Hong Kong (SimilarWeb, 2021). To evade possible influences due to variations of display on different devices, as noted by Miller et al. (2016), all emojis in the study were shown in JPEG format downloaded from the website of Emojipedia.

Pre-test for the selection of emoji stimuli

As some emojis may have different connotations for different people, and the connotations of emojis are continuously changing (Danesi, 2017; Gallud et al., 2018), I first conducted a norming pre-test to identify emojis that the younger and older groups would be likely to interpret in similar ways to ensure that the stimuli in the priming experiment would be emojis with well-agreed-upon affective valence. In line with Yang et al. (2021), an online survey titled “Personal Interpretation of Emojis” in traditional Chinese without a time limit was created with QuestionPro (available at https://www.questionpro.com/t/AUATuZpN2N). Respondents were required to rate the positiveness of 129 emojis (97 smileys and 32 hand gestures) on a 7-point Likert Scale individually, based on their subjective opinions.

A total of 29 valid responses were collected from Hong Kong native Cantonese speakers with frequent exposure to emojis within the university community (9 younger women, 5 younger men, 9 older women, and 6 older men), none of whom took part in the priming experiment afterwards. The average positivity score of each emoji among the two age groups was calculated respectively by averaging the Likert scale ratings across participants within the group, and the average positivity score among older people was subtracted from the average positivity score among younger people to quantify the extent to which a given emoji's interpretation differed between younger and older respondents. The smaller the absolute value of the difference, the more comparable the perceptions among the two parties are. Figure 1 summarizes the results of the pre-test, which were considered for picking emoji stimuli in the priming experiment afterwards.

Stimuli

I selected 80 emojis (59 smileys and 21 hand gestures) whose positivity difference scores differed by less than 0.6 between younger and older respondents and whose valences had average positivity scores across below 3 or above 4 on the 7-point Likert scale. Another 40 emojis were picked randomly from the remaining smiley and hand gesture emojis to be used as the primes for nonword filler targets, i.e. vocabulary items that do not exist in the world. Twenty non-smiley and non-hand-gesture emojis were utilized in the practice section to avoid any possible unwanted activation before the actual experiment. All emoji stimuli were displayed in JPEG format with 160 X 160 pixels on a white background during the experiment.

For word stimuli, I picked 40 vocabulary items for the 80 emoji primes from the latest version of the Chinese Valence-Arousal Words (CVAW) system, CVAW 4.0, which incorporates more than 5,000 affective Chinese vocabulary items with valence scores ranging from 1 (least positive) to 9 (most positive) (Yu et al., 2016). Another 200 non-affective Chinese words were used for i) 120 word stimuli, in-

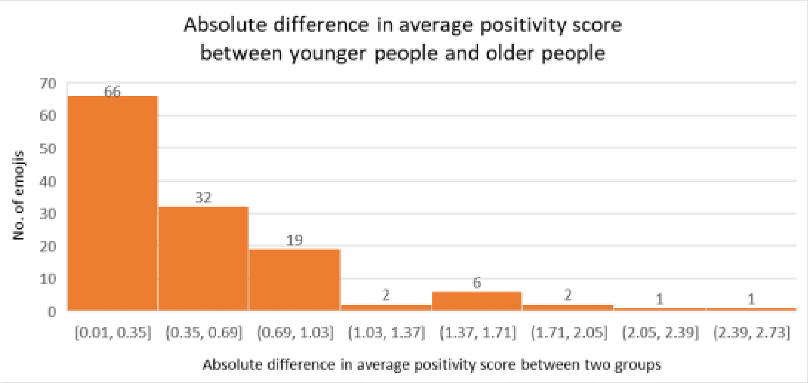


Figure 1. *A histogram summarizing the comparison of the average positivity score between the two groups, i.e. average positivity score of younger people of the emoji minus average positivity score of older people of the emoji (rounded to the nearest two decimal places).*

cluding 40 word targets, 40 congruent word primes, 40 incongruent word primes; ii) 40 word stimuli in the practice trials, encompassing 5 as word targets of congruent emoji primes, 5 as word targets of incongruent emoji primes, 10 as word primes for nonword filler targets, 10 as word primes for 5 congruent word targets and 5 incongruent word targets; as well as iii) 40 word primes for the nonword filler targets. The relatedness between word primes and targets is based on their meanings. Some examples are shown in Table 1. Only two-character Chinese vocabulary items were involved in the experiment to minimize the variations and limitations caused by the number of characters of the word stimuli. All word targets shown in SimSun, font size 20, in black on a white background in the experiment. The full list of primes and targets in different conditions can be found in the Appendices.

The design of emoji-word pairs considered the positivity scores of the pre-test and the mean of valence scores of the Chinese vocabulary items in the CVAW system. Emojis with a higher final average score were matched with terms with a higher valence rating to make up the congruent pairs and lower valence scores to make up the incongruent pairs. Some examples of the matching are summarized in Table 2.

Table 1: *Examples of the matching of word-word pairs*

Word target (English translation)	**Congruent word prime (English translation)**	**Incongruent word prime**
唇膏 (lipstick)	胭脂 (rouge blush)	烘乾 (drying)
太陽 (sun)	月亮 (moon)	平板 (tablet)
芭蕾 (ballet)	舞蹈 (dance)	癌症 (cancer)
運動 (sports)	足球 (football)	論文 (thesis)

A total of 200 trials evenly mixing emoji-word pairs and word-word pairs were incorporated in the experiment. The first 40 trials after the instructions at the beginning were for practice, followed by 160 randomized trials comprised of 20 congruent emoji-word pairs, 20 congruent word-word pairs, 20 incongruent emoji-word pairs, 20 incongruent word-word pairs, 40 emoji-nonword pairs and 40 word-nonword pairs. The stimuli were arranged into multiple lists in a Latin square design, so that no participant saw the same word or emoji twice and each word appeared in each condition. (For example, the target word 繽紛 “colourful” was preceded by an affectively congruent emoji for half of the participants and by an affectively incongruent emoji for the other half of participants).

Design and procedure

The experiment was conducted via DMDX (Forster & Forster, 2003). The participants’ task was to indicate, as quickly and accurately as

Table 2: *Examples of the matching of emoji-word pairs*

Word target (Eng. translation)	**Mean of valence scores in the CVAW system (out of 9)**	**Congruent emoji prime**	**Difference in average positivity scores (Younger VS. Older)**	**Final average positivity score (out of 7)**	**Incongruent emoji prime**	**Difference in average positivity scores (Younger VS. Older)**	**Final average positivity score (out of 7)**
繽紛 (colourful)	7.8	🥰	-0.46	6.30	😰	0.30	2.28
聰穎 (clever)	7.5	🤠	-0.55	5.06	🙁	-0.42	2.85
心 (heartbroken)	2.2	😓	0.038	2.55	😋	-0.22	5.82
違法 (illegal)	2.4	😞	-0.11	2.41	😗	-0.081	4.83

possible, whether each target word was a real Chinese word or a nonword. Participants made their responses by pressing the left or right Shift key only when they saw the target. They did not have to press anything when the prime was displayed. All primes were shown on screen for 360 milliseconds before the occurrence of the target, which disappeared when participants pressed a Shift key or the timeout duration (i.e. 2500 milliseconds) passed. After a 200-millisecond inter-trial interval, the prime of the next trial appeared.

Different coloured stickers were employed to assist participants in finishing the experiment quickly and accurately by lowering the efforts required to memorize the rules over the entire task. Participants were instructed to press the corresponding coloured rectangle stickers to indicate the validity of formed Chinese vocabulary items: a green sticker placed on the right Shift key to indicate correct items, and a red sticker placed on the left Shift key to indicate incorrect items. Usage of the green and red stickers on the two Shift keys was highlighted in the Cantonese introductory speech before the commencement of the experiment. During the experiment, no feedback was provided for participants' response times or errors in order to minimize distractions.

All recruited participants completed a short pre-experiment questionnaire in traditional Chinese on Google Forms to provide their demographic background information. Eligible participants were then invited to their most convenient indoor area without much background noise to take the priming experiment on a portable 14" ASUS X407U laptop. Participants were only told that they would work on a short task judging Chinese vocabulary; the invitation and instructions did not emphasize emojis or priming. To minimize the influence of stress on their judgements and reaction times (RT) during the task, the experimenters created a pressure-free environment by leaving the participants alone and unsupervised. The entire process took approximately 10 to 15 minutes per participant.

All accuracy and reaction times (RT) data were collected and gathered in .azk files for further analysis. Only trials with correct responses were included in the RT analysis, as the RT for targets involving invalid and incorrect responses (i.e. not reacting to the target within 2500 milliseconds, the time-out duration, or judging a non-word target as a correctly formed Chinese vocabulary item by pressing the inconsistent right Shift key or vice versa) may not reflect priming effects accurately. Data were visualized using R (R Core Team, 2022), Excel, and GraphPad Prism.

Results

Accuracy

The percentage of accuracy in responding to the targets in four conditions among younger and older age groups is depicted in Figure 2. Generally, the two groups achieved more than 80% accuracy when reacting to the targets in either condition during the experiment. However, older people tended to have a lower accuracy percentage in the emoji-word priming experiment (around 80% in both conditions) than in the word-word task (close to 90% in both conditions), while younger people, in general, attained over 90% accuracy in all four conditions.

Figure 2 shows that people responded more accurately to related targets than unrelated targets, particularly among older participants. Although priming effects were mostly predicted in reaction times, not in accuracy, the results indicate a consistency with a priming effect. In order to examine whether this apparent priming effect in accuracy is robust across participants, two graphs of individual priming effects in accuracy were made, as illustrated in Figure 3 (Weissgerber

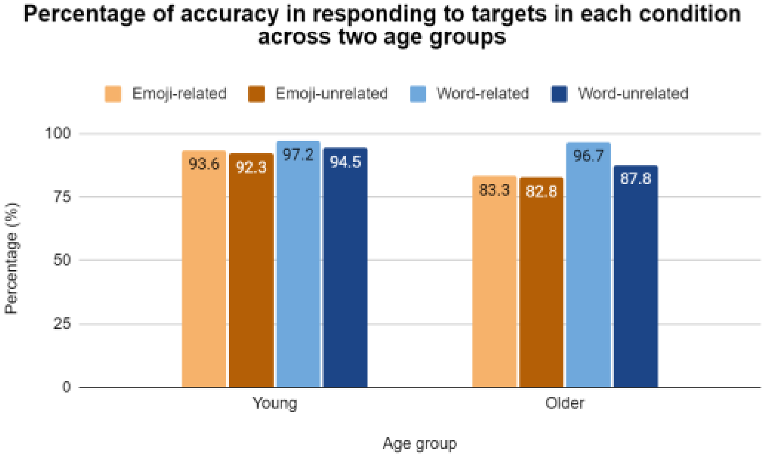


Figure 2. *Percentage of accuracy in responding to targets in each condition across two age groups.*

et al., 2015; Politzer-Ahles & Piccinini, 2018). Most priming effects in the accuracy of all participants are zero or close to zero (±5% - ±10%), and the apparent large priming effect for words among older participants was driven by just two participants and was not present in the others. I therefore concluded that there is not a substantial priming effect on accuracy rates in either group.

Reaction times

Figure 4 shows the average reaction time (RT) for each condition of younger people and older people, offering a broad picture of the results of the priming experiment. On average, both age groups seem to have priming effects on the word-word priming task. For younger people, the average RT of the targets with congruent word primes is 26 milliseconds faster than that of the targets with incongruent word primes, while for older people, the average RT of the targets with congruent word primes is 43 milliseconds quicker than that of the targets with incongruent word primes. However, for the emoji-word priming task, the two groups tend to react to targets congruent to the

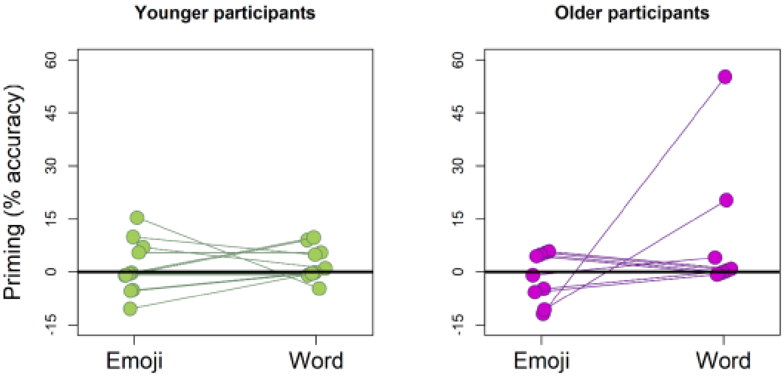


Figure 3. *The priming effects in accuracy (unrelated minus related) of each participant among the younger people group in both priming tasks. Dots above zero mean participants reacted more accurately in unrelated condition than the related condition. Dots are slightly jittered both horizontally and vertically to avoid completely overlapping one another.*

emoji primes 22– 25 milliseconds slower than those with incongruent emoji primes, suggesting that, on average, priming effects might not occur in the emoji-word priming task among the two age groups.

Two graphs of individual priming effects were also created to evaluate how robust these results are across participants (Weissgerber et al., 2015; Politzer-Ahles & Piccinini, 2018). Figure 5 illustrates the priming effects of each participant among the younger group in the two tasks. Among the 11 younger participants, the priming effects of nine participants (82%) are positive in the word-word priming task, demonstrating strong evidence that a semantic congruent word stimulus can facilitate the processing of the subsequent word stimulus among younger people. In the emoji-word priming task, the number of the positive and the negative priming effects tend to distribute evenly: five younger participants' priming effects are above zero, while the remaining six are the opposite, suggesting that there is no notable priming effect in the emoji-word priming experi-

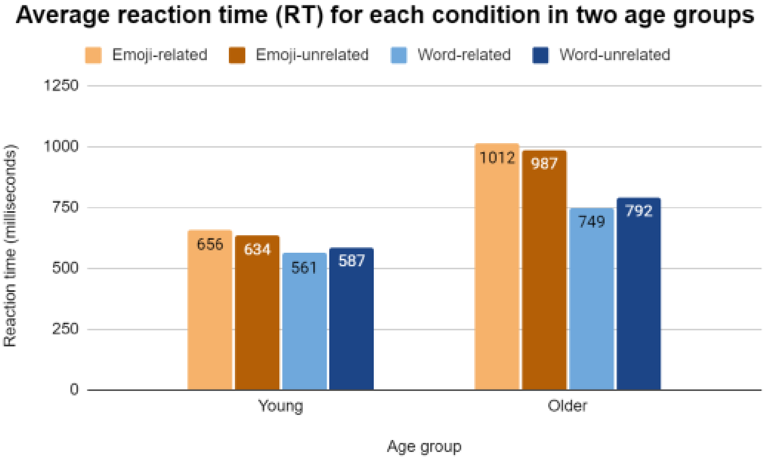


Figure 4. *Average reaction time (RT) for each condition in two age groups.*

ment. A valence-congruent emoji stimulus tends not to activate the processing of the ensuing word stimulus as a semantic congruent word among younger people.

The priming effects of each individual in the older group in the two tasks are depicted in Figure 6. Five out of nine participants show a positive priming effect in the word-word priming task. Although the remaining four reveal a negative priming effect, the numbers are close to zero, whereas the participants with positive priming effects tended to have quite large effects (note also that the magnitude of the priming effect for words is larger in this group than it is in the younger group, 43 ms vs. 26 ms). A semantic congruent word stimulus can also stimulate the processing of the succeeding word stimulus among older people. In the emoji-word priming task, five participants have a positive priming effect, while the other four possess a priming effect below zero. Similar to younger people, the number of individuals with a positive priming effect and those with a negative priming effect in the older group tend to distribute evenly. The priming effects of older people in the emoji-word priming task are not

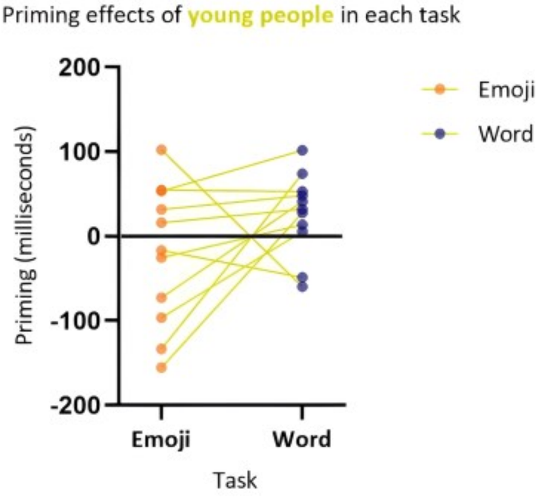


Figure 5. *The priming effects (unrelated minus related) of each participant among the younger people group in both priming tasks. Dots above zero indicate participants who had priming effects in the expected direction.*

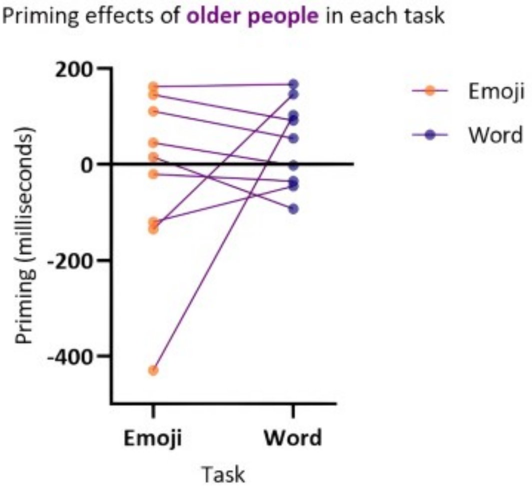


Figure 6. *The priming effects (unrelated minus related) of each participant among the older people group in both priming tasks. Dots above zero indicate participants who had priming effects in the expected direction.*

substantial, showing that an emotional word is not necessarily facilitated by a prior emoji with a congruent affective content among older people.

Table 3 briefly summarizes the priming effects of younger people and older people in the two priming tasks. The result presents that age influence was not identified in terms of the impact of emojis on processing the succeeding affective vocabulary items.

	Priming effect	
	Emoji-word priming task	**Word-word priming task**
Younger people	No notable priming effect	Positive
Older people	No notable priming effect	Tend to be positive

Table 3: *A summary of priming effects of the two age groups in the two tasks*

Discussion

The present study examined priming between emojis and words, as well as priming between word pairs, in both younger and older readers. For emoji-word pairs, neither group showed a priming effect: responses to words preceded by related emojis were no faster than responses to words preceded by unrelated emojis. On the other hand, for word pairs, both groups of participants showed priming effects, with faster reaction times to words preceded by related primes rather than unrelated primes.

The notion that age is not the only but perhaps one of the determinants causing miscommunication arising from emojis is reinforced, which is in line with previous studies suggesting that apart from age differences and the potential factors that this study attempted to evade (i.e. variations of cultural background and platform diversity), other multifarious factors can generate misunderstandings led by emojis, such as the unceasing development of the definition of emojis (Gallud et al., 2018), communication contexts (Gawne & McCulloch, 2019), personal habits and the social relationship between the sender and the receiver in the interaction (Herring & Dainas, 2020). Since numerous factors other than age may affect the influence of emojis on words, the differences within an age group seem more perceptible than the differences between the two age groups. This may explain the diverse distribution of priming effects in the emoji-word priming task within each age group in Figures 5 and 6.

Without a notable priming effect in the emoji-word priming task in both age groups, the assumption that emojis can function as adjectives is challenged. Since the outcomes suggest that a valence-congruent emoji tends not to activate the processing of the subsequent word stimulus as a congruent word prime, this research tends to contradict the belief that emojis are more than complementary and have the ability to somewhat replace words (e.g. Danesi, 2017; Thurlow & Jaroski, 2020; Albert, 2020). Rather, it agrees with the idea that emojis cannot replace words on a large scale (McCulloch, 2016) and are mainly "auxiliary means of communication" (Bai et al., 2019, p. 4). As a complement to written messages, when emojis are employed individually as a visual prime without a particular context as they were in the experiment, they may not possess a certain emotional content and tend not to facilitate the processing of the following vocabulary items with a congruent valence content linguistically.

The selection of lexical decision task for the emoji-word priming experiment might also influence the results. Although a lexical decision task, classifying whether the target is a real word or a nonword, can help reflect a reliable priming effect in the affective priming paradigm, the small magnitude limits the results, even with a rigorous control procedure (Ferré & Sánchez-Casas, 2014). Perhaps employing typical tasks, such as the evaluative decision task, in the affective priming paradigm, can help obtain more robust and significant results. The evaluative decision task, or the "affective word categorization task" (Yang et al., 2021) involves asking participants to determine whether the target is positive or negative. Moreover, the design of the word-word priming control task impacts the outcomes of priming effects since the relationship between primes and targets in the two experiments is different. In the emoji-word priming task, primes and targets are related in terms of meanings' positiveness, but they are congruent in terms of meanings in the word-word priming task, resulting in a more noticeable positive priming effect in the word-word task than in the emoji-word experiment.

An isolated priming experiment may not be sufficient to manifest age influence on the effect of emojis on the subsequent affective words exhaustively. In the research of Yang et al. (2021), no priming effects were demonstrated in the behavioural data of the priming experiment, but the capability of emojis to boost the attentional level in succeeding texts with a congruent meaning was concluded with the electroencephalogram (EEG) data, suggesting that congruent emojis do not affect the results of linguistic processing but enrich the attentional level and facilitate better processing due to fewer cognitive demands than incongruent emojis. A future study with proper control task design and EEG data collection is needed to investigate the effect of emojis on affective vocabulary items comprehensively.

Several other limitations of the experiment may have also hindered the results, such as the variation in venues where the experiment was conducted and the inconsistent backgrounds of the materials. The valence scores in the CVAW system developed in Taiwan might not apply to the perceptions of Hong Kong people pragmatically. As a result, some emoji-word pairs might not be congruent or incongruent for some participants, leading to inaccurate priming effects. Ideally, the experiment should be implemented in a designated language laboratory, with different pre-test tasks. Instead of only requiring participants to rate the positivity of each emoji with the 7-point Likert scale, the experiment should either ask respondents to score both the selected emojis and adjectives or request them to choose the most related adjectives in the emojis provided. This change might facilitate the design of the emoji-word trials in the priming experiment and ensure the pairs are congruent or incongruent in the Hong Kong context.

Finally, the present findings suggest some future research topics regarding the relationship between age and emojis. Interestingly, the average RT of the two conditions in the emoji-word priming experiment among older people are much longer than that of younger people, compared to the corresponding average RT of the two conditions in the word-word priming task, as depicted in Figure 4. It tends to be related to the findings of Koch et al. (2022) that younger people prefer employing emojis that help convey emotions while older people favour those representing people and objects. As younger people are more likely to use smiley and hand gesture emojis to give emotional meanings rather than to illustrate objects and people, older people, even those who claim to be frequent emoji users, may not be familiar with the selected emojis in the experiment. This unfamiliarity with the selected emojis in the experiment could have distracted them during the lexical decision task, ending up with a longer RT. However, this is a speculative suggestion, and a future study with a proper approach and design should be done to examine and provide a more concrete conclusion on the relevant issues.

Conclusion

In summary, the present exploratory study attempted to investigate whether age-based differences in the comprehension of emojis, especially smiley and hand gesture emojis, exist. By using an affective emoji-word priming experiment with a semantic word-word priming task as a control comparison after a pre-test grasping the latest general definitions of emojis in Hong Kong based on the reaction times (RT) and priming effects of each individual and on average, no notable priming effect was indicated in the emoji-word priming experiment for both age groups, which showed a certain extent of positive priming effects in the word-word priming task. The results suggest that emojis do not significantly impact the outcomes of linguistic processing. However, limitations of the experiment design and procedure might affect the outcomes, and the isolated priming experiment is not enough to conclude the effect of emojis on affective words besides linguistic processing. Future studies regarding age and emojis with comprehensive design and advanced statistical analysis methods are highly recommended.

References

Albert, G. (2020). Beyond the binary: Emoji as a challenge to the image-word distinction. In Thurlow, C., Dürscheid, C., & Diémoz, F. (Eds.), *Visualizing digital discourse: Interactional, institutional and ideological perspectives* (pp. 65–80). *De Gruyter Mouton.*

American Psychological Association. (n.d.). Affective. In *APA Dictionary of Psychology. Retrieved July 30, 2022, from https://dictionary.apa.org/affective*

American Psychological Association. (n.d.). Valence. In *APA Dictionary of Psychology. Retrieved July 30, 2022, from https://dictionary.apa.org/valence*

An, J., Li, T., Teng, Y., & Zhang, P. (2018). *Factors influencing emoji usage in smartphone mediated communications. Paper presented at 13th International Conference on Information, Sheffield, United Kingdom. https://doi.org/10.1007/978-3-319-78105-1_46*

Bai, Q., Dan, Q., Mu, Z., & Yang, M. (2019). A systematic review of emoji: Current research and future perspectives. *Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.02221*

Census and Statistics Department. (2020). *Usage of information technology and the internet by Hong Kong residents, 2000 to 2019. https://www.statistics.gov.hk/pub/B72010FA2020XXXXB0100.pdf*

Danesi, M. (2017). *The semiotics of emoji: The rise of visual language in the age of the internet. Bloomsbury Publishing.*

Daniel, J. (2021). *The most frequently used emojis of 2021. Unicode. https://home.unicode.org/emoji/emoji-frequency/*

Evans, V. (2017). *The emoji code: How smiley faces, love hearts and thumbs up are changing the way we communicate. Michael O'Mara Books.*

Ferré, P., & Sánchez-Casas, R. (2014). Affective priming in a lexical decision task: Is there an effect of words' concrete ness? *Psicologica: International Journal of Methodology and Experimental Psychology, 35(1),* 117–138.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35(1),* 116–124. *https://doi.org/10.3758/BF03195503*

Gallud, J. A., Fardoun, H. M., Andres, F., & Safa, N. (2018). *A study on how older people use emojis. Paper presented at the XIX International Conference on Human Computer Interaction, Palma, Spain. https://doi.org/10.1145/3233824.3233861*

Gawne, L., & McCullonch, G. (2019). Emoji as digital gestures. *Language@Internet, 17. https://www.languageatinternet.org/articles/2019/gawne*

Giannoulis, E., & Wilde, L. R. A. (2019). *Emoticons, kaomoji, and emoji: The transformation of communication in the digital age. Routledge.*

Herring, S. C., & Dainas, A. R. (2020). Gender and age influences on interpretation of emoji functions. *ACM Transactions on Social Computing, 3(2),* 1–26. *https://doi.org/10.1145/3375629*

Jaeger, S. R., Xia, Y., Lee, P. Y., Hunter, D. C., Beresford, M. K., & Ares, G. (2018). Emoji questionnaires can be used with a range of population segments: Findings relating to age, gender and frequency of emoji/emoticon use. *Food Quality and Preference, 68,* 397–410. *https://doi.org/10.1016/j.foodqual.2017.12.011*

Janiszewski, C., & Wyer, R. S. (2014). Content and process priming: A review. *Journal of Consumer Psychology, 24(1),* 96–118. *https://doi.org/10.1016/j.jcps.2013.05.006*

Karpinska, M., Kurzawska, P., & Rozanska, K. (2019). Emoticons: Digital lingua franca or a culture-specific product leading to misunderstandings? In E. Giannoulis, & L. R. A. Wilde (Eds.), *Emoticons, kaomoji, and emoji: The transformation of communication in the digital age* (pp. 67–81). *Routledge.*

Koch, T. K., Romero, P., & Stachl, C. (2022). Age and gender in language, emoji, and emoticon usage in instant messages. *Computers in Human Behavior, 126. https://doi.org/10.1016/j.chb.2021.106990*

López-Santamaría, L. M., Gomez, J. C., Almanza-Ojeda, D. L., & Ibarra-Manzano, M. A. (2019). *Age and gender identification in unbalanced social media. Paper presented at the 29th International Conference on Electronics, Communications and Computers, Puebla, Mexico. https://doi.org/10.1109/CONIELECOMP.2019.8673125*

McCulloch, G. (2016, June 29). *A linguist explains emoji and what language death actually looks like. The Toast. https://the-toast.net/2016/06/29/a-linguist-explains-emoji-and-what-language-death-actually-looks-like/*

Miller, H. J., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). *"Blissfully happy" or "ready to fight": Varying interpretations of emoji. Paper presented at the 10th International AAAI Conference on Web and Social Media, Cologne, Germany. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/viewPaper/13167*

Politzer-Ahles, S. J., & Piccinini, P. (2018). On visualizing phonetic data from repeated measures experiments with multiple random effects. *Journal of Phonetics, 70,* 56–69. *https://doi.org/10.1016/j.wocn.2018.05.002*

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

SimilarWeb. (2021). *Top Apps Ranking. https://www.similarweb.com/apps/top/google/store-rank/hk/communication/top-free/*

Sun, W. (2021). Analysis of pragmatic functions of "smile" emoji in Chinese WeChat communication between people of different ages. *Sino-US English Teaching, 18(8),* 207–215. *https://doi.org/10.17265/1539-8072/2021.08.001*

Thurlow, C., & Jaroski, V. (2020). 'Emoji invasion': The semiotic ideologies of language endangerment in multilingual news discourse. In Thurlow, C., Dürscheid, C., & Diémoz, F. (Eds.), *Visualizing digital discourse: Interactional, institutional and ideological perspectives* (pp. 45–64). *De Gruyter Mouton.*

Weiß, M., Bille, D., Rodrigues, J., & Hewig, J. (2020). Age-related differences in emoji evaluation. *Experimental Aging Research, 46(5),* 416–432. *https://doi.org/10.1080/0361073X.2020.1790087*

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLOS Biology, 13(4). https://doi.org/10.1371/journal.pbio.1002128*

Yang, J., Yang, Y., Xiu, L., & Yu, G. (2021). Effect of emoji prime on the understanding of emotional words—Evidence from ERPs. *Behaviour & Information Technology,* 1–10. *https://doi.org/10.1080/0144929X.2021.1874050*

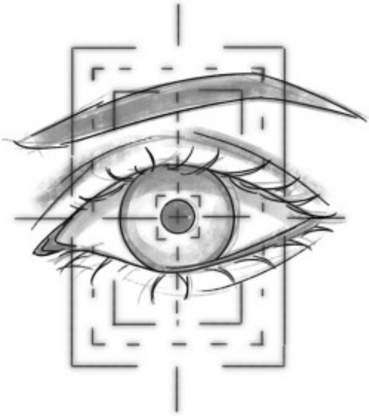
Yu, L. C., Lee, L. H., Hao, S., Wang, J., He, Y., Hu, J., Lai, K. R., & Zhang, X. (2016). *Building Chinese affective resources in valence-arousal dimensions. Paper presented at the 2016 Conference of the North American Chapter of the Associa tion for Computational Linguistics: Human Language Technologies, San Diego, California. https://doi.org/10.18653/v1/N16-1066*

Head
A single mind’s where we are stuck
To know others’ thoughts, we’re out of luck
  But research has shown
  Shared goals that are known
4Fs: to feed, fight, flee, and mate

Two Heads
The astonishing power of the brain
Where thoughts run through like a train
  But the head and the lips
  That sit at one’s hips
Force logical thought to chicane

Head (Part 1)

Angel Martin



The Effects of Mindfulness on Mind Wandering: An Eye-Tracking Study

Aleksandrs Baskakovs

Aarhus University

Department of Linguistics, Cognitive Science and Semiotics

Otto Sejrskild Santesson

Aarhus University

Department of Linguistics, Cognitive Science and Semiotics

Abstract

Mind wandering is a pervasive phenomenon, accounting for as much as 50% of the total waking lifespan. While not necessarily harmful, in a context where one engages in an attentionally demanding task, mind wandering can be costly. Preliminary studies have shown that mindfulness can reduce the number of mind wandering episodes. There is less evidence, however, for how mindfulness affects mind wandering episodes themselves. This is mostly due to mind wandering being treated as a dichotomy: either attention is decoupled from the external environment or it is not. In this study, we use eye-tracking to pilot a method for investigating mind wandering in a continuous, rather than binary, manner. Subsequently, we measure how mindfulness affects the intensity of mind wandering episodes. Six university students completed a reading task and had to report whether they were mind-wandering every time a probe popped up. Prior to the task, participants in the mindfulness group had to go through a 10-minute audio-based guided meditation. Eye-tracking was used to analyze 10 seconds of eye movements prior to when participants indicated that they were mind wandering. It was found that eye movement behaviours typical for mind wandering episodes (prolonged fixation duration and decreased fixation count) were generally lesser for mind wandering episodes of participants in the mindfulness condition compared to the control condition.

Keywords: **mind wandering, mindfulness, eye tracking.**

Introduction

Mind wandering

Prevalence and status quo of the field

Mind wandering is the ubiquitous phenomenon of when the attention drifts from the task at hand to other unrelated trains of thought (Christoff, 2012; Smallwood & Schooler, 2015). The contents of mind wandering can vary from fantasizing about winning the lottery to deliberate planning about future events, such as thinking about what to make for dinner (Barnett & Kaufman, 2020; Mooneyham & Schooler, 2013). Other typical examples of mind wandering include walking into a room and forgetting why you went there, thinking about family-related manners while driving to work, or reading a paragraph of text and then realizing that none of the read text had been meaningfully processed. The prevalence of mind wandering is so pronounced, that 96% of American adults experience mind wandering on a daily basis (Singer & McCraven, 1961), further underscored by the estimate of mind wandering occupying up to 50% of the contents of the awake mind (Barnett & Kaufman, 2020; Killingsworth & Gilbert, 2010; Klinger, 2009; Klinger & Cox, 1987; McMillan et al., 2013). It is therefore perplexing that it is only in the past decade that mind wandering has gained significant traction as a topic of study within the academic community (Smallwood & Schooler, 2006, 2015). A viable explanation to this might be that it is due to the reverberations from the behaviourist era of cognitive research, which had a pronounced skepticism directed toward the study of the inner workings of the mind, the scientific community has been more tentative toward studying mind wandering (Callard et al., 2012; Smallwood & Schooler, 2015).

Another challenge that the field faces is similar to the one of creativity, which has faced problems with incoherent terminology: there are a plethora of different terms for the same cognitive phenomenon (Barnett & Kaufman, 2020). A multitude of the words are synonyms of each other while some of the terms slightly differ from each other in their definitions. The terminology has such a lack of consistency that a multitude of researchers have in their article included an extensive nomenclature of the different terms used within the field (e.g., Christoff, 2012; Gruberger et al., 2011). This variability is showcased by the following, non-exhaustive list of terms that have been used in lieu of "mind wandering": "daydreaming, spontaneous thought, fantasy, zoning out, thought intrusions, task-irrelevant thoughts, perceptual decoupling, stimulus-independent thought, unconscious thought, internally generated thoughts, offline thought, incidental self-processing, undirected thought, and self-generated thought" (Barnett & Kaufman, 2020, p. 6). To address this complication, Barnett and Kaufman (2020) propose "mind wandering" to be the preferred term in onward research to build common phraseology within the field and to facilitate "current researchers to connect their work with the work of scholars who trod similar paths before them" (McMillan et al., 2013, p. 1). As a continuation, they proceed to posit that the two most essential components of mind wandering are intentionality (whether the mind wandering is deliberate or spontaneous) and plausibility (how close the contents of the mind wandering episode are to reality)" (Barnett & Kaufman, 2020, p. 7). However, multiple authors more so emphasize the stimulus-dependence aspect of mind wandering (Christoff, 2012; Mason et al., 2007; Smallwood & Schooler, 2015). Stimulus-independent thought (SIT) is the term used when the internal mental contents occur with little to no relation to external events. In other words, it is when the train of thought is decoupled from current sensory information (known as perceptual decoupling) (Antrobus, 1968; Christoff, 2012; Schooler et al., 2011; Teasdale et al., 1993). Conversely, stimulus-dependent thought (also referred to as stimulus-oriented thought) "reflects attention towards the current external sensory environment" (Christoff, 2012, p. 53).

There are other proposals on how the taxonomy of mind wandering should be structured (e.g., Smallwood & Schooler, 2015), but a full assessment of the different taxonomy used within the field is beyond the scope of this paper—the above delineation was simply to outline the status quo of mind wandering and the difficulties that the field faces. For the remainder of this paper, in alignment with the abovementioned proposal by Barnett and Kaufman, the term "mind wandering" will be employed when referring to the mental phenomenon of when one's thoughts and attention shift away from the task at hand.

The behavioural and neurocognitive response to mind wandering

Due to mind wandering being such a common mental occurrence, it has been an incentive for researchers to investigate the width and depth of how both behavioural and neurocognitive measures are affected. In their study investigating mind wandering's influence on general aptitude, Mrazek and colleagues (2012) found working memory capacity (WMC, measured with complex span tasks) and the reported amount of mind wandering episodes to be significantly negatively correlated, a relationship which since has been replicated numerous times (McVay & Kane, 2012; Robison & Unsworth, 2018). In addition, they found that measures of gF (fluid intelligence, measured using Raven's Progressive Matrices [RPM]) were significantly negatively correlated with mind wandering. Both of these measures were found to be correlated with the participants'

SAT (Scholastic Aptitude Test) scores. By employing structural equation modelling based on the measures of general aptitude and mind wandering occurrences, they extracted two latent variables: one denoting mind wandering episodes during the WMC and gF tasks and the other denoting general aptitude based on these two measures, along with SAT scores. The latent variable of mind wandering was found to explain 49% of the variance in the latent variable of general aptitude. By extrapolating these results, it is plausible that mind wandering would be correlated with deficits in cognitive performance in other areas, contingent on the predictive power that the aforementioned measures of general aptitude have on performance in a wide variety of contexts (Conway et al., 2008; Deary et al., 2007; Frey & Detterman, 2004; Kane et al., 2005; Rohde & Thompson, 2007; Schmitt et al., 2009; te Nijenhuis et al., 2007). This proposition is in line with mind wandering being linked to reduced reading performance and comprehension (McVay & Kane, 2012; Reichle et al., 2010; Smallwood, McSpadden, et al., 2008), reduction of retention of new information in academic settings (Seli, Wammes, et al., 2016; Smallwood et al., 2007), and reduced driving performance, including increased risk of causing traffic accidents (Galera et al., 2012; Yanko & Spalek, 2014) and medical malpractice (Smallwood et al., 2011).

A possible account of the interrelation between mind wandering and diminution in cognitive performance is that problem-solving and working memory involve recruitment of executive resources (Alvarez & Emory, 2006), which is also the case for mind wandering (Smallwood & Schooler, 2006). Mind wandering therefore competes with the task at hand, such as a lecture or a complex span task, for the coordination and control of working memory resources, which can result in performance impairment (Kane et al., 2007; Mrazek, Smallwood, Franklin, et al., 2012; Smallwood & Schooler, 2006). The prioritization of allocating cognitive resources on either the task at hand or mind wandering rely on a number of different factors, such as the incentive for attending the task at hand (e.g., if an individual is not motivated to attend a lecture, they are more likely to move their attention to task-unrelated thoughts that are more aligned with their goals) (Seli, Wammes, et al., 2016; Smallwood et al., 2007) or the demands of the task at hand (e.g., a proficient driver is more likely to switch their attention to unrelated thoughts, such as planning what to make for dinner, than to attend the undemanding task of driving the car) (Kane et al., 2007; Levinson et al., 2012; Mrazek, Smallwood, Franklin, et al., 2012; Smallwood & Schooler, 2006; Teasdale et al., 1995).

Multiple studies that have investigated the neurocognitive response to mind wandering have found evidence that the brain areas involved with mind wandering overlap with areas constituting the default mode network (DMN), a set of cortical regions where activity within is associated with the absence of a task or when the brain is "at rest" (Mason et al., 2007; Seli et al., 2016). Researchers have mapped the most prominent constituents of the DMN, which include the medial prefrontal cortex (medial PFC), the posterior cingulate/ precuneus region, and the temporoparietal junction (TPJ), and found strong evidence for neural activity in this network is associated with mind wandering episodes (Christoff et al., 2009a; Christoff, 2012; Fox et al., 2015; Mason et al., 2007; McKiernan et al., 2006; Raichle et al., 2001). Additionally, recruitment of the frontoparietal-control network, which comprises the dorsolateral PFC and the dorsal anterior cingulate (a network associated with executive control) (Schooler et al., 2011), has also been observed during mind wandering (Christoff et al., 2009a; Fox et al., 2015). This positive functional connectivity between these two networks "might represent a dynamic interplay whereby executive control regions guide, evaluate, and select among the various spontaneous streams of thoughts, memories, and imaginings offered up to consciousness by the DMN" (Fox et al., 2015, p. 617). Evidence from a brain imaging study has also showcased the negative functional connectivity between the DMN and the primary sensory cortices, which is an indication of the perceptual decoupling that coincides with mind wandering (Christoff, 2012; Schooler et al., 2011). This correlation is supported by results from a study conducted by Smallwood and colleagues (2008), where the amplitude of a late positive component of the event-related potential (ERP) (known as the P3), which is associated with task-relevant cortical processing, was found to be significantly reduced during mind wandering compared to on-task cognition. This trend was also present for both of the amplitudes of the visual P1 and auditory N1 ERP components, which are associated with sensory-level processing of visual and auditory information, respectively (Kam et al., 2011) (Figure 1).

The above findings of decreased neurocognitive response to external stimuli depict the cognitive mechanics behind mind wandering: When attention is shifted from the external sensory environment to internal trains of thought, due to an interplay between the DMN and executive function, it interferes with the online processing of sensory information. Exactly what role perceptual decoupling plays in mind wandering has yet to be assessed, although some speculations have been put forth: One possibility is that perceptual decoupling is necessary in terms of the generation and sustainment of mind wandering. According to this view, perceptual decoupling displays the "flexible reorganization of processes to facilitate a conscious focus on self-generated information" (Smallwood & Schooler, 2015, p. 501). Similar to when attention prioritizes one modality of sensory

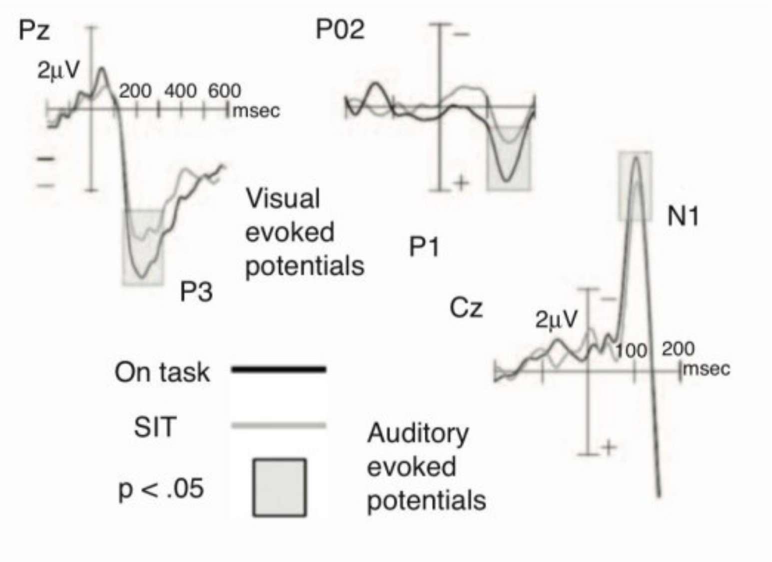


Figure 1. *Differences in task-evoked neural responses between SIT (stimulus-independent thought, i.e. mind wandering) and on-task cognition. EEG measurements exhibit lower amplitudes in the P3, P1, and N1 ERP components during mind wandering, indicating that the mental phenomenon tends to be decoupled from perceptual information from the external environment. Abbreviations: Cz, central midline electrode (according to the International 10–20 system for EEG recording, localized by the central region); Pz, posterior midline electrode (localized by the parietal lobe); PO2, posterior right lateral electrode (localized by the parietal -occipital lobe); SIT, stimulus-independent thought; EEG, electroencephalography; ERP, event-related potential. (Figure reprinted from Schooler et al., 2011).*

information over another to receive preferential processing (Posner & Petersen, 1990), perceptual decoupling may enable mind wandering by interfering with the processing of information unrelated to its contents (i.e., external sensory information). Another account is that perceptual decoupling is not a process whose main function is to insulate mind wandering from the external environment, but more so a corollary of the limited resources of the attentional system (Franklin et al., 2013; Smallwood & Schooler, 2015).

The neurocognitive activation profile of mind wandering is similar to the one of creative thinking, where the parallel recruitment of the DMN and executive brain regions is also present (Christoff, 2012; Kounios et al., 2006; Kounios et al., 2008; Subramaniam et al., 2009). This similarity illuminates one of the beneficial aspects of mind wandering, namely its relation to incubation, creative problem-solving, and ideation (Gable et al., 2019; Schooler et al., 2011). Evidence for this relationship was found through the meta-analysis conducted by Sio and Ormerod (2009), which showed that the creative benefits of incubation intervals are greatest when individuals are occupied by a non-demanding (and thus mind-wandering-inducing) task relative to either a demanding task or no task at all (Schooler et al., 2011). Another beneficial aspect of mind wandering is that since it is often occupied with autobiographical future thinking (D'Argembeau et al., 2011; Smallwood, Beach, et al., 2008), it is believed that this type of mental simulation can prepare the individual for different potential future outcomes, or in other words, plan ahead, thus functioning as a tool which is necessary to successfully navigate the complex social world" (Schooler et al., 2011, p. 321).

The case of mind wandering is therefore not simply something to be encapsulated as a purely detrimental phenomenon, but that it also comprises positive aspects. This is although the case for the majority of the research there has been on mind wandering: In their review of two decades of articles that addressed the costs and benefits of mind wandering, Mooneyham and Schooler (2013) found 30 papers that focused on the negative aspects of mind wandering whereas they only found 6 that addressed the possible benefits. One of the paradoxical features of mind wandering is that even though it poses some beneficial aspects, as outlined above, it is nevertheless detrimental to the goals of the moment, such as reading a book, completing a written assignment, or other tasks of everyday life. An important research question that has emanated from this conflict is whether it is possible to retain the positive aspects of mind wandering while diminishing the negative ones—to be able to tune the attentional system in order to both have the ability to engage in imaginative simulation while concurrently curtailing the detrimental consequences of losing focus on the goals of the moment. Initial research on mindfulness has shown auspicious results indicating it could serve as an efficacious technique for ameliorating the negative impact mind wandering has on cognitive performance, both in terms of reading comprehension and working memory capacity (Jha et al., 2010; Mrazek et al., 2013).

Mindfulness

Mindfulness is the notion of a mental mode that is broadly defined as "receptive attention to and awareness of present events and experience" (Good et al., 2016, p. 116). The concept dates back several millennia and has its roots in Buddhist traditions, where conscious attention and awareness are actively cultivated through mindfulness practice, typically in the form of meditation (Schmidt, 2011). Being in a mindful state of cognitive processing is hallmarked by attending stimuli simply "as they are", meaning that it is involved with experiential processing as opposed to conceptual processing: "It does not compare, categorize, or evaluate, nor does it contemplate, introspect, reflect, or ruminate upon events or experiences based on memory" (Brown et al., 2007, p. 213; Good et al., 2016). Furthermore, it is being aware of thoughts as *thoughts,* without engaging in them, but simply being a neutral observer of the contents of one's own mind (Brown et al., 2007). Another key aspect imputed to mindfulness is its attentional flexibility, which is the ability to direct one's attention to different levels. In other words, using visual perception as an analogy, one can choose to hyperfocus on one object in a visual scenery, such as an individual leaf on a tree, thus centring one's attention on one of the lowest levels. With attentional flexibility, the spotlight focus of attention can be widened from focusing on the branch, to the whole tree, and on to the whole forest, thus adeptly shifting between the different levels or scopes of attention. In that way, by using attention more mindfully, one can cultivate a more meta-aware and neutral view of one's own thoughts and emotions, and analyze them from a distance without acting upon them (Brown et al., 2007; Good et al., 2016). This phenomenological account of increased attentional control is congruent with the findings of mindfulness practice being linked to betterment of attentional control in terms of reductions in habitual allocation of attention and attention to distracting information (Jha et al., 2007; Tang et al., 2007; Wadlinger & Isaacowitz, 2011; Zylowska et al., 2008). Neurocognitive measures of mindfulness are in accordance with this finding as brain imaging studies have shown activation in brain regions are associated with executive function, namely the PFC and the ACC, of which selective attention is a component process (Alvarez & Emory, 2006; Zeidan, 2015). Another area that seems to benefit from mindfulness practice is cognitive capacity, showcased by observed increments in measurements of WMC (Jha et al., 2010; Roeser et al., 2013) and gF (Gard et al., 2014; Tang et al., 2007) followed by meditation, which is hypothesized to be due to the augmentation of the attentional control induced by mindful practice (Good et al., 2016; see also Jha, 2002; Redick & Engle, 2006). Consequently, these findings substantiate the positive relationship between mindfulness and cognitive performance (Mrazek et al., 2013; Smallwood & Schooler, 2015), since measures of WMC and gF are positively correlated with performance in a variety of contexts, as mentioned previously, such as SAT score (Smallwood et al., 2007; Smallwood & Schooler, 2006).

By juxtaposing the above findings with the cognitive correlates of mind wandering, as outlined in the previous section, it is possible to draw out similarities between the two mental states, although in opposite directions: Mind wandering is negatively correlated with scores of WMC and gF along with general performance, while mindfulness is positively correlated with these same measures. This seeming polarity indicates that these are two opposing constructs; this is further highlighted by the robust findings of significant dampening of activity in the DMN for all levels of mindfulness (Brefczynski-Lewis et al., 2007; Brewer et al., 2011; Farb et al., 2007, 2010; Garrison et al., 2013; Grant et al., 2011; Pagnoni, 2012; Tang et al., 2009; Taylor et al., 2013; Zeidan et al., 2011) along with increased activity in brain regions involved in sensory processing (Brewer et al., 2011; Hasenkamp & Barsalou, 2012; Manna et al., 2010; Zeidan, 2015), since mind wandering is linked to increased activity in the DMN and perceptual decoupling (i.e. curtailed neural response to sensory information). There is therefore substantial evidence to warrant an inquiry into whether mindfulness practice is a viable option to ameliorate the detrimental effects of mind wandering. Such an investigation has already taken place: Mrazek and colleagues (2012) found that mindfulness training significantly decreased the amount of mind wandering episodes whereas Rahl and colleagues (2017) showed that mindfulness training had beneficial effects on behavioural measures of mind wandering performance outcomes, thus establishing “a causal relationship between the cultivation of mindfulness and subsequent reduction in mind wandering” (Mrazek, Smallwood, & Schooler, 2012, p. 446). To give an account of the neural mechanisms behind this causal relationship between mindfulness practice and mind wandering, Brewer and colleagues (2011) hypothesize that the dampening of the activity in the DMN induced by mindfulness is the basis of the observed mitigation of mind wandering, although future research is needed to directly test whether this is the case or not.

Gauging effects

The previous section suggests mindfulness practice is promising in ameliorating mind wandering. It is unclear, however, how great the effect of mindfulness is on the reduction of mind wandering.

It has been previously shown (Mrazek et al., 2013) that mindfulness can reduce the number of mind wandering episodes in healthy individuals. It should be noted that this study investigates the effect of mindfulness on the occurrence of mind wandering episodes, rather than the intensity or strength of these episodes.

Indeed, most mind wandering research deals with mind wandering in a binary nature: whether the participant is currently mind wandering or not.

A false dichotomy?

The mind is generally described as consisting of different cognitive processes organized at different hierarchical levels, from early perceptual-motor processes to abstract representations at higher levels (Gazzaniga, 2009). Mind wandering, however, is dominantly treated as a dichotomy.

> [T]he dichotomy-hypothesis proposes that different levels of cognitive processing are decoupled from external input in an all-or-none fashion: during task focus all hierarchical levels of cognitive processing are coupled to the external environment, but when the mind wanders this coupling breaks down at all levels. (Schad et al., 2012, p. 180)

The above quote is from an article by Schad et al. (2012), appropriately titled “Mind wanders weakly, mind wanders deeply”. The study addresses the problem of the dichotomy hypothesis and introduces a model of graded attentional decoupling, which predicts that cognitive processing of external input fails at different hierarchical levels. This leads to different degrees of attentional decoupling—from weak to deep—which is supported by the findings of the study.

An additional argument for the continuous nature of mind wandering can be drawn from the dampening effects mindfulness has on activity in the DMN; since activation of the DMN can be measured in a continuous matter, one could argue that this should be extended to measuring mind wandering on different levels, since the two have shown to be positively correlated (Brewer et al., 2011; Raichle et al., 2001).

Furthermore, the continuous nature of mind wandering has been showcased in a tDCS study by Filmer et al. (2019). The effect of stimulation polarity and the intensity on mind wandering was investigated, where mind wandering was assessed via a probe asking "To what extent have you experienced task-unrelated thoughts prior to the probe? 1 (minimal) - 4 (maximal)" (Filmer et al., 2019). Indeed, it was found that increments in tDCS intensity led to linear increases in the reported extent of experienced mind wandering.

The prevalence of the dichotomy-hypothesis within the field of mind wandering can be attributed to the most commonly used method used for detecting mind wandering episodes, namely probes appearing during tasks with a dichotomous "Yes/No" answer option (e.g., Levinson et al., 2012; Uzzaman & Joordens, 2011).

Self-reports

The most widely used method for collecting mind-wandering data is the so-called *probe-caught* method, which involves sampling the participants for whether their focus is currently on- or off-task at random intervals during the task.

Whilst an inherently subjective measure, the probe-caught method has been shown to correlate with objective methods of mind wandering detection: performance measures such as reaction times (Cheyne et al., 2006), text comprehension (Smallwood, McSpadden, et al., 2008), memory (Risko et al., 2012) and physiological measures such as eye movements (Reichle et al., 2010; Smilek et al., 2010).

The probe itself often consists of a simple binary question on whether the subject's attention is currently directed or not directed toward the task. In some cases, binary response is replaced with a scale (e.g., a 5-point Likert scale, see Mrazek, Smallwood, & Schooler, 2012). Even more rarely do the measurements include participants' reports of the contents of their thoughts at the time of the probe (Klinger, 1984).

Another approach is the *self-caught* method, which is based on participants indicating, at any moment during the execution of the task, whether their attention has shifted (Giambra, 1989). The advantage of this method over the probe-caught method is that there can be an unlimited number of reports, as the reports are not limited by probe placement. However, this type of reporting requires the participants to be aware of their internal state, which is also referred to as meta-awareness (Schooler, 2002), which people often fail to achieve. This lack of self-awareness of mind wandering episodes is thought to be caused by mind wandering recruiting the brain regions that are necessary in order to notice its occurrence (Schooler et al., 2011).

Utilizing self-reports is a necessary measure required to detect mind wandering episodes. But it is only that—a measure for detection. In order to escape the dichotomy problem, self-report methods should be enriched with other sources of data (Hawkins et al., 2015). These would include posture (Seli et al., 2014), reading speed (Kopp et al., 2015), brain activity (Christoff et al., 2009), eye movements (Foulsham et al., 2013a), eye blinks (Frank et al., 2015), and pupil diameter (Franklin, Broadway, et al., 2013).

As seen above, in addition to industry-established *self-caught* and *probe-caught* techniques, mind wandering can be confidently detected using multiple ocular metrics, which speaks to the reliability and robustness of utilizing eye-tracking. Indeed, multiple studies (Bixler & D'Mello, 2014; Schad et al., 2012; Uzzaman & Joordens, 2011), since the mainstream introduction of the method by Reichle and colleagues (2010), have used eye-tracking to investigate mind wandering episodes. Considering the reliability of both methods, this study relies on both *reports by participants* and *eye-tracking metrics* to examine the effect mindfulness has on mind wandering.

Dissecting mind wandering using eye movements

The confidence in the ability to measure mind wandering using eye-tracking stems from decades of research on how eye movements are modulated by attention (Corbetta et al., 1998). What is specifically useful in the case of mind wandering detection, is that whenever attention shifts from an external stimulus (e.g., a text) to an internal stimulus (e.g., a thought), the change is reflected by overt embodied components whereby external input is blocked at the sensory level—the movement of the eyes (Just & Carpenter, 1980).

One of the theorized functions of eye blinks is to modulate the trade-offs between attention to mind wandering thoughts and to external, task-related stimuli. This trade-off is achieved physically by the closing of the eyelids and through the suppression of neural activity of visual processing before and after the actual lid closure (Smilek et al., 2010).

It has been shown that eye movements during reading exhibit a consistent pattern, which is impacted by lexical features such as word length and word frequency (Rayner, 1998). In an episode of mind wandering, the regularity of eye movements breaks down, and people tend to, for example, focus on low-frequency words for a shorter time span and vice-versa, indicating inadequate processing of the text, or "mindless reading" (Schad et al., 2012). Such propensity can be related to perceptual decoupling, since the decrement of time spent fixated on the infrequent words indicates that the attentional resources are being allocated to other processes unrelated to the external stimuli, i.e., mind wandering.

Whilst utilization of such non-standard variables has been productive for studying the ocular-motoric nature of lexical processing, standard variables of eye gaze, such as fixation duration and fixation count, have been shown to be differentiable between episodes of mind wandering and reading. People tend to have fewer and longer fixations during mind wandering episodes, indicating the state of being "zoned-out" and thus being interrupted from in-depth text comprehension (Reichle et al., 1998, 2010; Uzzaman & Joordens, 2011).

The field of mind wandering research has been dominantly fixed on viewing mind wandering categorically. As showcased in the previous sections, recent literature has justifiably argued for a more continuous look at mind wandering. Eye-tracking is perfectly suited to do that in an ecologically valid setting, as opposed to fMRI scans, where the participant is confined to complete the experiment in a noisy and poky setting. The employment of eye-tracking allows us to dissect mind wandering episodes moment-to-moment and compare one mind wandering episode to another one. Thus, it is possible to shed light on the question of how great the effect of mindfulness is on mind wandering.

Present study

This pilot study leverages previous insights from eye-tracking studies investigating mind wandering to 1) investigate the effect mindfulness has on mind wandering, and 2) showcase how eye-tracking can be used to look at mind wandering in a continuous, rather than binary, manner.

It has been shown that during mind wandering episodes, fixation duration tends to increase and fixation count tends to decrease (Reichle et al., 2010). It is also known that mindfulness reduces the activity of the default mode network, a brain area associated with mind wandering (Tomasino et al., 2013), and that oculomotor and attentional processes are tightly integrated at a neural level (Corbetta et al., 1998). Taking this into consideration, we predict that the effect mind wandering has on eye movement behaviours will be lessened in participants who underwent a guided meditation session prior to the experiment.

H1: Total fixation count during a mind wandering episode will be generally greater for participants in the mindfulness condition compared to the participants in the control condition

H2: Fixation duration during a mind wandering episode will be generally shorter for participants in the mindfulness condition compared to the participants in the control condition

Methods

Participants

Six undergraduates (50% female and 50% male, with a mean age of 23 years and a range of 22–25 years) from Aarhus University participated in the experiment. All subjects were proficient English speakers with normal, uncorrected vision. None of the participants had an in-depth familiarity with the text used in the study. The participants were assigned to either the experimental or control condition using a stratified random procedure.

Apparatus

An EyeLink 1000 eye-tracker (SR Research, Ottawa, Ontario, Canada) monitored the gaze location of participants' dominant eye during reading. The sampling rate was set at 500Hz. Participants viewed the stimuli binocularly on a monitor 60 cm away from their eyes; a chin and forehead rest was used to minimize head movements.

Materials

Mindfulness Exercise

A pre-recorded, 10-minute long, audio-based guided meditation from the *Waking Up*[1] application was played for the participants in the experimental condition. The specific meditation used in the experiment consisted of welcoming the participant, instructing the participants on how to direct and maintain their attention on their breathing, allowing the participant to practice on his/her own, and reminding the participant of the task (maintaining their attention on their breathing) every 2 to 3 minutes.

Participants in the control condition proceeded straight to the eye-tracker calibration routine and subsequent reading task.

Experiment

The experimental set-up was re-created from a study by Uzzaman and Joordens (2011), which was successful in inducing and observing mind wandering episodes in subjects, and differentiating between mind wandering and focused reading episodes using eye movement

---

[1] https://www.wakingup.com/

patterns. The only addition made to the experimental set-up was the inclusion of the meditation session. The experiment was programmed in Python (v 3.10.1) using the PsychoPy environment and conducted on a computer.

Stimulus

Participants read the first 12 pages of War and Peace by Leo Tolstoy, which were presented on a computer screen with a resolution of 1680 x 1050. The pages were read in a single sitting. Each participant read the text at their own pace, with a mean duration of 27.5 minutes to complete all 12 pages.

Procedure

A research assistant greeted the participants, outlined the procedure, and acquired informed consent. The participants in the experimental condition had to go through a 10-minute guided meditation. Participants in the control condition proceeded straight to the calibration of the eye tracker. Calibration involved locating the dominant eye of the participant and subsequent calibration of the eye tracker to the individual pattern of the participants' eye movements.

After calibration, participants proceeded to the reading part of the experiment. Before the showcase of the first page, participants were shown information about:

- What they are going to read
- How to switch between pages
- The description of the prompt that will appear at random intervals to ask whether the participant is zoned out
- How to answer the prompt
- The definition of mind wandering to support participants in answering the prompt accurately

Following the instructions, participants proceeded to read the 12 pages of the text at their own pace. Every 2 to 3 minutes (time sampled randomly from a uniform distribution) a probe would appear in the middle of the screen asking:

“ARE YOU ZONED OUT AT THIS PARTICULAR MOMENT?
Y/N”

The participants had to press the respective “Y” or “N” keys on the keyboard to answer the probe. The participants were unable to continue reading without answering the probe. The probe could only appear 20 seconds after the participants have shifted to the page, as a new page would refresh attention.

Variables

Eye-tracking

Six standard eye-movement variables were collected: fixation duration, fixation start (time), fixation end (time), and fixation position (X & Y coordinates). Whilst previous studies (Reichle et al., 2010; Uzzaman & Joordens, 2011) included additional non-standard variables such as within-word regressions and run-count to measure mind wandering, the general consensus is that the fixation duration and fixation count is a sufficient amount of eye-tracking metrics needed to assess mind wandering episodes.

Results

Behavioural measures

Table 1 shows descriptive behavioural measures for each participant: whether the participant was in the control or mindfulness group, the number of probes received, and the number of probes that caught the participant zoning out (probe-caught zone outs). In order to adjust for the difference between the total number of probes shown to the

Table 1: *Behavioural measures—the probe-caught ratio is calculated by dividing the number of probe-caught zone outs by the number of probes*

Participant ID	**Condition**	**Number of probes**	**Number of probe-caught zone outs**	**Probe-caught ratio**
1	Control	16	7	0.44
2	Control	12	3	0.25
3	Mindfulness	9	2	0.22
4	Mindfulness	6	4	0.66
5	Control	12	1	0.08
6	Mindfulness	8	0	0
Mean	-	*10.50*	*2.88*	*0.275*
Standard deviation	-	*3.56*	*2.48*	*0.24*

participant, the ratio between the number of probes and probe-caught zone outs is provided.

Eye movement measures

Measurements of participants' eye movements were filtered to include only the periods of 12 seconds preceding the answer to the probe until 2 seconds preceding the answer to the probe. The reason behind excluding the 2 seconds before the showcase of the probe is that we employ a conservative estimate that it might take up to 2 seconds to reflect about whether one was mind wandering and answer the probe. Moreover, previous research (Reichle et al., 2010) suggests that significant differences in eye movement patterns between mind wandering and normal reading episodes go as far back as 120 seconds prior to the display of the probe. One period includes multiple individual fixations with fixation coordinates, start time, end time, and total duration for each respective fixation. Furthermore, periods were categorized into either control or experimental conditions.

The two variables analyzed were fixation duration and fixation count. Whilst, as previously mentioned, other studies have used other standard and non-standard variables, fixation duration and fixation count have been shown to robustly differentiate episodes of mind wandering from normal reading (Foulsham et al., 2013b; Rayner, 1998; Reichle et al., 1998; Uzzaman & Joordens, 2011a).

All periods preceding probes answered as "N" (as in "not mind wandering at this particular moment") were filtered out, as this study aims to look at mind wandering episodes only. Thus, we are left with mind wandering episodes from both control and experimental groups.

We used a mixed-effects linear regression model to analyze the differences between the two.

To answer *H1*, a linear mixed effects regression model predicting *fixation count* from condition (control or mindfulness) was constructed. ID was included as a random intercept to account for individual differences. The syntax of the described model was the following:

Fixation Count ~ Condition + (1|ID)

A significant effect on participants' fixation count by condition, in the hypothesized direction, was found: b = 0.19 (SE = 0.09), z = 1.99, p < .05. The relationship is plotted in Figure 2.

To answer *H2,* a nearly identical linear regression model pre dicting *fixation duration* from the condition was constructed. syntax was the following:

Fixation Duration ~ Condition + (1|ID)

An effect in the hypothesized direction, approaching significance, was found by condition on participants' fixation duration, b = -0.14 (SE = 0.1341), t = -1.1, p > .05. The relationship is plotted in Figure 3.

Discussion

In this pilot study, we investigated the effects mindfulness meditation has on mind wandering episodes during reading. Mind wandering has been dominantly treated as a dichotomy, where attention is either decoupled from the external stimuli or not. Subsequently, the effects of potential interventions for mind wandering have been evaluated in terms of the number of occurrences of mind wandering episodes. The central aim of this study is to showcase eye-tracking as a viable method for investigating the continuous nature of mind wandering and to provide an in-depth look at the effects mindfulness meditation has on mind wandering.

This was achieved by first recreating an experimental set-up by Uzzaman and Joordens (2011) to induce and detect mind wandering episodes in participants during reading. Mind wandering episodes

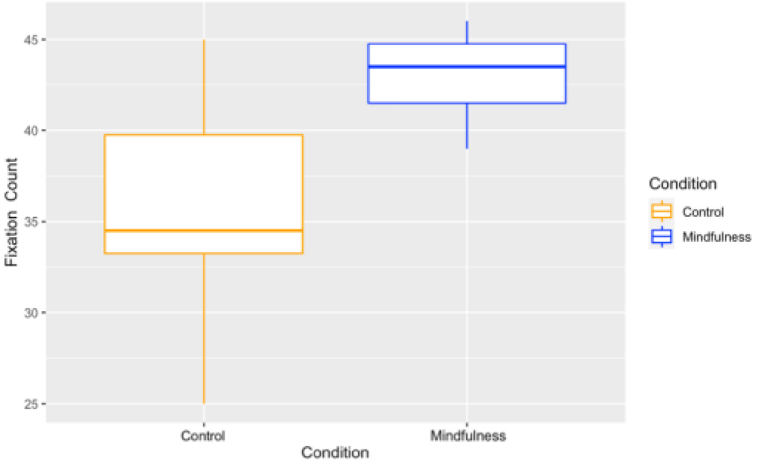


Figure 2. *A boxplot of fixations counts during mind wandering episodes in the two conditions. Condition 1 is the control condition, meaning that no meditation session was performed prior to the experimental task. Condition 2 is the mindfulness condition, meaning that a 10-minute guided meditation session was carried out prior to the experimental task. The Y-axis indicates the number of fixations made during the 10-second probe period (as described earlier in this section).*

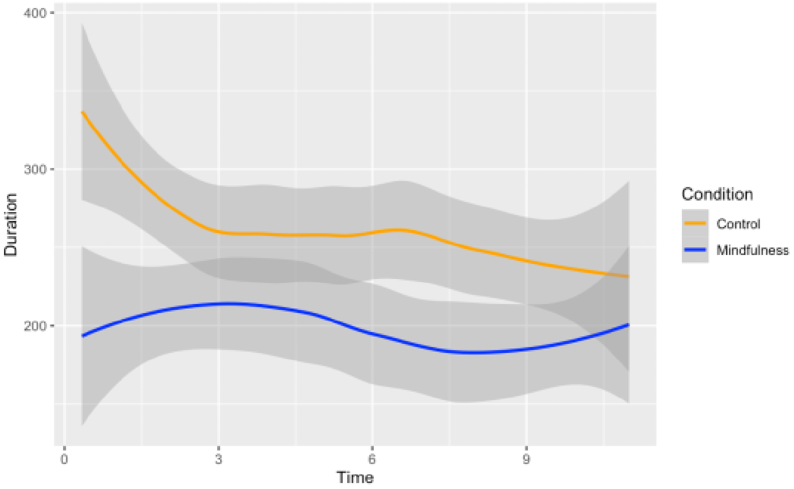


Figure 3. *Time series of fixation duration throughout mind wandering episodes in the two conditions. The X-axis is the time in seconds of the probe period, while the Y-axis specifies the fixation duration measured in milliseconds. The grey areas enwrapping the graphs of the two conditions are the associated confidence intervals of the fixation duration.*

were detected using the probe-caught method where participants are sampled at semi-random intervals about whether they are currently mind wandering or not. One limitation of the study is the lack of control for the expectancy effect in regard to the occurrence of the probes. Even though this experimental set-up has been successfully used in the past (Uzzaman & Joordens, 2011), a debrief with one of the participants revealed that the expectation of the appearance of the probes impacted the reading experience, negatively impacting the ecological validity of the experimental set-up. Participants were divided into control and mindfulness groups, where the mindfulness group completed a short, mindfulness meditation prior to the task. Finally, eye-tracking was used to analyze periods of 10 seconds prior to participants answering affirmatively to the probe.

Another limitation of the study is that the procedure used for inducing mind wandering was not validated. In order to validate the procedure, the 10s periods preceding probes which were answered as "N" by participants, meaning they did not mind wander at the time of the probe, would need to be compared to 10s periods of probes where the answer was affirmative. The procedure was, however, validated in previous studies (Uzzaman & Joordens, 2011).

It was predicted that the eye movement behaviours typical for mind wandering—increased fixation duration and decreased fixation count—would be of lesser magnitude for participants in the mindfulness condition.

Mindfulness for mind wandering

It was found that fixation count during mind wandering episodes in the mindfulness group was significantly greater than in the control group, supporting our first prediction (*H1*). In addition, fixation duration was found to be shorter in the mindfulness group, with the effect approaching significance, thus supporting to a certain extent our second prediction (*H2*). Both eye movement behaviour traits associated with mind wandering—increased fixation duration and decreased fixation count—were less extreme in the mindfulness condition, indicating the present effect of mindfulness.

These findings of reduced mind wandering as measured by eye movement behaviours are congruent with accounts that mindfulness training leads to reduced activation of the default mode network, an area often associated with mind wandering (Christoff et al., 2009). It is known that attentional and oculomotor processes are tightly integrated at the neural level (Corbetta et al., 1998). However, the exact relationship between the default mode network and eye movement behaviours remains unclear. Future work on neural correlates of mind wandering should take these into consideration and investigate how they interact with default mode network activity. Reduced activation of the default mode network was observed in long-term meditators and individuals who completed a two-week mindfulness training (Brefczynski-Lewis et al., 2007; Tang et al., 2009). Our results showcase that reduction in mind wandering can be achieved with as little as 10 minutes of meditation prior to the task. Future research should validate the results by directly measuring the activity of the default mode network post the 10-minute guided meditation session. If the efficacy of such a short mindfulness exercise is demonstrated, it would be important to compare the effects of a short mindfulness exercise versus a long one. One could imagine that mindfulness and gains associated with it follow a power law; each unit of gain requires progressively more effort or time. Comparing mindfulness exercises of different durations would potentially indicate the optimal duration for the practice. It is important to add that the inquiry suggested above would require a specific task context (such as working memory (Mrazek et al., 2013), emotional information processing

(Pavlov et al., 2015), or affective experience (Jha et al., 2010)), as we can not expect the effects to exactly be the same for tasks that engage different cognitive processes.

One obvious limitation of this pilot study was its small sample size of 6 participants, which is not enough to reliably determine whether the observed behaviours would generalize across the general population.

Approaching mind wandering continuously

Mind wandering is dominantly treated as a dichotomy and, subsequently, measured in a binary manner (Christoff, 2012; Levinson et al., 2012; Reichle et al., 2010; Smallwood & Schooler, 2015; Uzzaman & Joordens, 2011). Recently, graded and/or continuous models of mind wandering have been put forward (Mittner et al., 2016; Schad et al., 2012). New models, however, require new methods. When viewing mind wandering in terms of, for example, intensity, established binary self-reports of mind wandering are insufficient. Eye-tracking is a method perfectly suited to fill this gap. It is relatively inexpensive with high ecological validity and has been used extensively in mind wandering research. Albeit, mostly to research eye movement patterns of mind wandering when compared to non-mind wandering episodes. This pilot study suggests that the same behaviours can be used to compare mind wandering episodes to one another. Future research could rely on this method to evaluate other interventions for mind wandering, such as intelligent interfaces (D'Mello et al., 2016), mindfulness therapy with a specific focus on acceptance practice (Rahl et al., 2017), or to further investigate the link between depression and mind wandering (Killingsworth & Gilbert, 2010).

Efficacy of an online, single-session meditation

This study also makes a contribution to the question of whether online, app-based meditation sessions are effective, and specifically whether single sessions are effective in evoking increased mindfulness in participants.

While the efficacy of guided meditations by *Waking Up* have not been previously assessed, the specific meditation used for the purposes of the experiment (attention-to-breathing) is the primary meditation type used in Headspace, another mindfulness-based application whose efficacy has been shown in prior studies (Morrison Wylde et al., 2017; H. Taylor et al., 2022; Yang et al., 2018).

However, recently, a study by Mohd Zahid Juri (2022) showed no improvement in attention regulation in subjects who underwent a single session of guided meditation by Headspace.

Further investigation into the efficacy of single, guided meditation sessions is required to make confident conclusions on the matter.

Conclusion

This study aimed to investigate how mindfulness affects mind wandering measured by online external metrics provided by the eye-tracking equipment utilized. Even though we acknowledge several limitations to our experimental set-up, this pilot study showed preliminary findings of the effects that short mindfulness practice has on mind wandering, namely that the eye-tracking measures associated with mind wandering while reading (i.e., prolonged fixation duration and reduced fixation count), were found to be less pronounced in the experimental condition, where participants completed a 10-minute guided meditation prior to the commencement of the reading. This study thus builds on top of and combines previous research on measuring mind wandering with eye-tracking technology along with the ameliorating effects of mindfulness. Furthermore, it also goes beyond the established dichotomous model of mind wandering by drawing upon new evidence that points toward a novel, continuous understanding of the ubiquitous mental phenomenon. Future investigations will possibly provide more evidence for such a theoretical standpoint, such as a fully-fledged version of this pilot study.

Closing remark

We hope you did not mind wander too much while reading this paper. If you did, then don't worry—it happens to all of us.

References

Alvarez, J. A., & Emory, E. (2006). Executive Function and the Frontal Lobes: A Meta-Analytic Review. *Neuropsychology Review, 16(1),* 17–42. https://doi.org/10.1007/s11065-006-9002-x

Antrobus, J. S. (1968). Information Theory and Stimulus-Independent Thought. *British Journal of Psychology, 59(4),* 423–430. https://doi.org/10.1111/j.2044-8295.1968.tb01157.x

Barnett, P. J., & Kaufman, J. C. (2020). Mind wandering. In *Creativity and the Wandering Mind (pp. 3–23).* Elsevier. https://doi.org/10.1016/B978-0-12-816400-6.00001-8

Bixler, R., & D'Mello, S. (2014). Toward Fully Automated Person-Independent Detection of Mind Wandering. In V. Dimi trova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, & G.-J. Houben (Eds.), *User Modeling, Adaptation, and Personalization (Vol. 8538, pp. 37–48).* Springer International Publishing. https://doi.org/10.1007/978-3-319-08786-3_4

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007a). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences, 104(27),* 11483–11488. https://doi.org/10.1073/pnas.0606552104

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007b). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences, 104(27),* 11483–11488. https://doi.org/10.1073/pnas.0606552104

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y.-Y., Weber, J., & Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences, 108(50),* 20254–20259. https://doi.org/10.1073/pnas.1112029108

Brown, K. W., Ryan, R. M., & Creswell, J. D. (2007). Mindfulness: Theoretical Foundations and Evidence for its Salutary Effects. *Psychological Inquiry, 18(4),* 211–237. https://doi.org/10.1080/10478400701598298

Callard, F., Smallwood, J., & Margulies, D. (2012). Default Positions: How Neuroscience's Historical Legacy has Hampered Investigation of the Resting Mind. *Frontiers in Psychology, 3,* 321. https://doi.org/10.3389/fpsyg.2012.00321

Cheyne, J. A., Carriere, J. S. A., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition, 15(3),* 578–592. https://doi.org/10.1016/j.concog.2005.11.009

Christoff, K. (2012). Undirected thought: Neural determinants and correlates. *Brain Research, 1428,* 51–59. https://doi.org/10.1016/j.brainres.2011.09.060

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009a). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106(21),* 8719–8724. https://doi.org/10.1073/pnas.0900234106

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009b). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106(21),* 8719–8724. https://doi.org/10.1073/pnas.0900234106

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2008). Variation in Working Memory: An Introduction. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in Working Memory (pp. 3–18).* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0001

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Van Essen, D. C., & Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron, 21(4),* 761–773. https://doi.org/10.1016/S0896-6273(00)80593-0

D'Argembeau, A., Renaud, O., & Van der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Applied Cognitive Psychology, 25(1),* 96–103. https://doi.org/10.1002/acp.1647

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35(1),* 13–21. https://doi.org/10.1016/j.intell.2006.02.001

D'Mello, S., Kopp, K., Bixler, R. E., & Bosch, N. (2016). Attending to Attention: Detecting and Combating Mind Wandering during Computerized Reading. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems,* 1661–1669. https://doi.org/10.1145/2851581.2892329

Farb, N. A. S., Anderson, A. K., Mayberg, H., Bean, J., McKeon, D., & Segal, Z. V. (2010). Minding one's emotions: Mindfulness training alters the neural expression of sadness. *Emotion, 10(1),* 25–33. https://doi.org/10.1037/a0017151

Farb, N. A. S., Segal, Z. V., Mayberg, H., Bean, J., McKeon, D., Fatima, Z., & Anderson, A. K. (2007). Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience, 2(4),* 313–322. https://doi.org/10.1093/scan/nsm030

Filmer, H. L., Griffin, A., & Dux, P. E. (2019). For a minute there, I lost myself … dosage dependent increases in mind wandering via prefrontal tDCS. *Neuropsychologia, 129,* 379–384.

https://doi.org/10.1016/j.neuropsychologia.2019.04.013

Foulsham, T., Farley, J., & Kingstone, A. (2013a). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 67(1),* 51–59. https://doi.org/10.1037/a0030217

Foulsham, T., Farley, J., & Kingstone, A. (2013b). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 67(1),* 51–59. https://doi.org/10.1037/a0030217

Fox, K. C. R., Spreng, R. N., Ellamil, M., Andrews-Hanna, J. R., & Christoff, K. (2015). The wandering brain: Meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes. *NeuroImage, 111,* 611–621. https://doi.org/10.1016/j.neuroimage.2015.02.039

Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging, 30(2),* 266–278. https://doi.org/10.1037/pag0000031

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the Wandering Mind: Pupillometry of Spontaneous Thought While Reading. *Quarterly Journal of Experimental Psychology, 66(12),* 2289–2294. https://doi.org/10.1080/17470218.2013.858170

Franklin, M. S., Mrazek, M. D., Broadway, J. M., & Schooler, J. W. (2013). Disentangling decoupling: Comment on Small wood (2013). *Psychological Bulletin, 139(3),* 536–541. https://doi.org/10.1037/a0030515

Frey, M. C., & Detterman, D. K. (2004). Scholastic Assessment or g?: The Relationship Between the Scholastic Assessment Test and General Cognitive Ability. *Psychological Science, 15(6),* 373–378. https://doi.org/10.1111/j.0956-7976.2004.00687.x

Gable, S. L., Hopper, E. A., & Schooler, J. W. (2019). When the Muses Strike: Creative Ideas of Physicists and Writers Routinely Occur During Mind Wandering. *Psychological Science, 30(3),* 396–404. https:/doi.org/10.1177/0956797618820626

Galera, C., Orriols, L., M'Bailara, K., Laborey, M., Contrand, B., Ribereau-Gayon, R., Masson, F., Bakiri, S., Gabaude, C., Fort, A., Maury, B., Lemercier, C., Cours, M., Bouvard, M.-P., & Lagarde, E. (2012). Mind wandering and driving: Responsibility case-control study. *BMJ, 345(dec13 8),* e8105–e8105. https://doi.org/10.1136/bmj.e8105

Gard, T., Taquet, M., Dixit, R., Hölzel, B. K., de Montjoye, Y.-A., Brach, N., Salat, D. H., Dickerson, B. C., Gray, J. R., & Lazar, S. W. (2014). Fluid intelligence and brain functional organization in aging yoga and meditation practitioners. *Frontiers in Aging Neuroscience, 6,* 76. https://doi.org/10.3389/fnagi.2014.00076

Garrison, K., Santoyo, J., Davis, J., Thornhill, T., Kerr, C., & Brewer, J. (2013). Effortless awareness: Using real time neurofeedback to investigate correlates of posterior cingulate cortex activity in meditators' self-report. *Frontiers in Human Neuroscience, 7,* 440. https://doi.org/10.3389/fnhum.2013.00440

Gazzaniga, M. S. (2009). *The cognitive neurosciences.* MIT press.

Giambra, L. M. (1989). Task-unrelated thought frequency as a function of age: A laboratory study. *Psychology and Aging, 4 (2),* 136–143. https://doi.org/10.1037/0882-7974.4.2.136

Good, D. J., Lyddy, C. J., Glomb, T. M., Bono, J. E., Brown, K. W., Duffy, M. K., Baer, R. A., Brewer, J. A., & Lazar, S. W. (2016). Contemplating Mindfulness at Work: An Integrative Review. *Journal of Management, 42(1),* 114–142. https://doi.org/10.1177/0149206315617003

Grant, J. A., Courtemanche, J., & Rainville, P. (2011). A non-elaborative mental stance and decoupling of executive and pain-related cortices predicts low pain sensitivity in Zen meditators. *PAIN®, 152(1),* 150–156. https://doi.org/10.1016/j.pain.2010.10.006

Gruberger, M., Simon, E., Levkovitz, Y., Zangen, A., & Hendler, T. (2011). Towards a Neuroscience of Mind-Wandering. *Frontiers in Human Neuroscience, 5,* 56. https://doi.org/10.3389/fnhum.2011.00056

Hasenkamp, W., & Barsalou, L. (2012). Effects of Meditation Experience on Functional Connectivity of Distributed Brain Networks. *Frontiers in Human Neuroscience, 6,* 38. https://doi.org/10.3389/fnhum.2012.00038

Hawkins, G. E., Mittner, M., Boekel, W., Heathcote, A., & Forstmann, B. U. (2015). Toward a model-based cognitive neuroscience of mind wandering. *Neuroscience, 310,* 290–305. https://doi.org/10.1016/j.neuroscience.2015.09.053

Jha, A. P. (2002). Tracking the time-course of attentional involvement in spatial working memory: An event-related potential investigation. *Cognitive Brain Research, 15(1),* 61–69. https://doi.org/10.1016/S0926-6410(02)00216-1

Jha, A. P., Krompinger, J., & Baime, M. J. (2007). Mindfulness training modifies subsystems of attention. *Cognitive, Affective, & Behavioral Neuroscience, 7(2),* 109–119. https://doi.org/10.3758/CABN.7.2.109

Jha, A. P., Stanley, E. A., Kiyonaga, A., Wong, L., & Gelfand, L. (2010). Examining the protective effects of mindfulness training on working memory capacity and affective experience. *Emotion, 10(1),* 54–64. https://doi.org/10.1037/a0018438

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87 (4),* 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow Fluctuations in Attentional Control of Sensory Cortex. *Journal of Cognitive Neuroscience, 23(2),* 460–470.

https://doi.org/10.1162/jocn.2010.21443

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For Whom the Mind Wanders, and When: An Experience-Sampling Study of Working Memory and Executive Control in Daily Life. *Psychological Science, 18(7),* 614–621. https://doi.org/10.1111/j.1467-9280.2007.01948.x

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131(1),* 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Killingsworth, M., & Gilbert, D. (2010). A Wandering Mind Is an Unhappy Mind. *Science, 330(6006), 932.*

Klinger, E. (1984). A consciousness-sampling analysis of test anxiety and performance. *Journal of Personality and Social Psychology, 47(6),* 1376–1390. https://doi.org/10.1037/0022-3514.47.6.1376

Klinger, E. (2009). Daydreaming and fantasizing: Thought flow and motivation. In *Handbook of imagination and mental simulation (pp. 225–239).* Psychology Press.

Klinger, E., & Cox, W. M. (1987). Dimensions of Thought Flow in Everyday Life. *Imagination, Cognition and Personality, 7 (2),* 105–128. https://doi.org/10.2190/7K24-G343-MTQW-115V

Kopp, K., D'Mello, S., & Mills, C. (2015). Influencing the occurrence of mind wandering while reading. *Consciousness and Cognition, 34,* 52–62. https://doi.org/10.1016/j.concog.2015.03.003

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., & Jung-Beeman, M. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia, 46(1),* 281–291. https://doi.org/10.1016j.neuropsychologia.2007.07.013

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., & Jung-Beeman, M. (2006). The Prepared Mind: Neural Activity Prior to Problem Presentation Predicts Subsequent Solution by Sudden Insight. *Psychological Science, 17(10), 882–890. https://doi.org/10.1111/j.1467-9280.2006.01798.x*

Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The Persistence of Thought: Evidence for a Role of Working Memory in the Maintenance of Task-Unrelated Thinking. *Psychological Science, 23(4),* 375–380. https://doi.org/10.1177/0956797611431465

Manna, A., Raffone, A., Perrucci, M. G., Nardo, D., Ferretti, A., Tartaro, A., Londei, A., Del Gratta, C., Belardinelli, M. O., & Romani, G. L. (2010). Neural correlates of focused attention and cognitive monitoring in meditation. *Brain Research Bulletin, 82(1–2),* 46–56. https://doi.org/10.1016/j.brainresbull.2010.03.001

Mason, M. F., Norton, M. I., Horn, J. D. V., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science.* https://doi.org/10.1126/science.1131295

McKiernan, K. A., D'Angelo, B. R., Kaufman, J. N., & Binder, J. R. (2006). Interrupting the "stream of consciousness": An fMRI investigation. *NeuroImage, 29(4),* 1185–1191. https://doi.org/10.1016/j.neuroimage.2005.09.030

McMillan, R., Kaufman, S., & Singer, J. (2013). Ode to positive constructive daydreaming. *Frontiers in Psychology, 4,* 626. https://doi.org/10.3389/fpsyg.2013.00626

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141(2),* 302–320. https://doi.org/10.1037/a0025250

Mittner, M., Hawkins, G. E., Boekel, W., & Forstmann, B. U. (2016). A Neural Model of Mind Wandering. *Trends in Cognitive Sciences, 20(8),* 570–578. https://doi.org/10.1016/j.tics.2016.06.004

Mohd Zahid Juri. (2022). *Mindful, Neurotic, or Both: Efficacy of Online Single-Session Mindfulness.*

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 67(1),* 11–18. https://doi.org/10.1037/a0031569

Morrison Wylde, C., Mahrer, N. E., Meyer, R. M. L., & Gold, J. I. (2017). Mindfulness for Novice Pediatric Nurses: Smartphone Application Versus Traditional Intervention. *Journal of Pediatric Nursing, 36,* 205–212. https://doi.org/10.1016/j.pedn.2017.06.008

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness Training Improves Working Memory Capacity and GRE Performance While Reducing Mind Wandering. *Psychological Science, 24(5),* 776–781. https://doi.org/10.1177/0956797612459659

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General, 141(4),* 788–798. https://doi.org/10.1037/a0027968

Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2012). Mindfulness and mind-wandering: Finding convergence through opposing constructs. *Emotion, 12(3),* 442–448. https://doi.org/10.1037/a0026678

Pagnoni, G. (2012). Dynamical Properties of BOLD Activity from the Ventral Posteromedial Cortex Associated with Meditation and Attentional Skills. *Journal of Neuroscience, 32(15),* 5242–5249. https://doi.org/10.1523JNEUROSCI.4135-11.2012

Pavlov, S. V., Korenyok, V. V., Reva, N. V., Tumyalis, A. V., Loktev, K. V., & Aftanas, L. I. (2015). Effects of long-term meditation practice on attentional biases towards emotional faces: An eye-tracking study. *Cognition and Emotion, 29(5),* 807–815. https://doi.org/10.1080/02699931.2014.945903

Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience, 13(1),* 25 –42.

Rahl, H. A., Lindsay, E. K., Pacilio, L. E., Brown, K. W., & Creswell, J. D. (2017). Brief mindfulness meditation training reduces mind wandering: The critical role of acceptance. *Emotion, 17(2),* 224–230. https://doi.org/10.1037/emo0000250

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98(2),* 676–682. https://doi.org/10.1073/pnas.98.2.676

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124 (3),* 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Redick, T. S., & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology, 20(5),* 713–721. https://doi.org/10.1002/acp.1224

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105(1),* 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010a). Eye Movements During Mindless Reading. *Psychological Science, 21(9),* 1300–1310. https://doi.org/10.1177/0956797610378686

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010b). Eye Movements During Mindless Reading. *Psychological Science, 21(9),* 1300–1310. https://doi.org/10.1177/0956797610378686

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday Attention: Variation in Mind Wandering and Memory in a Lecture: Mind wandering. *Applied Cognitive Psychology, 26(2),* 234–242. https://doi.org/10.1002/acp.1814

Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(1),* 85–98. https://doi.org/10.1037/xlm0000444

Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., Oberle, E., Thomson, K., Taylor, C., & Harrison, J. (2013). Mindfulness training and reductions in teacher stress and burnout: Results from two randomized, waitlist-control field trials. *Journal of Educational Psychology, 105(3),* 787–804.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35(1),* 83–92. https://doi.org/10.1016/j.intell.2006.05.004

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition, 125(2),* 179–194. https://doi.org/10.1016/j.cognition.2012.07.004

Schmidt, S. (2011). Mindfulness in East and West – Is It the Same? In H. Walach, S. Schmidt, & W. B. Jonas (Eds.), *Neuroscience, Consciousness and Spirituality (pp. 23–38).* Springer Netherlands. https://doi.org/10.1007/978-94-007-2079-4_2

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology, 94(6),* 1479–1497. https://doi.org/10.1037/a0016810

Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6(8), 339–344. https://doi.org/10.1016/S1364-6613(02)01949-6*

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences, S1364661311000878.* https://doi.org/10.1016/j.tics.2011.05.006

Seli, P., Carriere, J. S. A., Thomson, D. R., Cheyne, J. A., Martens, K. A. E., & Smilek, D. (2014). Restless mind, restless body. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(3),* 660–668. https://doi.org/10.1037/a0035260

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-Wandering With and Without Intention. *Trends in Cognitive Sciences, 20(8),* 605–617. https://doi.org/10.1016/j.tics.2016.05.010

Seli, P., Wammes, J. D., Risko, E. F., & Smilek, D. (2016). On the relation between motivation and retention in educational contexts: The role of intentional and unintentional mind wandering. *Psychonomic Bulletin & Review, 23(4),* 1280–1287. https://doi.org/10.3758/s13423-015-0979-0

Singer, J. L., & McCraven, V. G. (1961). Some Characteristics of Adult Daydreaming. *The Journal of Psychology, 51(1),* 151–164. https://doi.org/10.1080/00223980.1961.9916467

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135(1),* 94–120. https://doi.org/10.1037/a0014212

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the Brain: Mind Wandering Reduces

Cortical Analysis of External Events. *Journal of Cognitive Neuroscience, 20(3),* 458–469. https://doi.org/10.1162/jocn.2008.20037

Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an under recognized influence on educational performance. *Psychonomic Bulletin & Review, 14(2),* 230–236. https://doi.org/10.3758/BF03194057

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008a). When attention matters: The curious incident of the wandering mind. *Memory & Cognition, 36(6),* 1144–1150. https://doi.org/10.3758/MC.36.6.1144

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008b). When attention matters: The curious incident of the wandering mind. *Memory & Cognition, 36(6),* 1144–1150. https://doi.org/10.3758/MC.36.6.1144

Smallwood, J., Mrazek, M. D., & Schooler, J. W. (2011). Medicine for the wandering mind: Mind wandering in medical practice: Mind wandering in medical practice. *Medical Education, 45(11),* 1072–1080. https://doi.org/10.1111/j.1365-2923.2011.04074.x

Smallwood, J., & Schooler, J. W. (2006). The Restless Mind. *Psychological Bulletin, 132(6),* 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology, 66(1),* 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010a). Out of Mind, Out of Sight: Eye Blinking as Indicator and Embodiment of Mind Wandering. *Psychological Science, 21(6),* 786–789. https://doi.org/10.1177/0956797610368063

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010b). Out of Mind, Out of Sight: Eye Blinking as Indicator and Embodiment of Mind Wandering. *Psychological Science, 21(6),* 786–789. https://doi.org/10.1177/0956797610368063

Subramaniam, K., Kounios, J., Parrish, T. B., & Jung-Beeman, M. (2009). A Brain Mechanism for Facilitation of Insight by Positive Affect. *Journal of Cognitive Neuroscience, 21(3),* 415–432. https://doi.org/10.1162/jocn.2009.21057

Taberna, S. (2018, March 23). Sunshine [Tumblr]. *Kitten Rain.* https://hoppip.tumblr.com/post/172179523878/i-close-my-eyes-and-let-the-sun-shine-this-one

Tang, Y.-Y., Ma, Y., Fan, Y., Feng, H., Wang, J., Feng, S., Lu, Q., Hu, B., Lin, Y., Li, J., Zhang, Y., Wang, Y., Zhou, L., & Fan, M. (2009). Central and autonomic nervous system interaction is altered by short-term meditation. *Proceedings of the National Academy of Sciences, 106(22),* 8865–8870. https://doi.org/10.1073/pnas.0904031106

Tang, Y.-Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., Yu, Q., Sui, D., Rothbart, M. K., Fan, M., & Posner, M. I. (2007). Short-term meditation training improves attention and self-regulation. *Proceedings of the National Academy of Sciences, 104(43),* 17152–17156. https://doi.org/10.1073/pnas.0707678104

Taylor, H., Cavanagh, K., Field, A. P., & Strauss, C. (2022). Health Care Workers' Need for Headspace: Findings From a Multisite Definitive Randomized Controlled Trial of an Unguided Digital Mindfulness-Based Self-help App to Reduce Healthcare Worker Stress. *JMIR MHealth and UHealth, 10(8),* e31744. https://doi.org/10.2196/31744

Taylor, V. A., Daneault, V., Grant, J., Scavone, G., Breton, E., Roffe-Vidal, S., Courtemanche, J., Lavarenne, A. S., Marrelec, G., Benali, H., & Beauregard, M. (2013). Impact of meditation training on the default mode network during a restful state. *Social Cognitive and Affective Neuroscience, 8(1),* 4–14. https://doi.org/10.1093/scan/nsr087

te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence, 35(3),* 283–300. https://doi.org/10.1016/j.intell.2006.07.006

Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., & Baddeley, A. D. (1995). Stimulus-independent thought depends on central executive resources. *Memory & Cognition, 23(5),* 551–559. https://doi.org/10.3758/BF03197257

Teasdale, J. D., Proctor, L., Lloyd, C. A., & Baddeley, A. D. (1993). Working memory and stimulus-independent thought: Effects of memory load and presentation rate. *European Journal of Cognitive Psychology, 5(4),* 417–433. https://doi.org/10.1080/09541449308520128

Tomasino, B., Fregona, S., Skrap, M., & Fabbro, F. (2013). Meditation-related activations are modulated by the practices needed to obtain it and by the expertise: An ALE meta-analysis study. *Frontiers in Human Neuroscience, 6.* https://doi.org/10.3389/fnhum.2012.00346

Uzzaman, S., & Joordens, S. (2011a). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition, 20(4),* 1882–1886. https://doi.org/10.1016/j.concog.2011.09.010

Uzzaman, S., & Joordens, S. (2011b). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition, 20(4),* 1882–1886. https://doi.org/10.1016/j.concog.2011.09.010

Wadlinger, H. A., & Isaacowitz, D. M. (2011). Fixing Our Focus: Training Attention to Regulate Emotion. *Personality and Social Psychology Review, 15(1),* 75–102. https://doi.org/10.1177/1088868310365565

Yang, E., Schamber, E., Meyer, R. M. L., & Gold, J. I. (2018). Happier Healers: Randomized Controlled Trial of Mobile Mindfulness for Stress Management. *The Journal of Alternative and Complementary Medicine, 24(5),* 505–513. https://doi.org/10.1089/acm.2015.0301

Yanko, M. R., & Spalek, T. M. (2014). Driving With the Wandering Mind: The Effect That Mind-Wandering Has on Driving Performance. *Human Factors, 10.*

Zeidan, F. (2015). The Neurobiology of Mindfulness Meditation. In *Handbook of mindfulness: Theory, research, and practice (pp. 171–189).* The Guilford Press.

Zeidan, F., Martucci, K. T., Kraft, R. A., Gordon, N. S., McHaffie, J. G., & Coghill, R. C. (2011). Brain Mechanisms Supporting the Modulation of Pain by Mindfulness Meditation. *Journal of Neuroscience, 31(14),* 5540–5548. https://doi.org/10.1523/JNEUROSCI.5791-10.2011

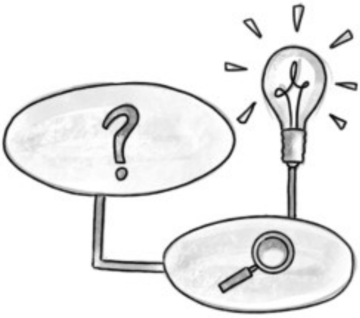
Zylowska, L., Ackerman, D. L., Yang, M. H., Futrell, J. L., Horton, N. L., Hale, T. S., Pataki, C., & Smalley, S. L. (2008). Mindfulness Meditation Training in Adults and Adolescents With ADHD: A Feasibility Study. *Journal of Attention Disorders, 11(6),* 737–746. https://doi.org/10.1177/1087054707308502

H3ad
Were to be stung by a bee
You’d no doubt yell ‘Owee’
  Certain scientists would ask
  Were you up to the task
Was that scream shaped like bouba or kiki

Forehead
That machine encased in the forehead
Meat circuits making hope and dread
  And when you’re asleep
  Is it electric sheep
Or broken consciousness leaving you dead

Head (Part 2)

Angel Martin



Analogical Insight and Recategorization

Ningxin Li

University of Toronto

Department of Psychology

Abstract

Although insight is recognized by most cognitive psychologists, the mechanisms and rationale behind its action remain unclear. Based on existing theories, this article proposes a new potential theory for understanding the role of insight. The theory suggests that insight recategorizes problems by discovering analogies. Correct categorization can lead to new and useful information, bring cognitive agents to effective search domains, and help find the right solution. Among this, prior experience and training are necessary for correct categorization. Lack of prior experience or relevant experience may result in the inability to perform effective recategorization and thus may fail to bring correct answers. Categorization theory also fits in with other theories. It can explain the difference between insight problems and non-insight problems and can be proved by several classical experiments to explain different phenomena.

Keywords: **analogies, insight, recategorization.**

Introduction

Most people have experienced in their lives that, after struggling with a problem for a while, they suddenly had a flash of inspiration and noticed a new way to find the solution. Although it is hard to describe in words, psychologists have given this "AHA!" phenomenon a name: insight. To be specific, insight is a sudden, unexpected thought that positively affects problem-solving, making the problem solver feel easeful and confident about an otherwise intractable problem (Topolinski & Reber, 2010). At the same time, problems that cannot be worked out using general or intuitive strategies and require insight to solve them are called insight problems (Shen et al., 2016).

Scholars have reached a relatively consistent view on the definition of insight, but there are still disputes over the details and mechanism behind the insight. Gestalt psychologists like Kohler (1929) believe that insight is like "looking into" a problem, which means paying attention to the problem and restructuring it. Maier (1931) proposed that when dealing with the problem, the relevant prior knowledge will transfer to the current problem and benefit the problem-solving process, which implies that insight might be related to previous experience. However, Karl Dunker (1945) believed that fixation is the blinding effect of past experience and prevents us from solving the problem, and the center of insight is how to reconstruct or formulate the goal or problem. Insight is an essential part of problem-solving and is often associated with innovation (Kheirandish & Mousavi, 2018). A better understanding of insight will not only help us understand general problem-solving processes but may also potentially improve the likelihood of innovation or the ability to solve complex problems. As for how insight leads people to the correct solution, this essay proposes a concept named categorization to explain how insight works after reviewing all of these theories. Notably, categorization is a long-standing concept in cognitive psychology. If we can show that insight and categorization are somehow related, then we can transfer the vast amount of research that already exists on categorization to the insight problem and thus discover more potential explanations.

The definition of categorization here is similar but has subtle differences with the object categorization. In object categorization, people can put things with similar properties into the same category according to their perception and predict other properties through the category (Rosch et al., 1976). Cognitive agents may solve problems in a similar way. For example, by obtaining additional information about the solution by categorizing problems. I claim a thesis that analogical insight can help cognitive agents recategorize difficult problems and search for relevant information and correct solutions in the correct categorization. This article will verify the existence of categorization in problem-solving based on several classical experiments and give a potential explanation of how it relates to insight. Before starting the verification, I will first comb through the relevant content from past studies and explain the complete process of categorization.

Literature Review of Categorization and Problem-Solving

Categorization in Solving Non-Insight Problems

Think back to the math tests that students complete in school. Rather than simply asking what 20 times 3 equals, the question is more likely to ask: If there are twenty students in the class and each student has three oranges, how many oranges are there in total in class? Teachers around the world tend to use semantic alignment to help students understand mathematical problems, using heuristics to help students relate situational models to mathematical models (Tyumeneva et al., 2017). The use of heuristics means that students can solve problems with limited time and information, using conventions or standard routines from the past (Kheirandish & Mousavi, 2018). In Tyumeneva et al.'s study (2017), mathematics textbooks would use discrete objects, such as marbles, to represent integer problems, and objects that can be continuous, such as water and temperature, to represent decimal and fraction problems. When objects with semantically symmetric relationships, such as tulips and roses, are mentioned, people are more likely to associate them with addition and subtraction problems. In contrast, when objects that have semantically asymmetric relationships, such as tulips and vases, are mentioned, people are more likely to associate them with multiplication and division problems. Although such mathematical problems are non-insight problems, they are relevant for analyzing the solution of general problems. We can conclude from these examples that the semantic information in the questions can help one categorize the problems by analogy.

The benefits of categorizing problems are similar to that of categorizing objects. Just as categorizing objects allows for predicting their properties (Rosch et al., 1976), categorizing problems allows for having more information relevant to solving them. For example, categorizing a problem under a semantic scenario model as a multiplication problem allows cognitive agents to be accessed to multiplication tables, multiplication exchange laws, multiplication distribution laws, and other laws or solution techniques. This theory

presents a pattern that is similar to the information differentiation in object categorization and recognition. In object categorization, the categorization can be classified into different levels. If an object is at the basic level, then it is likely to have a superordinate level and some subordinate levels (Murphy & Brownell, 1985). For example, the superordinate level of a chair could be furniture, while the subordinate level could contain barstools, quad chairs, rocking chairs, gaming chairs, and so on. The different levels of categorization contain different information. The subordinate level has more specific information and can list more attributes to the object, while the superordinate level contains more information that distinguishes it from other categories, but also makes the subordinate categories less distinctive (Murphy & Brownell, 1985). Therefore, if problems are miscategorized, the information that can help solve them will not be obtained and processed correctly, and the problem will become difficult. If this theory is applied to the insight problem, the reason why it is difficult to solve may be due to the wrong categorization. I will provide more details and discuss this hypothesis further in the subsequent parts of this article.

Categorization in Solving Insight Problems

Reviewing analogical insight is a good starting point to understanding how insight can help achieve ingenious categorization and bring the cognitive agents to the right answer. In many experiments which examine insight, scholars have found that participants rarely transfer relevant information to an insight problem without explicit prompts (Needham & Begg, 1991). The lack of spontaneous analogical transfer may lead to the inability of cognitive agents to put valid information and insightful questions into the same category, and therefore the problem cannot be solved correctly.

Two classic cases of insight can prove this thesis. The first case is Duncker's radiation problem (Duncker, 1945). In this scenario, doctors need to treat tumors with laser radiation, but the needed laser power can also damage other normal tissues. The medical problem about radiation is difficult for most people to address because few of them have information about radiation and medicine. However, a story of military action could cause insight (Gick & Holyoak, 1980). When attacking a fortress, the entire army cannot follow the same path. The army was therefore divided into small groups, which marched along different roads and eventually met up at the target. After providing hints about the military problem of the army occupation of the fortress, the participants were able to draw an analogy between the radiation problem and the military problem and solve it (Gick & Holyoak, 1980). In the process, participants likely placed both problems into the same category and therefore received more information about the dispersed transport. Another example is Rutherford's discovery of the nucleus. Although other theories already existed, Rutherford's pioneering analogy between the hyperbolic motion path of a particle and the motion path of a comet led to a theory that was later verified to be correct (Dietrich, 2010). Where most people are unable to make the connection between atoms and comets, Rutherford was able to see the two problems as the same category and apply the known information of comets to the atomic problem. This unique analogical insight made it possible for him to crack the puzzle. Some researchers have shown that analogy and categorization are highly similar and that they both contain similarity-based migration mechanisms (Ramscar & Yarlett, 2003). There are also studies that unify the two, arguing that analogy-making is a type of categorization, while spontaneous analogy generates insight and new knowledge (Dietrich, 2010). Thus, the argument that what insight brings to a cognitive agent may be a new understanding of problem categorization, which in turn helps to solve the insight problem, is reasonable. Next, I will apply this theory in several classical experiments.

Application of Categorization in Classic Insight Problems

Changing Categorization in Mutilated Checkerboard Problem

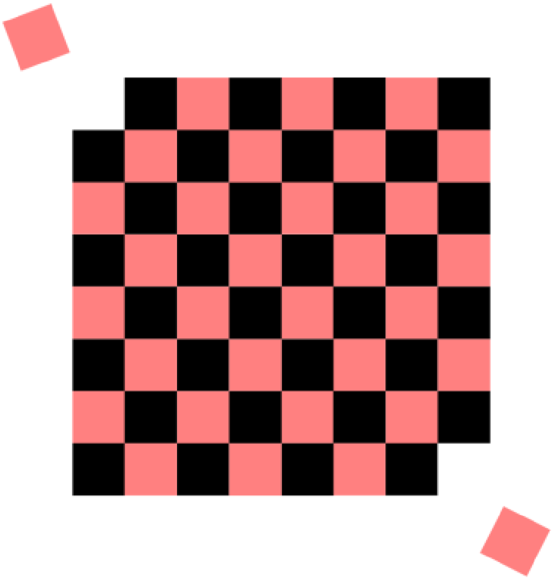


Figure 1. *The mutilated checkerboard problem.*

In the mutilated checkerboard (MC) problem experiment, the two opposite corners of the 8 x 8 board were removed, and subjects were asked to cover the entire board with rectangular dominoes which could cover two adjacent squares (Kaplan & Simon, 1990). If this was not possible, subjects needed to explain the reason. In the experiment, the researchers gave four representations of the board to manipulate the cues. The four representations included a completely blank board, a board colored in black and white, a board marked with "black" or "pink" on adjacent squares, and a board marked with "bread" or "butter" on adjacent squares. As the researchers predicted, the bread-and-butter hint group solved the problem the fastest, followed by the black-and-pink group, the black-and-white group, and the blank group. The researchers believed that the participants were inspired by the problem setting and the experimental cues, which made them realize that two adjacent squares of the checkerboard represented some parity. As a result, the insight about parity helped participants extend their searching space by changing the representation of the board, prompting them to solve the problem.

This explanation is reasonable, but how insight motivated subjects to change the representation of the problem and how it expanded the search domain still needs to be discussed. Categorization theory can further explain this process. Using the bread-and-butter group as an example, it is more likely that insight is what brought them to an analogy. After comparing adjacent grids to bread and butter, subjects would find it more appropriate to categorize the MC problem into the parity problem rather than into a coverage problem. In fact, how they name this category of problem is not the most important. The crucial point is that information about the parity is pulled out after the subjects have recategorized the questions. Subjects may recall knowledge about parity, such as pairwise occurrences and remainders. If the problem is not recategorized, subjects may still use the knowledge and information used to solve the covering problem to try to solve the MC problem, and then they will not be able to approach the correct solution.

Analogical insight and categorization also explain the differences created by the four sets of board representations. Since bread and butter was a more common parity imagery, it was more likely to allow for analogy and knowledge transfer, which also accelerated the rethinking of categorization. In contrast, the blank group, due to the lack of examples that could be used for analogy, made it difficult for subjects to spontaneously form analogical insights and to recategorize.

The example of the MC problem brings up an additional thought. If subjects do not have knowledge about parity, will they never be able to solve the problem? Intuitively, the answer is yes. Because without knowledge about the category of parity, subjects cannot classify MC problems as parity problems and cannot obtain new information. Regarding this assumption, the 9–dots problem and its training may give more support.

Changing Categorization in 9–Dots Problem

In their paper on examining fixation, Weisberg and Alba (1981) conducted a series of experiments on the 9-dots problem. This experiment gave a three-in-three nine-point diagram and asked the participants to connect all the dots with four straight lines without lifting the pen. Weisberg and Alba (1981) found that despite being given hints to get out of the square and being told that they had exhausted all possibilities in the square, the subjects still had difficulty finding the correct answer. However, if some simple training is provided to familiarize subjects in advance with how to solve concatenation problems in a non-dot-to-dot pattern, the success rate will increase. This phenomenon is consistent with the analogous insight and recategorization about parity in the MC problem. Recategorization of problems and gaining additional information was possible only after subjects had known the basics of parity and were able to see analogies in the hints. However, most people have only experienced solving dot-to-dot pattern concatenation problems, for example in a cell phone gesture password unlocking, and they lack experience with non-dot-to-dot patterns. This goes some way to explaining why training is helpful in solving insight problems. Past experience in fact provides the necessary conditions for analogy and categorization, increasing the likelihood that the subject will correctly recategorize

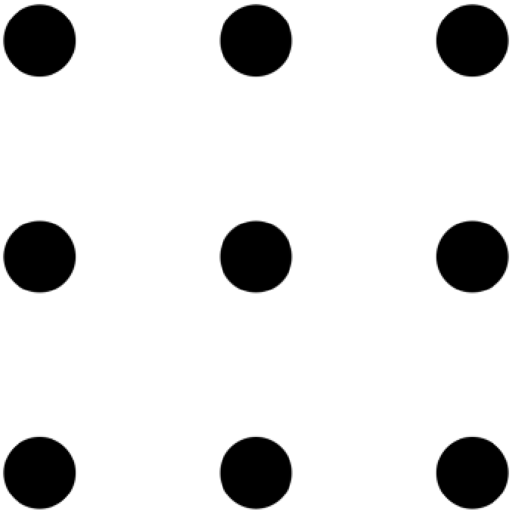


Figure 2. *The 9-dots problem.*

the problem. In their paper, Weisberg and Alba (1981) argue that removing fixation does not produce an immediate solution, and they refute the theory that fixation makes the insight problem difficult. But the delay in the emergence of a solution could also be attributed to the fact that, even if the problem is correctly categorized, the cognitive agents still need time to search for and integrate the solution under the correct categorization.

In Weisberg and Alba's follow-up experiments (1981), they also tried different training situations. They found that if the training provided was still in a dot-to-dot pattern but had the same shape as the actual problem, subjects could still improve their problem-solving skills. Conversely, if the training was in a dot-to-dot pattern but had a different shape than the actual problem, not only did it not improve problem-solving ability, but it had a negative effect. The researchers explained this by the fact that past experience is transferred to the present problem-solving strategy. This result demonstrates in another way that analogical insight allows for the recategorization of the problems. In the case of the dot-to-dot pattern with the same shape training sets, the insight brought to the subjects by past experience was not to categorize the 9-dots problems into categories beyond the dot-to-dot pattern (or beyond the square), but simply to categorize the training and actual problems as connected problems of the same shape. Thus, the insight expected by the researcher, which led to recategorization, did not occur during problem solving. Based on the same logic, in the case of the dot-to-dot pattern with the different shape training sets, subjects could neither recategorize the 9-dots problem into a beyond dot-to-dot pattern nor recategorize it into the same shape pattern as in the past experience, so the problem-solving ability decreased.

In a more recently conducted set of experiments, scholars have similarly demonstrated that the lack of a priori experience is one of the major factors in making the insight problem more difficult (Kershaw & Ohlsson, 2004). We can observe that a priori experience is necessary for resolving insights, and there are studies that show that it is also necessary for correct categorization (Bornstein & Mash, 2010). Therefore, we can demonstrate that there may be a connection between solving insight problems and categorization.

Changing Categorization and the "Small-World" Model

Schilling's "small-world" model and theory (2005) have similar logic to the categorization thesis. She argues that insight emerges with atypical associations in problem solving, which create representational shortcuts in the recombination and the search of problems. In her article, the recombination of problems means that the cognitive agents make a connection between two closely related ideas in order to find a solution. This connection may be unexpected, which is what insight brings to the "AHA!". And in the categorization thesis, recategorization behaves like recombination. In fact, the recategorization brought by an analogy insight is also the rapid abstraction and combination of two seemingly unrelated problems, so as to categorize the problems that cannot be distinguished at first sight into the correct category and obtain relevant information.

Several of the experimental examples used by Schilling (2005) can likewise be explained by categorization. For example, the apes were given a hoe and food was placed out of their reach. The apes were more likely to learn to use the hoe to reach the food if the researchers let them play with the stick and discovered how it functioned. Schilling's (2005) theory suggests that apes discovered the functional similarities between hoes and sticks, created associations and recombined them, and learned to solve the problem of obtaining food. And the categorization thesis could explain that apes gained analogical insight in using sticks and solved the problem by dividing the hoe and the food problem into a category which was used to reach the food. Other examples can also be explained using similar logic. Thus, the small-world model and categorization thesis can be viewed as different interpretations of the same set of logics, and both are valuable.

Connections to Other Theories

The categorization thesis can also fit in well with other theories. Firstly, most of the problems that are difficult to be solved belong to ill-defined problems. Ill-defined problems may have several drawbacks, such as unclear definitions of the initial state or goal state, or unclear potential pathways. However, the recategorization can transform the ill-defined problem into a well-defined problem. For example, in the MC problem, it is challenging for participants to identify all potential pathways to solve the problem. Also, they don’t know whether the answer to the problem is possible or impossible, so the goal state is also not clear. But after mastering the details of parity, participants narrowed down their pathway of actions and realized that the goal state is to prove the parity of the board. At this point, the MC problem becomes well-defined.

Second, some scholars argue that experts who are proficient in a particular field may have a more difficult time solving insight problems in that field because they are bound by their fixed experiences (Schilling, 2005). This is consistent with the characteristics of categorization. As mentioned above, categorization includes many different levels. Experts usually have a very detailed knowledge of their field, so they may tend to focus on the subordinate level rather than the superordinate level when categorizing problems. Thus, despite having accurate and detailed knowledge, it is more difficult for experts to see similarities between some categories, and make it more difficult to gain analogical insight.

Analogical insight and categorization theory can likewise be supported by physiological findings. It has been shown that, based on fMRI and EEG, sudden insights emerge when the resolver sees previously unseen connections (Bowden et al., 2005). This "connection" is likely to represent the occurrence of association, which is evidence for the existence of analogical insight. Both the small-world model and categorization thesis explain the process by which such associations occur, which is by relating features of other problems or things to the problem at hand. Recategorization, then, is likely to occur when dealing with insight problems.

Categorization theory can also explain the difference between insightful and non-insightful questions. As demonstrated in the 9-dots problem, training improves the ability of cognitive agents to solve insight problems and reduces reaction time (Weisberg & Alba, 1981). Automation may occur if a lot of training is repeated (Schilling, 2005). This automated phenomenon in fact transforms insight questions into non-insight questions specific to individuals, such as students who can automatically relate water and marbles to the characteristics of decimals and integers after a few years of math classes.

Conclusion and Discussion

Overall, the thesis suggests that the insight mechanism may be related to the cognitive agent's categorization of the problem. In the case of non-insight problems, the cognitive agent's categorization of the problem based on previous experience is usually fast and correct. However, for insight problems, the categorization of the problem may be vague or misleading. Therefore, the intuitive categorization is likely to be wrong, making it difficult for the cognitive agent to obtain information relevant to problem-solving. Thus, in previous experiments, miscategorization caused many participants to search in domains that did not contain the correct answer, resulting in a solution that could not be found even if exhausted. At this point, the solution will emerge if participants realize that they are miscategorizing the problem and put it back into the correct category by analogy. This unconscious act of making analogies and correctly categorizing them may explain the generation of insight.

It is noteworthy that each cognitive agent may have different methods and strategies for categorizing problems, and this feature is very similar to the logic by which people categorize objects. Experts in a particular problem category are more likely to focus on the subordinate level rather than the superordinate level when categorizing problems. This is because they have too much knowledge about their familiar categorization, making it harder to detect errors in their superordinate level of categorization. It is also more difficult for them to gain analogical insight and change their problem categorization. On the other hand, if cognitive agents lack knowledge of the correct category to which the problem belongs, they cannot recategorize the problem and obtain more valid information through the new categorization.

As mentioned above, the MC problem, the 9-dots problem, the small-world theory, and other theories are consistent with the categorization theory. Thus, we can tentatively conclude that the thesis that analogical insight allows cognitive agents to recategorize problems has more information about the problem and being able to solve them is potentially correct. The establishment of the categorization theory means we can transfer theories about object categorization to the study of insight to further explore the human cognitive system. This knowledge transfer can help us better understand how cognitive agents approach problems. In addition, potential benefits may arise from the study of the insight principle. For example, it has been suggested that the realization of innovation may involve transforming unknown relations into always relations (Kheirandish & Mousavi, 2018), which is similar to the theory of recategorization. If the categorization theory proves to be correct, then we may be able to improve people's learning and innovation abilities by understanding problem categorization. If the understanding of insight is applied to computer science, we may be able to improve existing algorithms. In linguistics, the word-based hints used in the previous experiments may help expose more findings on how people process words semantically.

However, more research on this theory can still be conducted in the future. For example, experiments could be conducted to demonstrate whether this recategorization exists, or more physiological studies could be conducted to explain the insight phenomenon. Possible experiments and studies include self-report and brain activity studies. For example, in subsequent experiments, researchers could ask participants to describe or record their thought processes before and after the "AHA moment" to analyze whether they are involved in categorizing the problem. We can also compare the brain regions that are active during insight problem solving with those that are active during object categorization to determine if the same brain structures are involved. If we can show that there is a process overlap between categorization and insight, then the categorization theory is likely to be valid.

References

Bornstein, M. H., & Mash, C. (2010). Experience-based and on-line categorization of objects in early infancy. *Child Development, 81(3),* 884–897. *https://doi.org/10.1111/j.1467-8624.2010.01440.x*

Bowden, E., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences, 9(7),* 322–328. *https://doi.org/10.1016/j.tics.2005.05.012*

Dietrich, E. (2010). Analogical insight: Toward unifying categorization and analogy. *Cognitive Processing, 11*(4), 331–345. *https://doi.org/10.1007/s10339-010-0367-7*

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58*(5), i–113. *https://doi.org/10.1037/h0093599*

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12(3),* 306–355. *https://doi.org/10.1016/0010-0285(80)90013-4*

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22(3),* 374–419. *doi:10.1016/0010-0285(90)90008-R*

Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 3–13. *http://dx.doi.org.myaccess.library.utoronto.ca/10.1037/0278-7393.30.1.3*

Kheirandish, R., & Mousavi, S. (2018). Herbert Simon, innovation, and heuristics. *Mind & Society, 17*(1-2), 97–109. *https://doi.org/10.1007/s11299-019-00203-6*

Köhler, W. (1929). Insight. In *Gestalt psychology. essay, Liverright.*

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12*(2), 181–194. *https://doi.org/10.1037/h0071361*

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(1), 70–84. *https://doi.org/10.1037/0278-7393.11.1.70*

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition, 19*(6), 543–557. *https://doi.org/10.3758/bf03197150*

Ramscar, M., & Yarlett, D. (2003). Semantic grounding in models of analogy: An environmental approach. *Cognitive Science, 27*(1), 41–71. *https://doi.org/10.1207/s15516709cog2701_2*

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439. *https://doi.org/10.1016/0010-0285(76)90013-x*

Schilling, M. (2005). A "small-world" network model of cognitive insight. *Creativity Research Journal, 17*(2), 131–154. *https://doi.org/10.1207/s15326934crj1702&3_2*

Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology, 9*(6), 523–538. *https://doi.org/10.1002/acp.2350090605*

Shen, W., Yuan, Y., Liu, C., & Luo, J. (2016). In search of the “aha!” experience: Elucidating the emotionality of insight problem-solving. *The British Journal of Psychology, 107*(2), 281–298. *https://doi.org/10.1111/bjop.12142*

Topolinski, S., & Reber, R. (2010). Gaining insight into the “aha” experience. *Current Directions in Psychological Science, 19* (6), 402–405. *https://doi.org/10.1177/0963721410388803*

Tyumeneva, Y. A., Larina, G., Alexandrova, E., DeWolf, M., Bassok, M., & Holyoak, K. J. (2017). Semantic alignment across whole-number arithmetic and rational numbers: Evidence from a Russian perspective. *Thinking & Reasoning, 24*(2), 198–220. *https://doi.org/10.1080/13546783.2017.1374307*

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of “fixation” in the solution of several “insight” problems. *Journal of Experimental Psychology: General, 110*(2), 169–192. *doi:10.1037/0096-3445.110.2.169*

Champion or Globetrotter? Investigating the Relationship Between Global/ Local Bias and Open-World Video Game Behaviour

Benjamin Johnson

Brock University

Department of Psychology

Abstract

Despite research identifying relationships between attention, perception, and video game behaviours, no studies have examined gameplay in relation to attentional breadth – whether one sees “the forest” or “the trees”. Drawing on attentional breadth research, I conducted an observational study investigating the relationship between global/local bias and gameplay behaviour. I hypothesized that global bias, as calculated by global scores on the hierarchical shapes task (HST), would be positively correlated with the breadth of game exploration in an open-world computer game, as calculated by the range of activity attempts, area visits, and character interactions. Twenty university students participated remotely by playing a 75-minute screen-recorded computer game. Additionally, participants completed the HST, a computerized cognitive task measuring naturally-occurring attentional breadth, followed by questionnaires assessing extraversion, openness to experience, and attitudes of horizontal/vertical individualism and collectivism. As predicted, there was a positive correlation between global bias and breadth of explorative gameplay, but the correlation was weak and not statistically significant. However, after controlling for extraversion, openness to experience, video game experience, and attitudes of horizontal/vertical individualism and collectivism, attentional breadth explained 10-20% unique variability in participants’ global and local gameplay behaviours. These results suggest that naturally-occurring attentional breadth can predict behaviour in an open-world video game, independent of certain personality traits and social attitudes..

Keywords: **attentional breadth, global/local bias, extraversion, openness to experience, individualism and collectivism, video game behaviour.**

Introduction

Human individuality is humorously conveyed by the old saying, “You are absolutely unique. Just like everyone else.”[1] In the field of psychology, this uniqueness is explored by studying individual differences. Individual differences are empirical representations of the ways by which psychological traits vary from person to person (see more in Stangor, 2012). Personality traits are commonly studied as individual differences variables (e.g., Ashton & Lee, 2001, 2007). Nonetheless, a highly broad range of traits aside from personality can be measured in individual differences studies. For example, cognitive variables such as attention and working memory have been used as individual differences measures in research studies (e.g., Nunez et al., 2015; Moosbugger et al., 2006), especially in relation to educational studies and health sciences (e.g., Kyndt et al., 2012; NICHD Early Child Care Research Network, 2005).

Global/Local Cognitive Processing

Attentional breadth is a cognitive trait that represents a person’s tendency to view “the forest or the trees”. People perceive visual stimuli on a “global” level (the forest) or on a “local” level (the trees). Specifically, global bias refers to a holistic perceptual style such that one allocates their attention more broadly to the item or scene as a whole. On the other hand, local bias refers to a perceptual style that focuses on the elements that comprise an item or scene by allocating attention more narrowly. This trait naturally differs between individuals and is measured using computerized tasks.

Attentional breadth was first studied in 1977 by Navon, who

---

[1] Although widely quoted, there is no author to whom this quotation can be originally attributed. See more at: https://quoteinvestigator.com/2014/11/10/you-unique/

constructed the Navon Letters Task (NLT), one of the most recognized measures of global/local bias (see Figure 1a for an example of a congruent and an incongruent Navon letter). Navon's research demonstrated a global precedence effect where individuals typically prioritize the global level over the local level. Individuals show faster reaction times when identifying letters that are congruent (i.e., the same letter at the global and the local levels) versus incongruent (i.e., different letters at the global and local levels), and this congruency effect increases when they are asked to report the local level than when they are asked to report the global level. The greater interference by the global level suggested that global processing received greater attentional priority.

Kimchi and Palmer (1982) constructed another measure of attentional breadth called the Hierarchical Shapes Task (HST). On each trial of the HST, participants are shown stimuli consisting of a standard hierarchical shape item above two comparison hierarchical shape items from which participants are asked to select the one they think best matches the standard (see Figure 1b). Each image was composed of smaller squares or triangles forming a larger square or triangle. One comparison item matches the standard at a local level whereas the other matches the standard at a global level. The researchers calculated global bias by taking the total number of selections that matched at the global level. Thus, a higher number of global selections indicates a higher degree of global bias. Attentional breadth as measured by the NLT and HST remains relatively stable and constant over time within an individual (Dale & Arnell, 2013). This consistency is such that some people have a consistent bias to allocate attention to the global level or the local level, while others have no bias, and this may vary with the nature of the stimuli (Poirel et al., 2008).

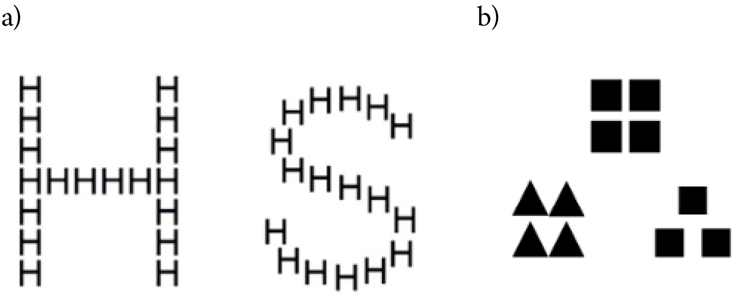


Figure 1. *a) An example from the Navon Letters Task of a congruent item on the left and an incongruent item on the right; b) An example from the Hierarchical Shapes Task of a target item (top), a global selection (left), and a local selection (right).*

Previous literature has established how attentional breadth has been associated with more than just the visual system. Rather, researchers have discovered evidence on how it relates to cognitive control and social behaviour. Taking a clinical approach, researchers have found that local bias tends to be prevalent among individuals affected by certain mental illnesses such as schizophrenia (Bellgrove et al., 2003), obsessive-compulsive disorder (Yovel et al., 2005), and autism spectrum disorder (Scherf et al., 2008). However, local bias is associated with positive attributes as well. Past experimental research has shown that experiencing a pleasant stimulus can increase local bias by narrowing attentional breadth (Gable & Harmon-Jones, 2008).

Regarding brain activity, several researchers have detected associations between cortical activity measurements and global/local processing. For example, Fink and colleagues (1996) investigated the neural underpinnings of global/local processing and found that local processing was emphasized in the left hemisphere, whereas global processing was emphasized in the right hemisphere. Extending these findings, Gable and Harmon-Jones (2008) observed that the combination of positive affect and high approach motivation was associated with not only narrow attentional breadth, but also higher left hemisphere activity in the frontal lobe. Expanding on this research, Pitchford and Arnell (2018) revealed that activation of cognitive control mechanisms during an effortful control task narrows attentional breadth in those with high approach-motivation. Specifically, when individuals express high motivational approach tendencies, they reduce their attentional breadth if exerting high self-control, and this relationship was found to be distinct from the link between left frontal lobe activity and local processing. Relating to the local bias associated with schizophrenia mentioned earlier, Choi and colleagues (2014) found that certain event-related potentials (ERPs) during local processing differ in those with schizotypal traits compared to a control group, suggesting that psychopathology influences attentional breadth as a result of differences in brain activity.

Global/Local Bias and Real-World Behaviour

Moving beyond behaviours observed in lab settings, global/local processing has also been shown to predict real-world social behaviours, such as public health behaviours. Specifically, a recent individual differences study on attentional breadth found that greater global bias can predict greater compliance with COVID-19 public health guidelines (Mendonca, 2021). This study found a positive correlation between the HST global bias score and one's compliance with public health guidelines after controlling for political affiliation and other factors. Similarly, other studies have found that asking participants to attend to the global level of hierarchical stimuli led to them making more monetary donations to charity than when participants were asked to attend to the local level of these same stimuli (Mukherjee et al., 2018). Another study showed how global bias relates to holding prosocial attitudes (Colzato et al., 2010). This study showed that greater global processing on the NLT was associated with greater compassion and prosociality self-reported through a social value orientation task. This evidence suggests that perceiving something globally may relate to acting globally. Nonetheless, the literature has not yet studied attentional breadth in relation to virtual behaviours, such as video game behaviour.

Video Game Behaviour

Video games have taken the interest of countless researchers, often to examine cognition and behaviour. One reason is because video games evolved from being purely a source of entertainment to being useful in many real-world applications. For example, specialized video games are being more frequently applied in educational settings (Hutchison, 2007), where these games can engage students and sometimes improve learning. Similarly, video games are being applied in healthcare settings (Halldorsson et al., 2021), where they are used in special therapy programs. There are a number of reasons why video game technology can provide these benefits. First, video game experiences can help promote skills such as hand-eye coordination, working memory, and other cognitive skills (Benoit et al., 2020). Second, games can provide rich learning experiences due to their interactive features and high interest among youth (Granic et al., 2014).

Video game behaviours often reflect who we are. This notion is confirmed by research on video games that shows how in-game behaviour often reflects the personality and preferences of the player (Teng et al., 2012). Zeigler-Hill and Monica (2015) studied avid video game players and correlated their HEXACO personality scores in relation to video game behaviour. Participants self-reported their attitudes, behaviours, and feelings involved with the video games they play on a regular basis. Their findings suggest that video game behaviours not only align with personality traits, but they are also consistent and predictable.

Scholars are increasingly investigating the reciprocal nature of cognition and behaviour in relation to video game experiences. Numerous studies have covered the role of players' attention in video game performance. Bavelier and Green (2003) studied how video game experience can boost cognitive traits such as selective attention. First, they noted that video game players exhibited greater accuracy compared to controls on tasks such as the attentional blink, suggesting they had more efficient attentional deployment over time. Gamers also outperformed controls on the "useful field of view task", suggesting they had more efficient attentional deployment over space. Next, they exposed non-gamers to video game training with action video games, which markedly improved their selective attention on these tasks compared to a control group who trained on the special puzzle video game, Tetris. In terms of disordered cognition, a study by Segev and colleagues (2017) examined participants' video game behaviour in relation to the presence of cognitive symptoms, finding that such symptoms correlated with externalizing behaviours in a video game. These studies demonstrate how cognitive traits can predict video game behaviour.

Doodle Champion Island Games

The video game selected for this study was "Doodle Champion Island Games" (hereafter "Champion Island"), a Google Doodle video game published on July 24th, 2021, to coincide with the Tokyo 2020 Summer Olympics.[2] In Champion Island, players explore a world filled with seven sport mini-games, hundreds of citizens, and dozens of side-quests. The sport mini games are based on sports featured in

---

[2] The updated August 24th version was used for the present study: https://www.google.com/doodles/doodle-champion-island-games-august-24

the Tokyo Summer 2020 Olympics: rock climbing, artistic swimming, archery, marathon, rugby, table tennis, and skateboarding. Champion Island's character and world design are based on concepts from Japanese culture, folklore, and mythology. Participants are free to choose where to travel, what sports to play, and other choices. The reason for selecting Champion Island was because my research seeks to involve unconventional, non-commercial video games in order for cognitive science to foster knowledge on how the relevance and cultural role of video games can be broadened beyond traditional, commercial contexts.

The Present Study

As noted above, individual differences in global/local bias, as measured with the HST, have been shown to predict compliance with real world public health behaviours such as social distancing and mask wearing during the COVID-19 pandemic (Mendonca, 2021). Also, as noted above, video gameplay reflects characteristics of the player (Zeigler-Hill & Monica, 2015). The Champion Island open-world video game is by far the most complex interactive Google Doodle created to date (Walsh, 2021), and one where participants can create their own experience through choosing which aspects of the game they complete. Therefore, the question addressed in this work was whether individual differences in global/local bias, measured by the HST, could predict Champion Island gameplay behaviours given that gameplay may relate to whether players perceive opportunities and objectives in the game primarily from a global level or from a local level. I hypothesized that attentional breadth would predict breadth of gameplay behaviour such that higher global bias on the HST would positively correlate with more globally-focused gameplay as define by a greater number of areas explored, interaction with a greater number of characters, more games played, more quests undertaken and completed. Moreover, I hypothesized that global bias on the HST would negatively correlate with locally-focused gameplay, as calculated by the participants' maximum and average number of attempts per sport. Finally, I predicted that extraversion and openness to experience would positively correlate with NPC Interactions and Total Areas Visited, respectively.

Methods

Participants

Twenty-six university students were recruited through the SONA psychology research platform using a convenience sampling approach. Six participants were removed due to various exclusion criteria (see Results), leaving seven male participants and 13 female participants (Mage = 21.3 years; SD = 2.72) for analysis. English proficiency was a requirement for participation, and was verified by the researcher. Participants were compensated with SONA research participation credit. Because this was a correlational study, all participants performed the same tasks under the same conditions. Notably, participants completed all tasks via remote virtual participation using their own devices. The study was approved by the institution's Research Ethics Board. In accordance with ethical standards, all participants provided informed consent prior to participating. Coders signed a confidentiality agreement, participated in training, and used a manual to complete their coding duties.[3]

Materials

Video Game: Champion Island

Using their own computer, all participants played the Champion Island video game individually for 75 minutes while their screen was

---

[3] This study was pre-registered, including methods and hypotheses, which can be viewed at the following link: https://aspredicted.org/blind.php?x=XG6_MPV

recorded. Coders later transcribed the recording by watching the entire gameplay recording while filling out a checklist and tally chart to track relevant behaviours (see Appendix A for checklist used by coders). The tasks were selected based on in-game progress as well as actions that relate to the psychological constructs of interest (e.g., global/local bias and individualism-collectivism). For consideration of data inclusion/exclusion, notes were taken if any interruptions occurred (e.g., internet disconnection).

Sports Participation. Seven different sports can be attempted in the game. The number of sports attempted was scored out of 7 ("Sports Attempted"). The number of times each sport was attempted was tallied and averaged across the sports that were attempted at least once such that higher numbers reflected greater attempts/sport played (AvgSportAttempts). The total number of sport attempts was also calculated (TotalSportAttempts) as was the highest number of times any single sport was played (MaxSportAttempts). A scroll was obtained for each sport after successfully winning the event the first time, and the number of scrolls earned was also recorded (ScrollsEarned). We also recorded the average number of attempts required to earn a scroll (PreScrollAttempts), and the average number of times a sport was played subsequent to earning the scroll, for example to earn a higher score (PostScrollAttempts).

Trophy/Quest Participation. The game featured numerous side-quests to complete and NPC characters with which to interact. The number of interactions with an NPC was totalled (TalkNPC). Quests could be completed for their own sake, or to earn trophies. Quests were initiated when the player engages with an NPC, but not all engagements with an NPC resulted in a quest opportunity, and not all quests were undertaken, therefore the total number of quests accepted was counted as a separate variable

(TrophiesAttempted). Completing a quest automatically resulted in earning a trophy. Therefore, the number of trophies earned (TrophiesEarned) also measured the number of completed quests.

Quests involved several steps and participants may have only completed a portion of the steps without completing the quest. The criteria for each step were defined objectively so that they were coded the exact same way for each participant. A step was defined by a single, explicit action that was taken in the game to affect progress in the quests. Accepting the quest was considered 'Step 0' because it activated the rest of the steps, but it was not considered making the same degree of progress as the other steps. The checklist for quests consisted of a positive integer that listed the number of steps completed on a specific quest by the end of the gameplay duration (TrophySteps) which was recorded for exploratory analyses. If a quest had not been accepted, then that item of the checklist was left blank. Finally, when an NPC had to be convinced of something through specific dialogue choices, then talking to them and convincing them were considered separate steps because the actions and consequences were independent.

Exploration. The coding checklist included whether each of the seven main areas were visited as well as seven side areas and six unlockable areas. How many of the 14 readily accessible areas had been visited was recorded as "TotalAreas". Coders also tallied the number of times a participant visited their Team Base if they chose a team.

Composite Global and Local Gameplay Measures. For the operational definition of global gameplay, participants' overall score of in-game global bias was estimated with a composite global measure calculated as the average of the standardized measures (z-scores) of number of different sports attempted (Sports Attempted), number of quests accepted (Trophies Attempted), number of areas visited (Total Areas), and number of NPC interactions (NPC Interactions) in that for each of these scores higher scores are thought to indicate more global bias. The main connection to the global bias processing style is that participants perceive opportunities and objectives in the game holistically. This means that they have a broad outlook on their possibilities in terms of sports and quests – examining the bigger picture – rather than fixating on specific opportunities, which resembles the local bias processing style. A composite local gameplay measure was also calculated from the average of the z-scores for max and average number of times each sport was played.

Background Questionnaire

Participants typically vary in age, gender, and task-related experiences (e.g., video games and competitive sports). Thus, a background survey was developed to collect these data points (see Appendix B) immediately prior to participants completing the attentional breadth task and the attitude and personality questionnaires.

Attentional Breadth Cognitive Task

The Hierarchical Shapes Task (HST; Kimchi & Palmer, 1982) was used to measure local global bias. In the task, participants viewed an example trial, then completed 24 experimental trials (see Figure 1b). For each trial, participants were shown three stimuli consisting of a target stimulus centred above two adjacent stimulus options. Each stimulus was composed of 3 or 4 small circles, squares, triangles, or crosses (local level) forming the shape of a larger square or triangle (global level). Participants were asked to select, based on their initial instinct, which of the two options best matched the target shape. Of the 24 trials, 12 were filler trials intermixed with 12 test trials. Filler trials had an objectively correct answer: the correct option matched the target shape at either the global or local level, and the incorrect option did not match the target at either level. The test trials included two subjectively correct options, where one option matched the target at the local level (e.g., both were made of small squares as in the right option in Figure 1b), and the other choice matched the target at the global level (e.g., both formed a larger square as in the left option in Figure 1b). For both filler and test trials, the order of options (whether global/local or the correct/incorrect appear on the right or left) changed throughout the task to minimize confounds. Selecting the option that matched the target at a local level indicates that the participant had a bias for narrow attentional processing on that trial. Conversely, selecting the global level option indicates a bias for broad attentional processing. Participants' overall HST score for global bias was calculated by adding the number of global selections on the test trials (from 0 to 12) such that a higher HST score signifies greater global bias.

HEXACO Personality Domains Questionnaire

Scores on "extraversion" and "openness to experience" from the HEXACO personality inventory (Lee & Ashton, 2007) were collected using the 10 questions from each domain (Lee & Ashton, 2009). Extraversion measures one's tendency to enjoy social interaction and hold generally positive self-esteem. Openness to experience measures an individual's fondness of unconventional ideas and novel experiences. All items were answered on a 5-point scale, ranging from strongly disagree (1) to strongly agree (5). After reverse-scoring negatively-keyed items, an average score for each of the two domains was calculated.

Horizontal-Vertical Individualism and Collectivism Questionnaire

The Horizontal-Vertical Individualism and Collectivism (HVIC; Triandis & Gelfand 1998) questionnaire measures collectivism and individualism in two forms: horizontal and vertical. The questionnaire included 4 statements for each of these categories where participants indicated the extent to which each statement accurately captured their attitude, feelings and behaviours on a 9-point scale ranging from never (1) to always (9). Scores were averaged within each subscale: horizontal individualism (HI), vertical individualism (VI), horizontal collectivism (HC), and vertical collectivism (VC).

Procedure

Data Collection

Through a Microsoft Teams meeting, the experimenter first provided a brief game tutorial for Champion Island that clarified that participants could play however they prefer. The experimenter provided the participant with the online questionnaire link along with their participant number. Prior to the experimenter leaving the video call, participants were asked to share their screen (via screen-sharing function of Teams) and turn off their audio/video during gameplay. Participants were asked to play the game for 75 minutes, then close the Teams meeting and open the link to complete the remainder of the study unrecorded using the Testable online experiment platform. The questionnaire consisted of background questions (see Appendix B), 24 trials of the HST, questions on the two HEXACO domains, followed by the HVIC questions.

Participants' data was excluded from the sample if certain disruptive events took place during the Champion Island gameplay: (a) their internet connection or computer was troubled for more than 15 minutes (non-consecutive); (b) they were inactive for more than 15 minutes (non consecutive); or, (c) they were not engaging properly, such as doing something else on their computer other than the video game. Additionally, participants' data was removed from the sample if they did not score at least 9 out of 12 on the HST filler trials where there are objectively correct answers and pass at least two out of the three attention checks embedded in the questionnaires.

Data Analysis

After data exclusions, coders carefully reviewed each gameplay recording in order to code all the variables involved in the analysis. As mentioned previously, coding the operationalized variables consisted of filling in the checklist and tracking sheet. This was a structured observational approach because the variables are predefined. Recording was done mechanically by computer software, but the coding was performed by humans. Thirty-five percent of the data files were coded by two coders so that reliability estimates could be calculated. Additionally, the correlations amongst the key global and local gameplay variables were examined to verify their suitability for inclusion in the composite measures.

To test the hypotheses, Pearson's r correlations were examined between HST global scores and HVIC scores, HEXACO scores, the composite global gameplay measure, and the composite local gameplay measure. Correlations between global and local gameplay measures and the personality measures were also examined. In addition to correlation analyses, two multiple regression analyses were conducted to examine whether attentional breadth scores on the HST could predict global and local gameplay composite measures over and above the HVIC scores, personality measures from HEXACO, and prior video game experience.

Results

Overall, six participants were removed from the dataset. Three participants failed the questionnaire attention checks, which suggested their questionnaire data were not reliable. One participant failed to obtain the minimum of 9 out of 12 correct responses on the filler trials of the HST, which suggested improper completion of the HST. One participant had previous progress on the game and their data were deemed to be compromised. The sixth removal was due to a recording glitch that lost over 15 minutes of footage.

Descriptive Statistics for Scale Measures

Descriptive statistics for the key measures of interest are listed in Table 1. The sample had considerable variability in the HST scores and extraversion, openness to experience, and HVIC questionnaire scores. All of these means fall in the mid-range of the scales, and the standard deviations and minimum and maximum values indicate considerable spread with no ceiling or floor effects. With regard to coded gameplay behaviours, all twenty participants attempted each of the seven sports, which means that all participants also visited each of the seven main areas on the game map. Hence, the two gameplay variables of Sports Attempted and Main Areas suffered from ceiling effects. On the other hand, participants showed a large range of scores for the rest of the gameplay measures, including the components of the global gameplay and local gameplay composite variables. Despite the Main Areas variable suffering from ceiling effects, Total Areas remained suitable for analysis as a component of the global gameplay composite variable because it consisted of the sum of Main Areas and Side Areas, the latter of which had sizeable variance within the sample.

Inter-rater Reliability for Coded Gameplay Variables

Gameplay data for seven participants (35%) were coded independently by two separate coders who produced two sets of scores that were examined to create a final version of scores. The intraclass correlations (ICC) were used to calculate inter-rater reliability. ICC is used as a valid measure of inter-rater reliability in research designs wherein only a subset of participants will be coded by multiple raters (Hallgren, 2012), and wherein pairs of raters change for each subject. ICC values ranged from .864 to .978 Average Sport Attempts had an

Table 1: *Descriptive Statistics for Variables of Interest*

	N	Mean	SD
HST	20	6.60	
Extraversion	20	3.05	
Openness	20	3.62	
Sports Attempted	20	7.00	
Trophies Attempted	20	10.95	
NPC Interactions	20	47.70	
Total Areas Visited	20	10.85	
Average Sport Attempts	20	4.78	
Max Sport Attempts	20	8.60	

Note. *Openness = Openness to Experience. Possible scores for each variable are the following: HST = 0 to 12; Extraversion and Openness to Experience = 1 to 5; HVIC dimensions (HI, HC, VI, VC) = 1 to 9; Sports Attempted = 0 to 7; Trophies Attempted = 0 to 24; Areas Explored = 0 to 14; NPC Interactions and Average/ Max Sport Attempts = any positive integer. Global Gameplay and Local Gameplay were calculated as the average z-score from a set of gameplay variables. As standardized variables, the mean is 0 and the standard deviation is 1. Therefore, they are not reported in this table.*

ICC of .978 (*p* < .001). Trophy Attempts had an ICC of .948 (*p* < .001). NPC Interactions had an ICC of .864 (p = .003). Side Area Visits had an ICC of .898 (*p* = .001). The greatest discrepancy between coders was on the number of NPC interactions, which was not surprising due to the fact that this variable had a wide range of scores (12 – 90 interactions based on the final version of scores). The most probable cause of the discrepancy can be accounted for by the coder not following the exact criteria for what constituted an interaction (see Methods). Because the two sets were not identical on all scores, the discrepancies needed to be addressed. For the number of NPC interactions, the average score from both sets was taken for the final version. For all other gameplay variables, discrepancies were resolved by the researcher re-evaluating the game recording and deciding on the score for the final version. After being resolved by the researcher, the reconciled values were used for the analyses.

Relationships Amongst Gameplay Measures

Sports Attempted and Main Areas Visited were not included in the global gameplay composite variable due to ceiling effects. As seen in Table 2, all remaining components from the global gameplay composite variable (Trophies Attempted, NPC Interaction, and Total Areas) had strong positive inter-correlations (*r* > .8) that were statistically significant (*p* < .001). Additionally, they all had moderate to strong negative correlations with the two components of the local gameplay composite variable (Average Sport Attempts and Max Sports Attempts). From these strong and significant positive correlations between the global gameplay composite variables, global gameplay measures demonstrated strong convergent and divergent construct validity and were deemed suitable for use in the global gameplay composite variable. Given that they were highly positively correlated with each other, and highly negatively correlated with components of the global gameplay variable, Max Sport Attempts and Average Sport Attempts were deemed suitable as the components of the local gameplay composite variable.

Relationships Between Gameplay Composites and Other Measures

Pearson's r for the global gameplay composite and HST scores showed a weak positive correlation (see Table 3) that was not statistically significant. The correlation between the local gameplay composite and HST scores was medium-strength and negative, but non-significant ($p = .112$). Global gameplay had a significant strong and negative correlation with local gameplay. HST score was highly negatively correlated with present video game experience, but this fell short of statistical significance. Additionally, past and present video game experience were non significant predictors of greater global gameplay and less local gameplay.

Table 2: *Correlation Matrix for Standardized Gameplay Variables*

	NPC Interactions	*Total Areas*	*Average Sport Attempts*	*Max Sport Attempts*
Trophies Attempted	.874**	.816**	-.696**	-.670**
NPC Interactions		.801**	-.590**	-.559*
Total Areas			-.432	-.507*
Average Sport				.895**

** $p<.01$, *$p < .05$

Regression Analysis

Video game experience and personality measures were somewhat related to global and local gameplay measures and HST scores. Therefore, controlling for this variability would allow examination of the relationship between global and local gameplay and HST over and above these measures. A simultaneous regression was conducted predicting global gameplay using the predictors of HST scores, extraversion, openness to experience, HVIC measures, past and the researcher re-evaluating the game recording and deciding on the score for the final version. After being resolved by the researcher, the reconciled values were used for the analyses.

Relationships Amongst Gameplay Measures

Sports Attempted and Main Areas Visited were not included in the global gameplay composite variable due to ceiling effects. As seen in Table 2, all remaining components from the global gameplay composite variable (Trophies Attempted, NPC Interaction, and Total Areas) had strong positive inter-correlations ($r > .8$) that were statistically significant ($p < .001$). Additionally, they all had moderate to strong negative correlations with the two components of the local gameplay composite variable (Average Sport Attempts and Max Sports Attempts). From these strong and significant positive correlations between the global gameplay composite variables, global gameplay measures demonstrated strong convergent and divergent construct validity and were deemed suitable for use in the global gameplay composite variable. Given that they were highly positively correlated with each other, and highly negatively correlated with components of the global gameplay variable, Max Sport Attempts and Average Sport Attempts were deemed suitable as the components of the local gameplay composite variable.

Relationships Between Gameplay Composites and Other Measures

Pearson's *r* for the global gameplay composite and HST scores showed a weak positive correlation (see Table 3) that was not statistically significant. The correlation between the local gameplay composite and HST scores was medium-strength and negative, but non-significant ($p = .112$). Global gameplay had a significant strong and negative correlation with local gameplay.

HST score was highly negatively correlated with present video game experience, but this fell short of statistical significance. Additionally, past and present video game experience were non significant predictors of greater global gameplay and less local gameplay.

Regression Analysis

Video game experience and personality measures were somewhat related to global and local gameplay measures and HST scores. Therefore, controlling for this variability would allow examination of the relationship between global and local gameplay and HST over and above these measures. A simultaneous regression was conducted predicting global gameplay using the predictors of HST scores, extraversion, openness to experience, HVIC measures, past and present video game experiences. As shown in Table 4, together the predictors explained 49.4% of the variability in global gameplay, $R = .703$, $F(9,10) = 1.085$, $p = .447$. Global bias as measured by the HST predicted 11.56% of unique variability in global gameplay over and above the factors of personality and video game experience (see Figure 3). This was not statistically significant, however, and there were no unique predictors.

A second simultaneous regression was conducted predicting local gameplay using the same predictors of HST, personality and video game experiences. As shown in Table 5, together the predictors explained 54.4% of the variability in local gameplay, $R = .737$, $F(9,10) = 1.087$, $p = .333$. HST Global bias predicted 22.37% of unique varia-

Table 3: *Correlation Matrix for Key Variables*

	HST	Global Play	Local Play	Extra	Openness	HI	HC	VI	VC	Past VG
HST	1									
Global Gameplay	.145	1								
Local Gameplay	–.272	–.628**	1							
Extra	.178	–.067	.294	1						
Openness	0.057	.039	.146	.318	1					
HI	–.009	–.160	.192	–.366	–.016	1				
HC	.262	.028	.231	.673**	.401	–.244	1			
VI	–.152	–.2.80	.164	–.364	–.294	.426	–.354	1		
VC	.067	–.181	–.051	.210	–.117	–.349	.401	–.053	1	
Past VG	–.091	.513*	–.435	–.040	.099	–.249	.053	–.026	0.52	1
Present VG	–.420	.327	–.192	–.293	.240	0.37	.015	–.041	.124	.507*

🟥 = Key Correlations

▨ = Other Correlations of Interest

Note. *Extra = Extraversion; Openness = Openness to Experience; Past VG = Past Video Game Experience; Present VG = Present Video Game Experience.* ** $p < .01$, * $p < .05$

bility in local gameplay over and above the factors of personality and video game experience (see Figure 4), but this was not significant, nor were any other predictors.

Discussion

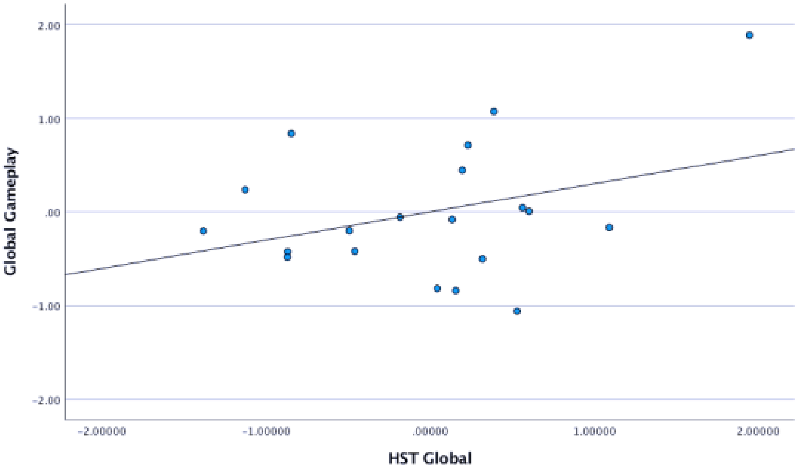
Hypothesis Testing

Attentional Breadth and Global/Local Gameplay

It was predicted that there would be a positive correlation between HST global scores and the global gameplay composite variable. The correlation was weak ($r = .145$) and not statistically significant, but was in the correct direction. Similarly, it was predicted that HST global scores could explain unique variability in global gameplay over and above extraversion, openness to experience, individualism-collectivism, and video game history. As predicted, HST had a fairly strong partial correlation with global gameplay once these measures were controlled for, but this was not statistically significant. The partial correlation of .340 was more noteworthy than the zero-order correlation of .145 because there was a high amount of variability in global gameplay explained by the other predictors (e.g., video game experience), which is excluded from the partial correlation calculation. My predictions were based on obtaining a larger sample size (approximately N = 60) with greater statistical power than the actual result. As this statistical power was not obtained, the p-values are less useful in validating the effects between the key variables of interest (e.g., HST scores and global/local gameplay behaviour measures).

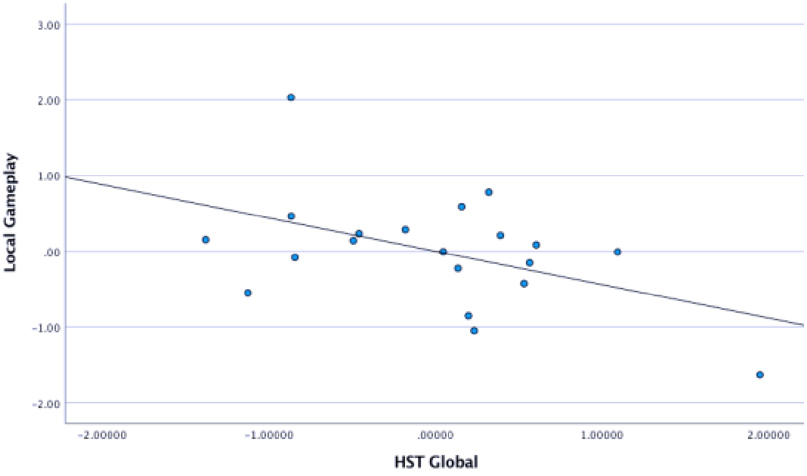
I predicted that there would be a negative correlation between HST global scores and the local gameplay composite variable. The correlation was relatively modest ($r = -.272$) and not statistically significant, but was in the right direction. Similarly, I predicted that

Figure 3: *Partial Scatterplot for HST and Global Gameplay with Line of Best Fit*



Note. *Both variables are shown as z-scores.*

Figure 4: *Partial Scatterplot for HST and Local Gameplay with Line of Best Fit*



Note. *Both variables are shown as z-scores.*

Table 4: *Regression Analysis Predicting Global Gameplay*

Predictor	Partial *r*	SE	*t*-value	*p*-value
HST	.340	.264	1.142	.280
Extraversion	–.011	.496	–.035	.973
Openness to Experience	–.254	.396	–.829	.427
Child VG	.364	.351	1.238	.244
Present VG	.297	.428	.982	.349

Note. *For all variables, N = 20. SE = standard error. * p<.05. ** p<.01. *** p<.001.*

Table 5: *Regression Analysis Predicting Local Gameplay*

Predictor	Partial *r*	SE	*t*-value	*p*-value
HST	–.473	.259	–1.699	.120
Extraversion	.181	.487	.581	.574
Openness to Experience	.116	.389	.369	.720
Child VG	–.403	.344	–1.392	.194
Present VG	–.119	.420	–.380	.712

Note. *For all variables, N = 20. SE = standard error. * p<.05. ** p<.01. *** p<.001.*

HST global scores could explain unique variability in local gameplay after accounting for the same personality measures and video game experience. As predicted, HST had a fairly strong negative partial correlation with local gameplay, but this was not statistically significant. Once again, the partial correlation of –.473 was more noteworthy than the zero-order correlation of –.272 because there was a high amount of variability in local gameplay explained by the other predictors such as past video game experience.

The mild relationship between HST global bias in predicting global gameplay does not directly support any existing findings. Even so, this relationship shows how acting "globally" (e.g., having positive, prosocial interactions with NPCs) may be related to the HST (over and above other predictors), which resembles existing research by Mendonca (2021) as well as Colzato and colleagues (2010) who show the link between attentional breadth and prosociality.

Personality and Gameplay

Contrary to expectations, there was no significant relationship between Extraversion and NPC interactions, ($r = -.051$, $p > .05$). One reason why these two variables are not correlated could be because voluntary interactions may reflect some participants pursuing a trophy rather than reflecting intrinsic inclination toward social interaction. For example, although the Trophy House provides hints to where trophy quests may be initiated, players might have talked to several NPCs before coming across the individual they are in search of for the quest.

Contrary to expectations, openness to experience did not positively correlate with areas visited ($r = -.033$, $p > .05$). The absence of a correlation could be because the effect of openness to experience was weaker than that of past video game experience, which caused participants to struggle more on the sports attempts, resulting in less time for exploring areas. Video game experience, past and present, were found to be related to gameplay measures. This relationship is sensible because video game experience allows one to be more successful at the sports, which decreases the need to repeat attempts in a manner that exhibits local gameplay. Simultaneously, this in-game success also provides more opportunity to make progress in quests and side areas in a manner that exhibits global gameplay. However, although a player may have the ability (i.e., due to video game experience) to make greater progress in the game, they may intrinsically enjoy playing a specific game which draws them to repeat sports even after the scroll is earned. Therefore, deeper analyses could be conducted using the gameplay measures of post-game sport attempts as well as bonus version sports attempts.

There were also some suggestive correlations between video game experience and personality measures. For example, extraversion and openness to experience were found to have weak-to-medium correlations with present video game experience (see Table 3). However, the present study did not produce any key results directly connecting personality to gameplay in Champion Island, contrary to past findings showing that personality could predict the nature of in-game play (e.g., Zeigler-Hill & Monica, 2015). While results suggest that participants' gaming background and the level of difficulty of Champion Island may play a role in whether or not a player behaves based on their gaming abilities or based on their personality and naturally occurring attentional breadth.

Limitations and Strengths

One strength of this study is how the remote participation aspect of the design is likely to have more ecological validity than if participants were to have completed all tasks and the gameplay within the lab. This is because the external validity would be higher when applying in natural settings given that people would typically be playing Champion Island in their own homes on their own computers, rather than in a lab. Another strength is that participants' gameplay behaviours were video-recorded and coded, which generally has greater accuracy than a self-report measure, such as if participants indicated what aspects of Champion Island they completed by filling in a questionnaire.

The present study contained numerous limitations on the precision of measurement techniques, threats to internal validity. To start, the past and present video game experience questionnaire was short and limited, yet this seemed to be a relevant factor that could be explored at a deeper level. For example, the number of hours per week and the number of years could have been collected to more deeply measure video game experience. This may be especially important given the perceptual effects of habitual video game engagement (e.g., Chopin et al., 2019). Another area with questionable measurement was the coded number of NPC interactions. This behaviour was difficult to quantify in all situations because the purpose was to measure intentional interactions, but this was not easily distinguished between moments when the player may have accidentally clicked a button that prompted the NPC dialogue.

One potential confound in the gameplay setup was that those with less video game experience may have struggled with the controls in a manner that impeded their expression of global/local gameplay. Similarly, a potential environmental bias influencing gameplay was that the game emphasized on the sports, and participants may have simply felt compelled to complete the sports prior to the side quests (e.g., the game description says, "earn all 7 scrolls to restore balance to the island," indicating that scrolls rather than trophies are the superior goal). Moreover, the 75 minute gameplay duration may have been too short for those who struggled with the controls. Accordingly, future research could consider providing a different amount of time, providing more support for game controls, or both. Alternatively, 75 minutes may be too long because there were ceiling effects for the Main Areas and Sports Attempted gameplay variables. An effective way to account for gameplay duration and video game experience would be to use a homogeneous set of participants with respect to video game experience (past or present).

For example, future studies can use a bunch of experienced gamers. and give them less time, or use a bunch of novice gamers and give them more time.

Another limitation was that participants played the game prior to completing the HST, so instead of measuring participants' naturally-occurring global bias, the HST may have measured participant's global bias in response to the video game challenges, which may have required exerting a lot of self-control, especially for participants less familiar video games, in turn manipulating attentional breadth (Pitchford & Arnell, 2018). The present study had several limitations on the generalizability of results. First and foremost, the sample size (*N* = 20) was inadequate in terms of statistical power and external validity of sample, due to the fact that it was insufficient in representing a normal distribution. The skew of the data prevents the error variance from cancelling each other out, resulting in less measurement accuracy and reliability. Additionally, some of the quests were not attempted by any participant, so not all aspects of this game were captured. Another limitation of this study are that only university students were looked at in this sample. Therefore, the results do not tell us how this gameplay relates to psychological traits of children and youth, an important demographic in video game research.

Implications and Future Directions

This study did not examine participants' behaviour during the sports mini-games. Future research could consider this behaviour, such as the use of strategies (e.g., power-ups in the rugby mini-game). Similarly, only one game was used for this study, and future research could examine video games from different genres. The present study employed a nonexperimental, single-group research design. In the future, an experimental research design with multiple conditions could be used. For instance, researchers can manipulate participants' global/local bias by priming them for global bias prior to them playing the game. This sort of experiment would test causal relationships between global attentional bias and in-game variables. Thus, a similar research question about this relationship could be answered by providing the very first experimental test on the gameplay behaviours in this specific game. Given the use of electroencephalography (EEG) on research in attentional breadth (e.g., Pitchford & Arnell, 2018), future research on video games and global/local bias could measure brain activity through EEG or similar methods. Similarly, local and global processing resemble "analytic" and "holistic" thinking respectively (Nisbitt et al., 2001). There may be games by which these patterns of thinking can be better examined.

On a broader level, other areas of cognitive research could be integrated with attentional breadth, attitudes, and personality in future video game research. Visual cognition is a related cognitive field with existing video game research involving eye-tracking (e.g., Chen & Tsai, 2015) and visual information processing (e.g., McColeman et al. 2020). Future research could investigate how these variables relate to each other as well as video game behaviour to better understand the cognitive mechanisms that relate to our virtual behaviour.

Through its artistic and technical richness, Champion Island shows how non-commercial video games can be used to celebrate culture and the global community (Walker, 2021). Games and gamification are becoming more effective engagement mechanisms beyond the typical "gamer" (e.g., Hassan, 2017). Thus, the research on this game can be continued to develop more results relevant to the real world, such as the potential of Champion Island to promote prosociality and global-minded thinking. For example, a future area of inquiry could be the extent to which this game can successfully promote prosocial attitudes and behaviours, such as cultural tolerance through appreciation of Japanese culture and folklore.

Conclusion

In summary, the present study aimed to examine whether attentional breadth can predict virtual behaviours over and above other psychological factors. This research question stemmed from existing research indicating how attentional breadth can predict real-world public health behaviours (Mendonca, 2021). Although many results were not statistically significant, the overall methodology provided interesting results and implications on how attentional breadth may be related to video game behaviours and experience. Finally, this study sets the stage for more cognitive research on open-access video games such as Champion Island.

References

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality, 15, 327–353.* https://doi.org/10.1002/per.417

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11, 150–166.* https://doi.org/10.1177/1088868306294907

Bavelier, D., & Green, C. S. (2003). Action video game modifies visual selective attention. *Nature, 423(6939), 534–537.* https://doi.org/10.1038/nature01647

Bavelier, D., & Green, C. S. (2019). Enhancing attentional control: Lessons from action video games. *Neuron, 104(1), 147–163.* https://doi.org/10.1016/j.neuron.2019.09.031

Bellgrove, M. A., Vance, A., & Bradshaw, J. L. (2003). Local-global processing in early-onset schizophrenia: Evidence for an impairment in shifting the spatial scale of attention. *Brain and Cognition, 51(1), 48–65.* https://doi.org/10.1016/S0278-2626(02)00509-2

Benoit, J. J., Roudaia, E., Johnson, T., Love, T., & Faubert, J. (2020). *The neuropsychological profile of professional action video game players. PeerJ, 8, e10211–e10211.* https://doi.org/10.7717/peerj.10211

Chen, Y., & Tsai, M. J. (2015). Eye-hand coordination strategies during active video game playing: An eye-tracking study. Computers in Human Behavior, 51, 8-14. https://doi org/10.1016/j.chb.2015.04.045

Choi, E. J., Jang, K. M., & Kim, M. S. (2014). Electrophysiological correlates of local-global visual processing in college students with schizotypal traits: An event-related potential study. *Biological Psychology, 96(1), 158–165.* https://doi.org/10.1016/j.biopsycho.2013.12.015

Chopin, A., Bediou, B., & Bavelier, D. (2019). Altering perception: The case of action video gaming. *Current Opinion in Psychology, 29, 168–173.* https://doi.org/10.1016/j.copsyc.2019.03.004

Colzato, L. S., Hommel, B., Van den Wildenberg, W. P. M. V., & Hsieh, S. (2010). Buddha as an eye opener: A link between prosocial attitude and attentional control. *Frontiers in Psychology, 1(SEP), 1–5.* https://doi.org/10.3389/fpsyg.2010.0015

Dale, G., & Arnell, K. M. (2013). Investigating the stability of and relationships among global/local processing measures. *Attention, Perception, and Psychophysics, 75(3), 394– 406.* https://doi.org/10.3758/s13414-012-0416-7

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature, 382*(6592), 626–628. https://doi.org/10.1038/382626a0

Förster, J., Liberman, N., & Shapira, O. (2009). Preparing for novel versus familiar events: Shifts in global and local processing. *Journal of Experimental Psychology: General, 138(3),* 383–399. https://doi.org/10.1037/a0015748

Freeseman, L. J., Colombo, J. & Coldren, J. T. (1993). Individual differences in infant visual attention: Four-month-olds' discrimination and generalization of global and local stimulus properties. *Child Development, 62(4),* 1191–1203. https://doi.org/10.1111/j.1467-8624.1993.tb04195.x

Gable, P. A. & Harmon-Jones, E. (2008). Approach-motivated positive affect reduces breadth of attention. *Psychological Science, 19(5),* 476–482. https://doi.org/10.1111/j.1467- 9280.2008.02112.x

Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *The American Psychologist, 69(1),* 66–78. https://doi.org/10.1037/a0034857

Halldorsson, B., Hill, C., Waite, P., Partridge, K., Freeman, D., & Creswell, C. (2021). Annual research review: Immersive virtual reality and digital applied gaming interventions for the treatment of mental health problems in children and young people: The need for rigorous treatment development and clinical evaluation. *Journal of Child Psychology and Psychiatry, 62(5),* 584–605. https://doi.org/10.1111/jcpp.13400

Hallgren K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8(1),* 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hassan, L. (2017). Governments should play games: Towards a framework for the gamification of civic engagement platforms. *Simulation & Gaming, 48(2), 249–267.* https://doi.org/10.1177/1046878116683581

Hutchison, D. (2007). *Playing to learn: Video games in the classroom.* Teacher Ideas Press. Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance, 8(4),* 521– 535. https://doi.org/10.1037/0096-1523.8.4.521

Kyndt, E., Cascallar, E., & Dochy, F. (2012). Individual differences in working memory capacity and attention, and their relationship with students' approaches to learning. *Higher Education, 64(3),* 285–297. https://doi.org/10.1007/s10734-011-9493-0

McColeman, C., Thompson, J., Anvari, N., Azmand, S. J., Barnes, J., Barrett, R. C. A., Byliris, R., Chen, Y., Dolguikh, K., Fischler, K., Harrison, S., Hayre, R. S., Poe, R., Swanson, L., Tracey, T., Volkanov, A., Woodruff, C., Zhang, R., & Blair, M. (2020). Digit eyes: Learning-related changes in information access in a computer game parallel those of oculomotor attention in laboratory studies. *Attention, Perception & Psychophysics, 82(5),* 2434–2447. https://doi.org/10.3758/s13414-020-02019-w

Mendonca, A. (2021). *Missing the forest for the trees: An exploration of the link between attentional breadth and COVID-19 guideline compliance* [Unpublished undergraduate thesis]. Brock University.

Moosbrugger, H., Goldhammer, F., & Schweizer, K. (2006). Latent factors underlying individual differences in attention measures: Perceptual and executive attention. *European Journal of Psychological Assessment, 22(3),* 177–188. https://doi-org/10.1027/1015- 5759.22.3.177

Mukherjee, S., Srinivasan, N., Kumar, N., & Manjaly, J. A. (2018). Perceptual broadening leads to more prosociality. *Frontiers in Psychology, 9(SEP), 1–10.* https://doi.org/10.3389/fpsyg.2018.01821

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9(3),* 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research, 43(1), 1*–32. https://doi.org/10.1007/BF00309635

NICHD Early Child Care Research Network. (2005). Predicting individual differences in attention, memory, and planning in first graders from experiences at home, child care, and school. *Developmental Psychology, 41(1),* 99–114. https://doi.org/10.1037/0012- 1649.41.1.99

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108(2),* 291–310. https://doi.org/10.1037/0033-295X.108.2.291

Nunez, M. D., Srinivasan, R., & Vandekerckhove, J. (2015). Individual differences in attention influence perceptual decision making. *Frontiers in Psychology, 8,* 18–18. https://doi.org/10.3389/fpsyg.2015.00018

Pitchford, B. & Arnell, K. (2018). Self-control and its influence on global/local processing: An investigation of the role of frontal alpha asymmetry and dispositional approach tendencies. *Attention, Perception, & Psychophysics, 81,* 173–187. https://doi.org/10.3758/s13414-018-1610-z

Poirel, N., Pineau, A., & Mellet, E. (2008). What does the nature of the stimuli tell us about the global precedence effect? *Acta Psychologica, 127(1), 1–11.* https://doi.org/10.1016/j.actpsy.2006.12.001

Scherf, K. S., Luna, B., Kimchi, R., Minshew, N., & Behrmann, M. (2008). Missing the big picture: Impaired development of global shape processing in autism. *Autism Research, 1(2),* 114–129. https://doi.org/10.1002/aur.17

Segev, A., Gabay-Weschler, H., Naar, Y., Maoz, H., & Bloch, Y. (2017). Real and virtual worlds alike: Adolescents' psychopathology is reflected in their videogame virtual behaviors. *PLoS ONE 12(7): e0181209.* https://doi.org/10.1371/journal.pone.0181209

Stangor, C. (2012). *Introduction to psychology.* Saylor Academy Open Textbooks. https://saylordotorg.github.io/text_introduction-to-psychology/index.html

Teng, C.-I., Jeng, S.-P., Chang, H. K.-C., & Wu, S. (2012). Who plays games online?: The relationship between gamer personality and online game use. *International Journal of e Business Research, 8(4),* 1–14. https://doi.org/10.4018/jebr.2012100101

Triandis, H. C. (1995). *Individualism and collectivism. Westview Press.*

Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality, 69(6),* 907–924. https://doi.org/10.1111/1467-6494.696169

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74(1),* 118–128. https://doi.org/10.1037/0022-3514.74.1.118

Walker, J. (2021, July 23). Google celebrates Olympics with an entire JRPG. *Kotaku.* https://kotaku.com/googles-olympic-doodle-is-an-entire-jrpg-1847348733

Walsh, M. (2021, July 23). Largest Google doodle ever is a playable sports video game. *Nerdist.* https://nerdist.com/article/largest-google-doodle-ever-is-a-playable-sports-video-game/

Yovel, I., Revelle, W., & Mineka, S. (2005). Who sees trees before forest?: The obsessive compulsive style of visual attention. *Psychological Science, 16(2),* 123–129. https://doi.org/10.1111/j.0956-7976.2005.00792.x

Zeigler-Hill, V., & Monica, S. (2015). The HEXACO model of personality and video game preferences. *Entertainment Computing, 11, 21–26.* https://doi.org/10.1016/j.entcom.2015.08.001

Preferential Unconscious Attention to Pictures Over Words in a Visual Recognition Paradigm

Sally Berson

University of Michigan

Department of Psychology

Olivia Ripley

Skidmore College

Department of Neuroscience

Alexandra Zehner

Barnard College,
Columbia University

Department of Neuroscience & Behavior

Elisa Papagallo

Washington University in
St. Louis

Department of Philosophy-Neuroscience-Psychology

Abstract

While prior literature has demonstrated the existence of unconscious attention, differences in processing disparate forms of visual stimuli have yet to be investigated. The present study aimed to determine whether unconscious attentional processing is preferential for pictorial or textual stimuli. Using an adaptation of Posner's (1980) spatial cueing paradigm, we tested subjects' (N = 25) ability to identify the location of unconsciously presented words and pictures. Participants were significantly faster and more accurate when locating consciously and unconsciously presented pictures, indicating preferential attention to pictorial over textual stimuli. This preference may be due to the increased evolutionary importance of pictures, or to a proposed greater allocation of cognitive processing networks (Paivio, 1986). In any case, the results of this study encourage further research of the influence of stimuli characteristics on unconscious attention, as it may prove relevant to academic and commercial fields alike.

Keywords: **spatial cueing paradigm, stimuli characteristics, preferential unconscious attention.**

Introduction

The relationship between consciousness and attention has been a subject of much debate (van Boxtel et al., 2010; Tallon-Baudry, 2012). Some theorists argue that they are entirely separate (Lamme, 2006; Cararra-Augustenborg, 2013), while others go so far as to equate the two (Baars, 2005; Damasio, 2003). However, over the last few years, there has been an abundance of research indicating that consciousness and attention are separate entities, supporting the existence of both conscious and unconscious attention (Heemskerk et al., 1996; Naccache et al., 2002; Kentridge et al., 2004; Montaser-Kouhsari & Rajimehr, 2004; Jiang et al., 2006; Sato et al., 2007; Chen et al., 2015; Prasad & Kumar Mishra, 2019). Accordingly, current research is centred around the characteristics of unconscious attention and its capabilities.

Researchers of unconscious attention tend to differentiate between bottom-up and top-down processes, particularly because the former is especially rapid to develop and dissolve in an unconscious attention context (Mulckhuyse & Theeuwes, 2010). Some have proposed that expectations and goals have an important impact on attentional orienting (Mulckhuyse & Theeuwes, 2010), while others advocate exogenous stimuli can successfully capture attention without explicit awareness (McCormick, 1997). Furthermore, some studies have even suggested that top-down and bottom-up processing may be inextricable due to mutual moderational influences (Chen et al., 2015). Despite these incongruencies, unconscious attention studies generally tend to use similar cueing methodologies, including the masked priming and spatial cueing paradigms (Prasad & Kumar Mishra, 2019; Posner, 1980). More than just validating the presence of unconscious attention, they can also be utilized to answer another lingering set of questions in this area of research: What aspects of a stimulus can be processed unconsciously, and are stimuli with certain characteristics processed preferentially?

Researchers have attempted to answer this question using the comparison of conscious attention to words versus pictures, with some researchers advocating for a picture superiority effect (Paivio & Csapo, 1973). For example, a study by Miller (2011) found that participants completing an attention task had faster reaction times for pictorial stimuli than textual stimuli. He proposes this may be because words are categorized as a unique form of visual stimuli due to their semantic content, leading to less efficient cognitive processing. However, Miller also proposes that task demands seem to have a significant effect on processing. Contrastingly, a study by Amrhein et al. (2002) did not find evidence for the picture superiority effect, and subsequently shed skepticism on pictorial stimuli's privileged processing. To our knowledge, no study has determined whether this attentional difference explicitly applies to unconscious stimuli. Therefore, this study seeks to fill that gap in the literature by examining the difference in unconscious attention for stimuli with different characteristics presented in the same task.

In the present study, we aim to determine whether unconscious attentional processing is preferential for pictorial or textual stimuli. This research question is assessed using an adaptation of the spatial cueing task from Posner (1980) that measures participants' reaction times (RT) and accuracy when identifying the location of unconsciously presented pictures and words. Our hypotheses are: (i) RT will be lower and accuracy will be higher for questions regarding consciously presented stimuli compared to questions regarding unconsciously presented stimuli; (ii) RT for questions regarding unconsciously presented pictures will be faster than that of unconsciously presented words; (iii) accuracy will be higher for questions regarding unconsciously presented pictures than that of unconsciously presented words. The latter two hypotheses were predicated on research of the picture superiority effect for conscious perception (Paivio & Csapo, 1973).

Methods

Participants

Participants were recruited from the Danish Institute for Study Abroad in Copenhagen. There were 25 in total, all between the ages of 20 and 22 (M = 20.88, STD = 0.53; 16 female, 8 male, 1 non-binary). Most participants had either normal or corrected-to-normal vision. Those without corrected-to-normal vision (N = 3) had myopia and could see the materials clearly without corrective lenses.

Paradigm

We used an adaptation of the spatial cueing design first described by Posner (1980). The paradigm was created using E-prime software (Psychological Software Tools, Pittsburgh, Pennsylvania, USA). Each participant sat 90 cm away from a computer monitor, with the height of the monitor adjusted to be centred in their visual field. Participants were first presented with a screen containing a black and white, 5” x 5” (12.7 cm x 12.7 cm), four-quadrant grid containing a 1” x 1” (2.54 cm x 2.54 cm) red cross in the centre for 1000 ms (Figure 1). Following this, a cue (1.5” x 1.5” (3.81 cm x 3.81 cm) red star) was randomly presented in one of the four quadrants for 1000 ms. The subject was instructed to direct their attention toward the cue. Two words and two images positioned randomly within the four quadrants then flashed for 45 ms, a duration just above the conscious perception threshold (Rolls, 2004). This timing enabled participants to consciously attend to the cued stimulus, while unconsciously attending to the three uncued stimuli. The words used were all four-letter nouns to reduce differences in complexity, and the images used were of singular objects, obtained from an independent artist. Subjects were thereafter asked to identify the location of one of the four stimuli using the W, C, M, and O keyboard keys, each corresponding with one of the four quadrants, as quickly and as accurately as possible. These four keys were chosen because, with stimuli presented for such a short duration, they needed to be physically separate (meaning participants used a different finger to press each key), spatially correspond to the quadrants, and be roughly equidistant. After answering this question, the blank quadrant with the red cross immediately returned, and the next trial began.

Participants completed two practice trials. In these practice rounds, subjects were only asked about stimuli in the cued quadrant, and they were given immediate feedback on the accuracy of their responses. All subjects thereafter completed a total of 24 real trials in randomized order. Eight of the trials required the participant to locate the stimulus that was in the cued quadrant. These trials (referred to as "cued" trials), therefore, had valid cues, and the participants' responses would reflect conscious attentional processing. For the other 16 trials (referred to as "uncued"), they were asked to identify the position of a stimulus in one of the three uncued quadrants. Thus, the cues were invalid, and the participants' responses reflected unconscious attentional processing. All subjects had been instructed before beginning that only questions asking about stimuli in the cued quadrant would count towards their final score. All four quadrants, stimuli types (pictures versus words), and directionality of cue versus question (i.e., top left to bottom left, top left to bottom right, etc.), were equally represented in the questions about uncued quadrants. Moreover, for each question, the stimulus was presented the exact way it was shown in the grid (i.e., when asking about a word, it was

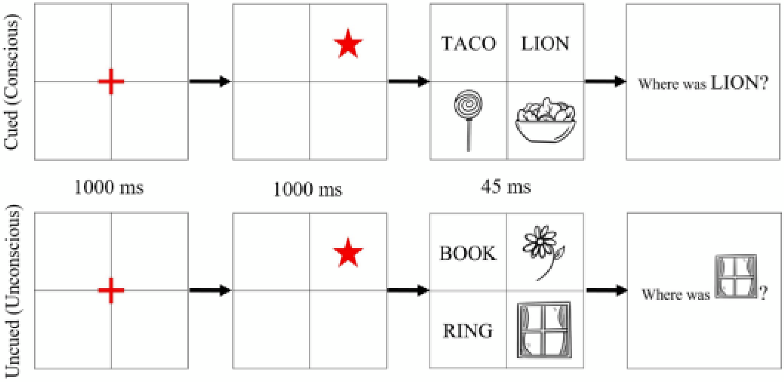


***Figure 1.** An example of our spatial cueing paradigm. "Cued" trials asked the participant to locate stimuli in the cued quadrant (valid cues, conscious processing). "Uncued" trials asked the participant to locate stimuli from any of the three uncued quadrants (invalid cues, unconscious processing).*

presented in the same font and size, and when asking about a picture, the picture itself was used in the question).

Data Analysis

Data analysis consisted of a combination of paired two-sample t-tests assuming equal variances, paired two-sample t-tests assuming unequal variances, simple linear regressions, and chi-square tests. Analysis and figures were generated on Excel Version 16.43. All computed means were simple arithmetic means.

Results

Data Validation

A metric of validity composed of two criteria was employed to assess the success of the paradigm. To meet the first criterion, participants must have performed with above-chance accuracy on the cued trials to confirm that conscious attention was utilized. 19 of the 25 participants satisfied this crucial criterion. The second layer of the metric required a significant difference in accuracy between trials with valid and invalid cues to likewise indicate dissociated forms of conscious and unconscious processing. Four of the remaining participants satisfied this criterion, leaving two subjects who neither performed above chance on the cued trials nor performed significantly differently in the two conditions. Because 92% of participants satisfied at least half of the validity metric, it was assumed that the paradigm functioned as planned and no participant was omitted.

Additionally, accuracy concerning exposure to the paradigm was modeled by linear regression with trial number as a predictor for accuracy. The model was statistically insignificant (R^2 = .02, F(1, 22) = .08, p < .05). This confirms that accuracy did not increase as a function of practice, and accuracy could be assessed through the measures of interest rather than as a product of learning.

Accuracy

Across all conditions, participants performed significantly better when asked to locate a picture (M = .6133, Var = .237) than a word [(M = .35, Var = .228), t(598) = 6.68, p < .0001].

As shown in Figure 2, this trend was maintained in the conscious cued condition [M(pictures) = .69, M(words) = .36, Var(pictures) = .216, Var(words) = .232, t(198) = 4.96, p < .001]. Likewise, in the unconscious uncued condition, responses to questions about pictures were significantly more accurate (M = .575, Var = .345) than responses to questions about words (M = .345, Var = .227, t(397) = 4.73, p < .001; Figure 2). The reported means refer to the percent of accurate responses for each participant across all 24 trials.

Reaction Time

Participants were overall faster when attempting to identify the location of a picture (M = 1.3, Var = 360.8) than that of a word (M = 1.6, Var = 400.8, t(598) = -6.38, p < .001). Similarly, reaction times were

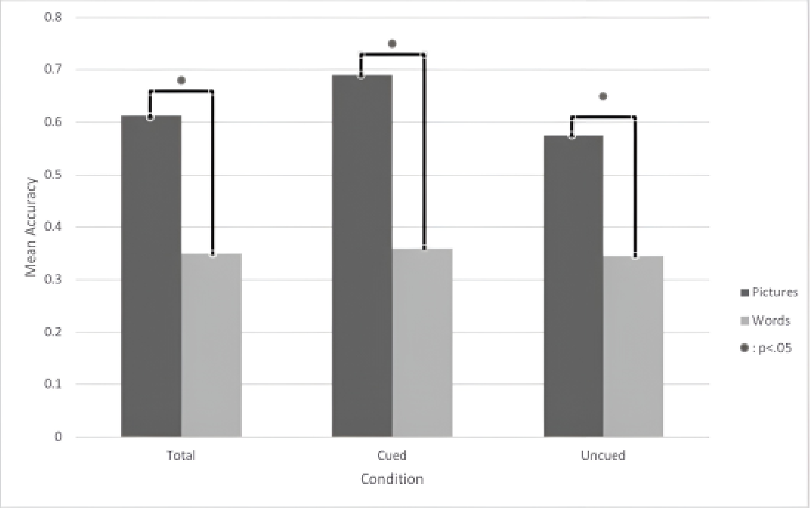


Figure 2. Accuracy in the general picture vs. word condition, as well as in the cued and uncued conditions. A significant difference was found between words and pictures in every condition.

lower for cued pictures (M = 1.2, Var = 264.6) than for cued words (M = 1.5, Var = 338.6, t(195) = -3.58, p = .0002). When identifying the location of figures that were unconsciously attended, participants were faster when asked about a picture (M = 1.3, Var = 408.5) than when asked about a word (M = 1.7, Var = 427.1, t(398) = -5.3, p < .001; Figure 3).

Quadrant Analysis

A post-hoc analysis of cue location revealed that when participants were prompted to look at the left side of the screen, they identified the location of the stimuli asked about with higher accuracy (M = .547, Var = .249) than when the cue was in one of the two right-side quadrants (M = .417, Var = .244, t(299) = 3.22, p = .0007). A chi-square analysis of each quadrant compared against the three others revealed that, for 3 out of the 4 tests comparing a left-sided cue to a right-sided cue, performance was significantly better when the cue was on the left side (Table 1). Conversely, there was no significant

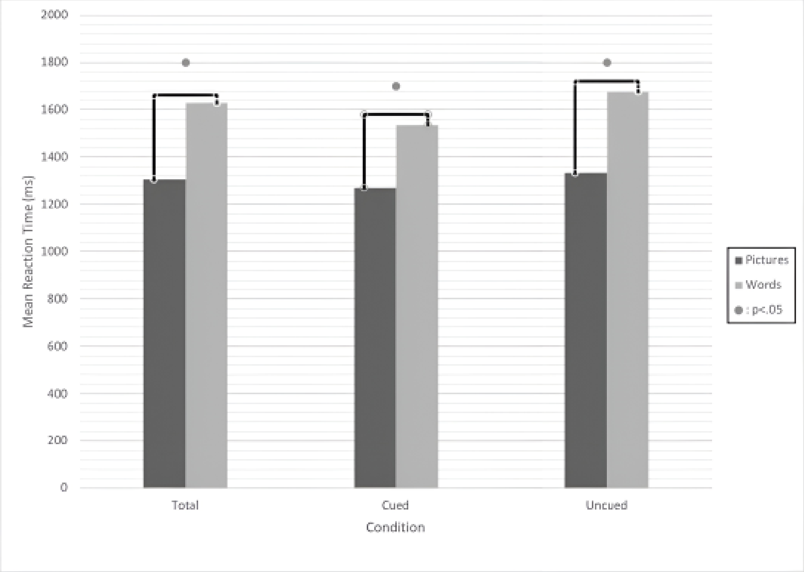


***Figure** 3. Mean reaction times in the general picture vs. word condition as well as in the cued and uncued conditions. A significant difference was found between words and pictures in every condition.*

mean difference in accuracy rate when the compared cues were unilateral and above/below each other.

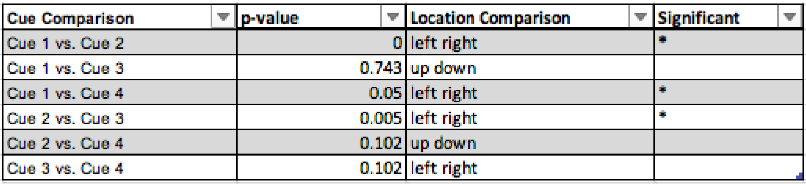
A closer look at the left-cue superiority effect depicted that improved performance when the cue was on the left-hand side applied when the figure asked about was a picture but not a word (Figure 4).

Therefore, only when the question asked about a picture, participants responded with greater accuracy when the cue was presented on the left side of the screen than when the cue was presented on the right side of the screen [(M = .69, Var = .215), t(298) = 3.336, p = .0004].

Discussion

These results contribute to a growing body of literature that corroborates the presence of unconscious attention and supports research suggesting consciousness and attention are distinct processes that likely rely on unique neural mechanisms (Chen et al., 2015). Moreover, as subjects were explicitly instructed to attend only to cued quadrants yet performed significantly better than chance when identifying the location of uncued picture stimuli, their performance supports previous findings demonstrating unconscious attention to exogenous cues (McCormick, 1997). Conversely, these results contradict previous studies suggesting task-relevant cues that match top-down goals are necessary to produce an efficient reaction time and high level of accuracy (Ansorge & Heumann, 2006).

Participants demonstrated preferential unconscious attention to pictures over words, as evidenced by their significantly greater accuracy and significantly lower response times when identifying uncued pictures compared to uncued words. There are several possible explanations for this finding. First, from an evolutionary perspective, filtering potential threats in the surrounding environment largely depends on processing images rather than words (e.g., an image resembling a snake presents more potential danger than the word “snake”). Therefore, unconsciously attending to pictures—and filtering whether they necessitate further conscious attentional resources—may have provided an evolutionary advantage and thus been selected for over time (Yorzinski, 2014). Paivio’s (1986) dual-coding theory may also explain the observed results. He argues that



Cue Comparison	p-value	Location Comparison	Significant
Cue 1 vs. Cue 2	0	left right	*
Cue 1 vs. Cue 3	0.743	up down	
Cue 1 vs. Cue 4	0.05	left right	*
Cue 2 vs. Cue 3	0.005	left right	*
Cue 2 vs. Cue 4	0.102	up down	
Cue 3 vs. Cue 4	0.102	left right	

Table 1: *Comparison of mean accuracy with respect to cue location. Quadrants and cue locations are labeled 1-4 clockwise. Significant differences ($p < 0.05$) occurred between cues that differed laterally*

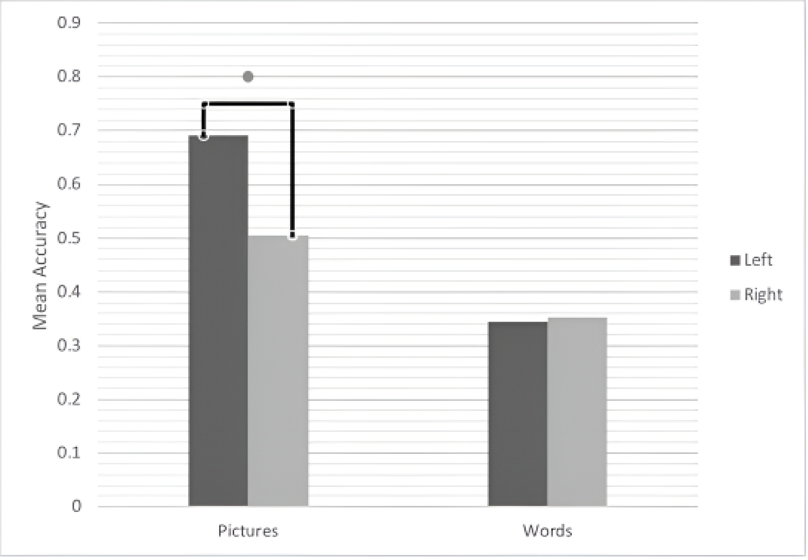


***Figure 4.** Mean accuracy for questions asking about pictures and words, depending on if the cue was on the left or right side of the screen. Performance was significantly better for questions regarding pictures when the cue was presented on the left. There was no significance in accuracy between left and right cues when the participant was asked to locate a word.*

information can be processed by verbal or visual networks. While textual stimuli can only be processed by verbal networks, pictorial stimuli can be processed by both, as pictures inherently have a verbal counterpart. Thus, there may be preferential unconscious attention for pictures because they are processed through two networks, allotting these stimuli more cognitive resources. It can therefore be hypothesized that either there is a lower threshold for unconsciously attending to pictorial stimuli, or that pictures may have a more robust and/or direct line of processing compared to words. It is also worth noting that within the cued stimuli, which were above the threshold for conscious perception, participants also displayed a higher accuracy and lower response time for pictures, suggesting preferential attentional processing for pictures over words even consciously.

One unanticipated result was that left-sided cues predicted a higher mean accuracy of response for questions regarding pictures. This finding could potentially be explained through the activation-orienting hypothesis for the contralateral direction of attention (Reuter-Lorenz et al., 1990), in conjunction with Kinsbourne's Orientation Model for right hemisphere (RH) attentional dominance (1993). According to the activation-orienting hypothesis, which is experimentally supported by Reuter-Lorenz et al. (1990), the brain directs attention contralaterally to the more activated hemisphere. This attentional directing mirrors the contralateral nature of visual processing. Kinsbourne (1993) corroborates the finding of Reuter-Lorenz et al. (1990) but elaborates that within the brain's vectors of attention to the contralateral visual field, the left hemisphere's (LH) vector is more strongly biased towards the right visual field (RVF), while the RH is more generalized to include both the LVF and RVF. As a result, Kinsbourne (1993) proposes that the RH is dominant for attention, as its vector directs attention to both visual fields, while the LH only directs attention to the RVF. In the context of our experimental paradigm and results, it is possible that left-sided cues activated the RH and therefore enabled attention to be directed to the entire visual field, resulting in a higher mean accuracy of responses. Conversely, right-sided cues activated the LH and directed attention almost exclusively towards the RVF, therefore lowering accuracy over the entire visual field. As pictures appear to be preferentially unconsciously attended to in comparison to words, this model likely had a disproportionately greater impact on those stimuli.

We anticipated that a potential limitation of our study would be ensuring that participants followed directions and oriented their attention to the cued quadrant (and, by extension, that the rest of the stimuli were attended to unconsciously). In response, we took multiple measures to mitigate this risk. First, we instructed participants before beginning that only questions regarding quadrants containing the red star would count towards their "final score". Second, the speed with which the successive trials were conducted likely prevented participants from forming and implementing a thorough strategy. Lastly, as all the stimuli were presented barely above the threshold for conscious perception, consciously attending to the cued stimulus and all three uncued stimuli simultaneously seems unlikely, if not impossible. These efforts appeared to have successfully mitigated this potential limitation, as accuracy did not increase as a function of trial number, confirming the absence of a learning curve or practice effect. Most participants' significantly higher rate of accuracy when asked about cued, in comparison to uncued stimuli, also endorses that most followed instructions. That being said, it is possible that the two subjects who did not satisfy the criteria for validity (as they did not have a significant difference between response accuracy to cued and uncued stimuli) may have been directing their attention toward uncued quadrants.

There are, however, additional limitations to our study that may have affected our results. To start, there is a possibility that once subjects realized questions were being asked about uncued quadrants, instead of remaining a completely bottom-up task, there was a shift to making all stimuli task-relevant and thus subject to top-down processing (even if still unconscious). We can therefore not be completely confident whether our findings extend only to bottom-up attention, or a combination that also includes top-down. Additionally, despite being an adaptation of the Posner cueing task, this paradigm is new and could benefit from further testing to ensure its validity. Moreover, although our images were obtained from an independent artist, and thus had a similar style, there were apparent differences in the complexity of each image. In addition to varying levels of complexity between pictures, many of the pictures also contained more features than the words, which may have influenced the participants' attentional orienting. As a future direction, we propose a study design that standardizes the complexity—as measured by the number of features—of the images and words used.

Future studies could also investigate the influence of emotional valence on unconscious attention. Earlier, we proposed the difference in attentional processing may be evolutionarily based; a study examining whether there are differences in unconscious attention toward attractive or aversive pictorial and textual stimuli could help evaluate the accuracy of that claim. Finally, regarding practical applications, the results of this study could be utilized by neuromarketing firms. Conscious and unconscious attention were preferentially directed towards pictures, indicating that graphics may be more crucial than text when advertising a product. Neuromarketing research could further examine how to best implement this principle to create more efficient marketing for businesses.

Conclusion

Although the presence of unconscious attention is becoming increasingly acknowledged—and is supported by our findings—the specificities of this form of processing have yet to be thoroughly studied. Our results build upon literature in this field by uncovering a novel component of unconscious attention: its preferential treatment of pictures over words. We have provided several explanations for this finding, as well as proposed an interpretation for why left-sided cues coincided with higher accuracy of responses to pictorial stimuli. Future research should investigate the impact of feature number, emotional valence, and other aspects to further elucidate the effect of stimuli characteristics on unconscious attention. Understanding how humans process stimuli at this most basic level will prove useful for both academic and practical purposes as results can be applied to a wide variety of subjects, from theories of consciousness to neuromarketing strategies.

References

Amrhein, P. C., McDaniel, M., & Waddill, P. (2002). Revisiting the picture-superiority effect in symbolic comparisons: Do pictures provide privileged access? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 843–857.

Ansorge, U., & Heumann, M. (2006). Shifts of visuospatial attention to invisible (metacontrast-masked) singletons: Clues from reaction times and event-related potential. *Advances in Cognitive Psychology, 2(1),* 61–76.

Baars, B. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Progresses in Brain Research, 150,* 45–53.

Carrara-Augustenborg, C. (2013). Endogenous feedback network: Summary and evaluation, in *The development of a comprehensive model of human consciousness,* 15–23, Ph.D Thesis, University of Copenhagen.

Chen, X. U., Ran, G., Zhang, Q. I., & Hu, T. (2015). Unconscious attention modulates the silencing effect of top-down predictions. *Consciousness and Cognition, 34,* 63–72.

Damasio, A. (2003). Feeling of emotions and the self. *Ann. N.Y Acc.. Science, 1001,* 253–261

Heemskerk, J., DiNardo, S., & Kostriken R. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383,* 334–337.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender-and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, 103*(45), 17048–17052.

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia, 42*(6), 831–835.

Kinsbourne M. (1993). "Orientational bias model of unilateral neglect: Evidence from attentional gradients within hemispace." *Unilateral Neglect: Clinical and Experimental Studies, edited by I.H Robertson & J.C. Marshall. Erlbaum.*

Lamme, V. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences 10*(11), 494–501.

Miller, P. (2011). The processing of pictures and written words: A perceptual and conceptual perspective. *Psychology, 2*(7), 713–720.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 168–180.

Montaser-Kouhsari, L., & Rajimehr, R. (2004) Attentional modulation of adaptation to illusory lines. *Journal of Vision, 4*(6), 434–444.

Mulckhuyse, M., & Theeuwes, J. (2010) Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta Psychologica 134*(3), 299–309.

Naccache, L., Blandin, E., & Dehaene, S. (2002) Unconscious masked priming depends on temporal attention. *Psychological Science, 13*(5), 416–42.

Paivio, A. (1986). Mental representations: A dual-coding approach. New York: Oxford University Press.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery and dual encoding? *Cognitive Psychology 5,* 176–206.

Posner, M. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32,* 3–25.

Prasad, S. & Kumar Mishra, R. (2019). The nature of unconscious attention to subliminal cues. *Vision, 3*(38).

Reuter-Lorenz, P .A., Kinsbourne, M., & Moscovitch, M. (1990). Hemispheric control of spacial attention. *Brain and Cognition, 12*(2), 240–266.

Rolls, E. (2004). Consciousness absent and present: A neurophysiological exploration. *Progress in Brain Research, 44,* 95–106.

Sato, W., Okada, T., & Toichi, M. (2007). Attentional shift by gaze is triggered without awareness. *Experimental Brain Research, 183*(1), 87–94

Tallon-Baudry, C. (2012). On the neural mechanisms subserving consciousness and attention. *Frontiers in Psychology 2.*

van Boxtel, J. J. A., Tsuchiya, N., & Koch, C. (2010). Consciousness and attention: On sufficiency and necessity. *Frontiers in Psychology 1.*

Yorzinski, J. L., Penkunas, M. J., Platt, M. L., & Coss, R. G. (2014). Dangerous animals capture and maintain attention in humans. *Evolutionary Psychology, 12*(3), 534–548.

5head
Ordinary language philosophy
Calls problems overthinking atrocity
  I don't mean to be crass
  But you’re stuck up your ass
And this poem roasts academic pomposity

Head (Part 3)

Angel Martin

Effect of Monetary Reward on Manual Automatic Imitation

Luqman Azmir

University College London

Division of Psychology and Language Sciences

Abstract

Automatic imitation is an ongoing topic of research across various domains including neuroscience, social psychology, and cognitive psychology. It is unclear how automatic the process really is, but previous studies suggested that motivation as measured with reward manipulation could discern between intentional and automatic imitation. The present study investigated the influence of reward manipulations on the automatic imitation effect. Specifically, we conducted an online study with a between-subjects design (N = 83) comparing the performance between participants with monetary incentives (reward group) and participants with no monetary incentives (no -reward group) through a stimulus-response compatibility (SRC) key-pressing task. We found an automatic imitation effect, demonstrated by faster response times for congruent trials than incongruent trials, but no effect of the reward manipulation is observed in this study. We concluded that automatic imitation was not modulated by reward suggesting that the imitation process was automatic.

***Keywords:* imitation, rewards.**

Introduction

Automatic imitation

Imitation is a common behaviour in humans and animals alike and has garnered scientific interest to understand its nature. The term "imitation" refers to any circumstance in which an observer performs the same action as what they observe, including both intentional and involuntary copying (Chartrand & Bargh, 2002; Catmur, 2016). Behavioural studies postulate a connection between imitation and motor neural networks in the brain, or else ubiquitously known as motor mimicry or imitation. Automatic imitation has been investigated in various domains such as perceptuomotor processing of speech (Jarick & Jones, 2009; Adank et al., 2018), emotional processing (Mondillon et al., 2007; Butler et al., 2016), motivational stimulation (Soutschek et al., 2014), and social context (Pan & Hamilton, 2015). In these studies, an individual observing an action performed would have the tendency to involuntarily imitate the observed action, given that the visual depiction of the action and of self-motor program matches (Catmur, 2016). The central debate surrounding imitation focuses on its propensity to be "automatic", and has received considerable interest among researchers to distinguish between a controlled imitation and an automatic imitation by their underlying cognitive processes.

Automatic imitation refers to the phenomenon that motor planning of an observed action is activated independent of intentional, top-down cognitive control processes resulting in covert imitation (Brass et al., 2001; Heyes, 2011). Neuroimaging studies have pointed to areas of the brain considered as part of the mirror neuron system. In this system, a set of brain areas are activated both during observation and execution of specific actions; for example, during speaking and listening (Rizzolatti et al., 2001; Cross & Iacoboni, 2014). In behavioural experiments, automatic imitation can be elicited through the stimulus-response compatibility (SRC) experimental paradigm which is described in detail below (Brass et al., 2000).

Previous research has substantially expanded the understanding of automatic imitation of observed action in a variety of contexts, such as when primed with pro-social vs anti-social sentences (Wang & Hamilton, 2012). In Wang & Hamilton (2012), they investigated whether priming pro-social and anti-social sentences could influence imitation in a finger-tapping task; a larger automatic imitation effect was found for anti-social sentence priming than for the pro-social sentence priming. Another study compared possible vs impossible finger movements in an SRC task performed by a biomechanical hand, as well the effect of attentional weighting, i.e., when subjects were aware of both possibilities, on automatic imitation effect (Longo et al., 2008). Interestingly, automatic imitation was elicited for both possible and impossible finger movements, but the effect vanished when subjects were aware of the impossible movements. Longo and colleagues (2008), through this finding, proposed that finger movements were coded in terms of the goal, not the manner in which the action is performed. The automaticity of imitating observed action has been supported in these studies in which they demonstrated how social contexts could influence the effect of imitating congruent and incongruent hand or finger movements. Pan and Hamilton (2015) investigated automatic imitation in a rich social environment with virtual characters (VC) and balls performing compatible actions. They found an automatic imitation effect (measured by faster response times to stimuli shown) for compatible trials, and that the effect was larger for action performed by VC compared to balls. A major consideration of past evidence is about discerning an imitation which is initiated by cognitive control to an imitation which is automatic, independent of intentional and deliberate cognitive control.

Research in automatic imitation is imperative to understanding the concentration required to suppress the urge to imitate upon observing someone else act. Scerrati et al. (2017) reported the role of spatial attention in the coding of spatial position of a stimulus in a Simon's task, where a right/left-located visual stimuli of a square with a frame of varying colours were presented to participants. They were asked to discriminate against either the colour of a square or the colour of a square's frame using a left or right index finger keypress. This resulted in a Simon effect, in which the response times were faster when the required response is spatially congruent with the stimulus location, showing that performance is facilitated almost automatically for spatially congruent stimulus-response pairings. It is necessary to clarify that the term 'automaticity' is defined in terms of the level of processing of which very little cognitive control occurs, and in the context of imitation task, a response follows automatically once a compatible and task-relevant stimulus are instantiated (Moors & de Houwer, 2006).

The conundrum regarding higher-level cognitive processing in the context of action perception and motor planning is a difficult one, as the role of attention control remains ambiguous and without a strong consensus in the present literature. Heyes and Catmur (2021) proposed that mirror neurons are involved in perceiving actions with low-level processing (e.g., precision of hand grip), but not those requiring higher-level processes (e.g., inferring actions of others). Indeed, the most prominent discussions of automaticity describe three main characteristics of this concept: effortless, stimulus driven, and operates outside awareness (Laberge & Samuels, 1974; Schneider & Shiffrin, 1977).

Motivational account and relationship with automatic imitation

Given that the underlying process governing automatic imitation is still unclear and under debate, studying reward might explain if automatic imitation is truly automatic or if it can be manipulated by an intentional decision process. Motivation, which is driven by social and psychological factors, can be modelled in many ways but is categorized into two famously known archetypes: intrinsic (autonomy and purpose) and extrinsic motivations (reward and punishment; Reiss, 2012). The extrinsic aspect was of our interest as it was presumably easier to manipulate through experimental design. Additionally, Soutschek et al. (2014) demonstrated the association between monetary reward and conflict processing in a Stroop task, requiring participants to respond to the colour of a word whilst ignoring its meaning. The main finding of their experiment was that the response time did not differ between the congruent and incongruent trials when there were monetary incentives. The performance-dependent paradigm of the experiment could enhance focus of attention and perhaps cognitive control, which would subsequently facilitate conflict resolution for the incongruent trials. In the present work, we were interested in the extrinsic factor of motivation (monetary incentives) in controlling reward sensitivity, and its effects on a goal-directed behaviour. Specifically, we tested the influence of monetary incentive on automatic imitation effect, examining whether the prospect of monetary reward could inhibit or reduce the automatic imitation effect, which would be reflected by a larger reaction time difference between congruent and incongruent trials.

Recent evidence from (Krebs & Woldorff, 2017b) demonstrated that motivation could modulate cognitive processes that subsequently facilitate automatic imitation, i.e., a larger response time difference between congruent and incongruent trials. Similarly, Krebs et al. (2010) observed that reward prospect improves processing of task-relevant information resulting in faster and more accurate response in a colour-naming Stroop task. A subsequent study by Krebs et al. (2013) observed event-related potentials (ERP) in a similar task paradigm. They found that behavioural facilitation in potential-reward trials was linked to early fronto-central and occipital ERP modulations, indicating increased attention for task-relevant stimulus components related to reward prediction. Furthermore, the attentional reinforcement appeared to regulate the temporal dynamics of conflict processing, allowing behavioural interference from task-irrelevant stimuli in potential-reward trials to be reduced. This finding is consistent with a study by Prével et al. (2021), which tested the effect of non-instructed prospects of monetary gains and positive affect stimuli on cognitive control using the AX-Continuous Performance Task (AX-CPT) in which participants to respond to a prompt (X) only when it is preceded by a cue (A) using either the left or right arrow-key to respond. The results showed a significant improvement in the task following positive outcomes (monetary reward and positive affective stimuli), suggesting that non-instructed outcomes possibilities may influence cognitive control. Therefore, we would predict that an SRC effect would appear under reward manipulation through an SRC task which measures the reaction time difference between congruent and incongruent stimulus-response trials.

On the contrary, if automaticity is independent of cognitive control processes, we would expect that automatic imitation would not be affected by motivation as is argued in several studies. Genschow et al. (2022) through a series of experiments including an imitation-inhibition in a finger lifting task reported that a group membership, i.e., in- and out-group memberships in a social environment did not affect automatic imitation. Despite not finding any social influence on automatic imitation, their findings have an influence on the motivational account of automatic imitation in social modulations as it posits that a sense of affiliation (group memberships) does not enhance motivation to imitate others in a social interaction. The generalizability of published studies concerning motivational influence on automatic imitation is still lacking and requires a systematic approach to contribute to the current motivational theories of imitation.

Present study

Nevertheless, most previous studies have yet to elucidate the role of motivation in modulating automatic imitation effect in an SRC paradigm. Subsequently, there are two aims in the present study:

I. To replicate the automatic imitation effect of congruent stimulus in an SRC finger tapping task from Brass et al. (2000) in an online setting.
II. Establish whether participants with a prospect of reward, i.e., monetary incentive (reward group) would perform differently than participants with no incentive (no-reward group) in the SRC task.

The goal of the research was to test the main hypothesis that the automatic imitation effect is stronger for participants with monetary incentives (reward group) compared to those without monetary incentive (no-reward group) in the SRC task. We predicted that if automatic imitation can be modulated by cognitive processes, an enhanced performance is expected, i.e., automaticity is more diminished in the reward manipulation compared to in the no-reward manipulations.

We utilized an online implementation of the SRC task to test our predictions. Automatic imitation was assessed through an SRC finger tapping task, which was a modified version of (Brass et al., 2000; Brass et al., 2001). To manipulate the reward group, we instructed our subjects at the beginning of the task that they would be rewarded with a bonus payment of £2 if they performed in the top 10 percent in terms of accuracy and speed. While the no-reward group was not given such monetary incentive, participants in this group were compensated the same amount as the reward group at the end of the study. We assessed the reaction time (RT) for a response following a prompt accompanied with a video distractor which could either match the prompted response action (compatible) or mismatch it (incompatible).

Methods

Participants

The study had initially recruited a total sample size of 151 participants through the Prolific platform. After subsequent drop-outs, opt-outs of the study, etc., only 83 participants (41 in the reward group and 42 in the no reward group) remained who successfully completed the study. We later excluded four additional participants from further analysis due to technical issues while completing the study, and another two for high error rates (ERs; See Table 1). The final analysis consisted of 39 participants in the reward group (26 females; mean age = 24.08 years, SD = 3.63), and 38 in the no reward group (28 females; mean age = 24.32 years; SD = 3.93). Participants gave consent of their participation, and were paid equally upon completion of the study. All were native-English speakers, no history of language related disorders, had corrected-to-normal vision, right-handed and 18 to 30 years of age. The study was approved by UCL's Research Ethics Committee with ethics code 15365.001.

Materials

The experiment was administered online on the Gorilla platform, where participants completed a SRC finger tapping task adapted from Brass et al. (2001) displayed on their computer screen. Participants were instructed to respond to prompts presented on a computer screen by a left or right arrow key press. As shown in Figure 1A, the symbolic cues ("!" and "#") were displayed simultaneously on top of a video distractor of upright index or middle finger tapping. The distractors matched or mismatched the key press response prompted by the symbolic cues. Congruency refers to compatibilities in perceived and performed actions, where the video distracter matches the prompt in a trial, while in an incongruent trial, the video distractor and prompt are mismatched. An attention check task was included to monitor participants' attention and active

participation in this online study. The task was embedded in the SRC task, whereby randomly once within each block a trial displayed either 3, 5, or 9 dots, and participants had to answer the number of dots displayed in the trial (see Figure 2). The trials were important to ensure that participants paid attention throughout the experiment.

The videos were filmed using a Canon Legria video camera and edited in Final Cut Pro. Clips were extracted representing the actor's left hand in a neutral position, with the left of middle finger down (see Figure 1B). The hand images were rotated 90 degrees counterclockwise and displayed as 50% of their original size. Each stimulus video lasted 3,000 ms, beginning with the hand in resting position for 500 ms or 740 ms (jitter) before finger movement (lasting 1000 ms), and ending with the hand in resting position (lasting 1500 ms or 1250 ms). The prompts were size 63, font Helvetica in white. The files were saved in .mov format before being exported as .mp4 files.

Table 1: *Exclusion criteria and numbers at each stage of the study*

Study stage	**Initial *n***	**Exclusion criteria**	***n excluded***
Main study	151	Withdrew from the study	57
		Did not complete the experiment	4
		Failed the attention check	7
		Total exclusions	**68** (45.03%)
Data processing	83	Error rate >3 SD from the group mean	2
		Technical issues during the experiment	4
		Total exclusions	**6** (7.23%)

Note. *n = number of participants; final n = 77 (39 in the reward group, 38 in the no reward group).*

Procedure

The study was conducted online entirely, in which participants completed a computer-mediated experiment. A between-subject design was employed to investigate the effects of congruency (congruent vs. incongruent) and reward condition (incentive vs. no incentive) on reaction times (RTs) in the SRC task. The task consists of six blocks of 40 counterbalanced and randomized trials each, which in total took on average 16 minutes to complete. Two versions of stimulus associations were created to counterbalance the symbol-action pair-

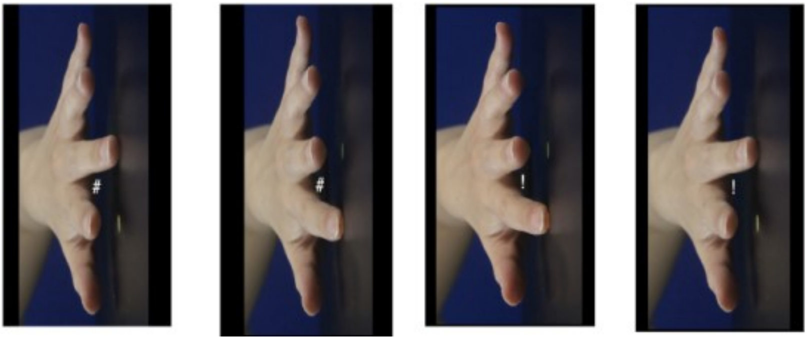

Congruent Incongruent Congruent Incongruent

Figure 1a. *Four possible end-frame of congruent and incongruent trials for "#" and "!" prompts in the first type of stimulus association.*

ings. The stimulus in the first association required a left arrow key press response to a "!" prompt and a right arrow key press response for a "#" prompt, whereas the prompt and response was reversed in the second stimulus association, i.e., a "!" prompt required a right arrow key press response. Participants were assigned to either one of the stimulus associations. Additionally, a prompt onset was controlled for at a jitter of 500 and 740 milliseconds after a trial begins to avoid any possible interference effects. Each participant participated in either the reward or no reward group. In the reward group, participants were told about monetary incentive (£2 bonus) if they performed in the top 10 percent in terms of speech and accuracy while the other group was not informed of such incentive. All of them were paid equally (£4 in total), regardless of the condition, at the end of the study.

The distinction between reward and no-reward groups was that, in the reward group, participants were told of the monetary incentive at the beginning of the task. In each group, participants completed the information sheet and consent form before proceeding to the task. Following this, the participants were informed of the SRC task structure together with training trials to familiarize themselves with

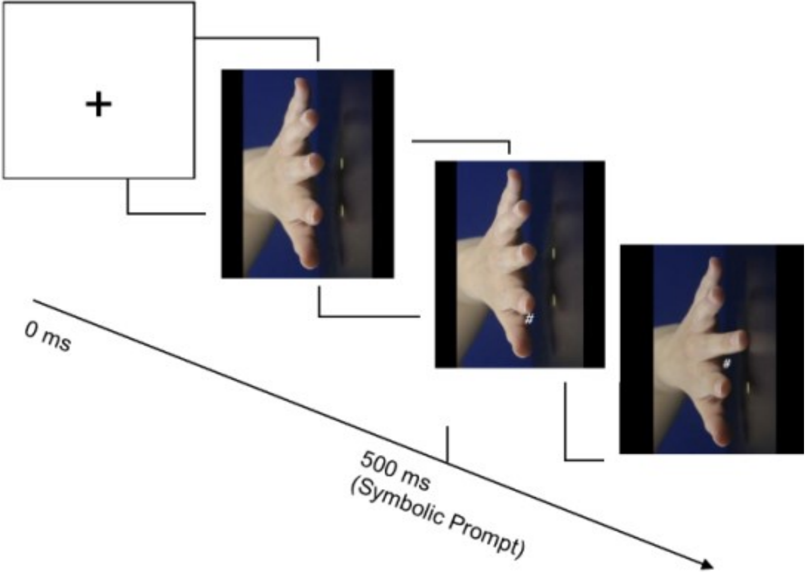


Figure 1b. *An example of timeframe of a congruent trial in a testing session for the first stimulus association, a "#" prompt and right middle finger tapping video distractor pair, with the prompt presented at 500 ms.*

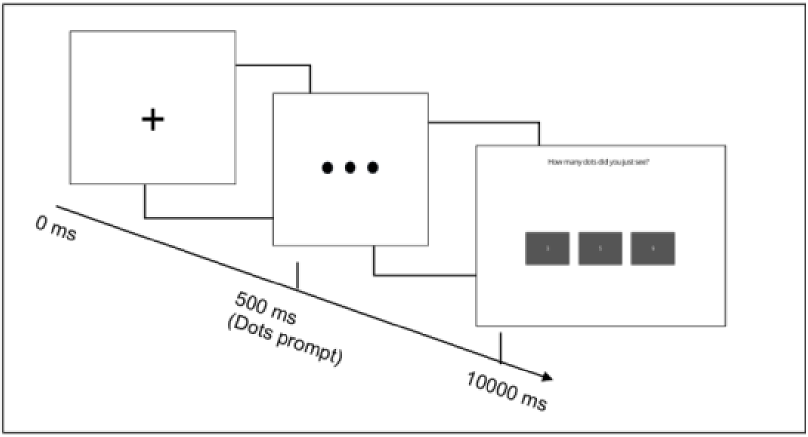


Figure 2. *A timeline of attention check trial presented after a series of SRC task trials. The video lasted for 10,000 ms, with the prompt presented at 500 ms. In the last frame, participants had to choose out of three options which matched the prompt displayed.*

the stimuli and compatible responses before proceeding to the test phase. In the first stimulus association, half of the participants were instructed to respond with a left arrow key press using their right index finger when a "!" symbol appeared, and a right arrow key in response to the appearance of a "#" symbol using the right middle finger. Video stimuli displayed a symbolic cue in between the right index and the middle fingers. In a congruent trial, a symbol "!" is matched with a right index finger video stimulus, and this prompt must be followed by a right arrow key press, signifying a compatible response (Figure 1B). In an incongruent trial however, a video stimulus is incompatible to a symbolic cue, i.e., an index finger tapping distractor was displayed for a "!" cue. Conversely, in the second stimulus association, the instructions of the prompt were reversed for the rest of the participants, i.e., they had to respond with a left arrow key press using right index finger when a "#" prompt appeared, whereas a right arrow key press response using the right middle finger when a "!" prompt appeared. Manual responses were recorded automatically, with RT being the onset of key press relative to the onset of response prompt. An accurate response was determined by a key press as indicated by a prompt shown in a trial.

Analysis

Participants with error rates greater than three standard deviations away from the group mean and participants who perform <=80% of 'catch trials' (three per block, 18 trials in total) were excluded from the analyses. Only correct trials were included in analysis, meaning that correct trials which fall within three median absolute deviations (MADs) from participants' median RT in each condition (for correct trials) were included, whereas incorrect responses and responses with RTs <200 ms or >1000 ms were removed from analysis. The effects of congruency (congruent vs incongruent) and monetary incentives (reward vs no reward) on the RT for accurate trials in the SRC task were investigated using a linear mixed model (LMM). A comprehensive model was structured to account for the random effect of by-participant random intercepts for all fixed factors and interactions, as well as congruency (congruent vs incongruent), and monetary incentives (reward vs no-reward). In this approach, a forward model construction technique is employed, as recommended by Barr et al. (2013), starting with random effect only, with successive addition of a fixed effect and interactions until the best model is obtained. Then, an analysis of variance (ANOVA) test is conducted to determine if the best fit model significantly improved its previous model. Aside from the p-value based stepwise selection, the model with lowest Bayesian Information Criterion (BIC) value was chosen as it supported the model with better quality relative to another model, and it penalised model complexity more heavily than the Akaike Information Criterion (AIC; Rouder et al., 2016). The hypotheses of the study were:

I. The RTs for congruent trials would be faster than the incongruent trials, demonstrating an automatic imitation effect.
II. There would be a substantial difference in RT in the SRC task between the participants in the reward group and those in the no-reward group.
III. The automatic imitation effect would significantly be stronger in the reward group and the no-reward group.

It was predicted that the model with all main effects and interactions, as well as random intercept of participants would be a significant fit to the data. The key interaction central to our hypotheses was between condition and congruency, as it signifies whether the automatic imitation effect could be influenced by the prospect of monetary gains. The following information were recorded from each trial:

- Reaction time (RT): The time from the prompt onset (i.e., symbolic cue) to the time when the participant responds with a key press, measured automatically with Gorilla in milliseconds (ms).
- Accuracy: Whether a key press response matched the prompt given; 0 (incorrect) and 1 (correct)

Results

The data was exported from Gorilla (in a .csv format) and was analyzed with R (RStudio 2021.09.0 version). Prior to analysis, participants with technical issues, error rates >3 standard deviation (SD), and incorrect trials were excluded from the data. Additionally, the median absolute deviation (MAD) was calculated for each participant and condition, and RTs which deviated by 3 MAD were removed, as suggested by Leys et al. (2013). After the errors and outliers were omitted, RTs distribution were calculated for each condition, congruency, and jitter (see Figure 3 and Table 2 below).

Main Analyses

A linear mixed model (LMM) analysis was conducted in R using lme4 package and lmerTest package, which implemented the Satterthwaite's method to generate p-values for significance testing and degree of freedom estimates for mixed models (Bates et al., 2015; Kuznetsova et al., 2017). The model includes RT with conditions, congruency, and jitter as main predictors, condition and congruency as interaction terms, and the random factor of participants to test our hypotheses. We ran a first model with only a by-participant random factor to observe if random effects of participant influence the RTs prediction. Unlike previous studies (Genschow et al., 2022; van den Berg et al., 2014) which reported comparison of means for each group, the inclusion of random effects of participants would enhance our prediction of means and variance for a general population that were not part of this study. In this model, 43.81% of the variance is explained by the random effect of participants.

A similar model with addition of fixed effects of condition, congruency, and jitter was created. In this model, there were significant main effects of conditions (β= -30.18, t = -2.10, p = .0394), congruency (β = -8.69, t = -7.67, p < .001), and jitter (β = -0.089, t = -18.76, p < .001). The random effect of participants explained 43.40% of variance which was not explained by the fixed effect model. We compared the current model with fixed and random effects with the previous model with only random effect using ANOVA, which applied the maximum likelihood (ML) method for model selection. The purpose of the test was to determine whether a more complex model was significantly better to capture the data than a simpler model (Phillips, 2017). We found a significant result, p < .001, and lower BIC value, BIC = 183902, indicating that the addition of condition, congruency, and jitter fixed factors significantly improved the model fit.

We further tested the interaction between condition and congruency as these factors were central to our hypothesis that automatic imitation could be mediated by reward. In the third model, we fitted an interaction effect between condition and congruency. The results showed that the main effects were significant; condition (β = -30.31, t = -2.10, p = .0393), congruency (β = -8.81, t = -5.41, p < .001), and jitter (β = -0.089, t = -18.76, p <.001). However, the interaction between condition and congruency was not significant (β = 0.23, t = .100, p = .920). The random effect of participants in this model explained 43.39% of the variance which was not explained by this model. An ANOVA test was conducted to compare between this model with interaction of congruency and conditions. The result was not significant, p = .9203, this model with interaction term had a higher BIC value, BIC = 183911, hence, adding an interaction term of con-

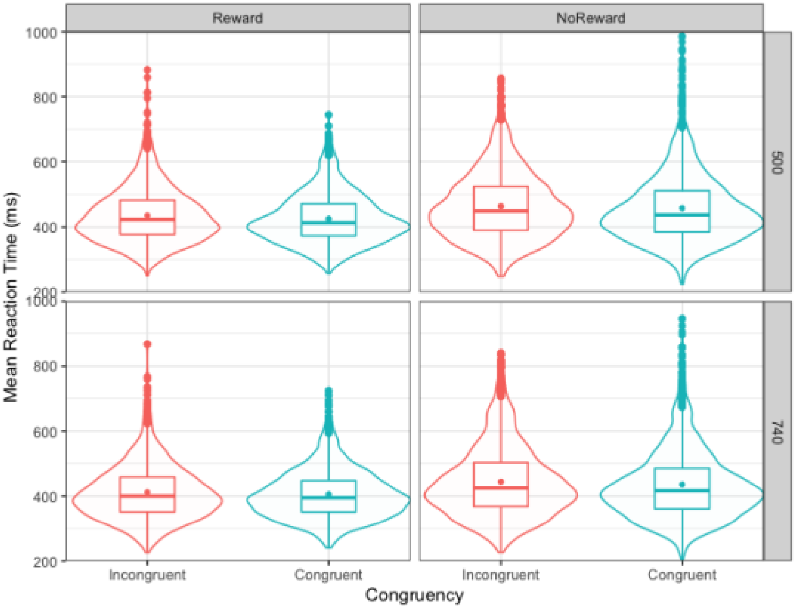


__Figure__ 3. Reaction times distribution in each condition, congruency, and jitter (the dots within each boxplot represented means).

gruency and condition did not improve the model fit.

Taken together, the overall results suggest that the second model with fixed effects of condition, congruency, and jitter, with the random effect of participants, was the best model for our data. In summary, our results suggest no evidence of monetary incentives on the automatic imitation between the reward prospect and stimuli congruency, contrary to our prediction that the automatic imitation effect would be stronger for the reward group which received monetary incentive than the no-reward group.

Table 2: *Mean reaction times in milliseconds (ms), SOA = stimulus onset asynchrony or jitter, standard deviations (SD), and standard errors (SE) for each group*

Group		SOA (ms)	RT (ms)	SD (ms)	SE (ms)
Reward	Congruent	500	424.7	76.75	1.673
		740	404.8	74.14	1.620
	Incongruent	500	435.0	84.20	1.862
		740	411.4	85.15	1.874
No-reward	Congruent	500	457.2	109.2	2.482
		740	435.6	109.5	2.458
	Incongruent	500	463.3	102.7	2.340
		740	444.2	109.3	2.473

Discussion

The study's aim was to determine whether automatic imitation is stronger for congruent trials than incongruent trials in an SRC key-pressing task, and if this effect was modulated by a reward manipulation. The study adapted an SRC finger tapping task by Brass et al. (2000; 2001) in combination with a monetary incentive reward manipulation.

Automatic imitation effect

It was hypothesized that the RT for congruent trials would be faster in the congruent trials than the incongruent trials. Numerous studies have demonstrated the automatic imitation effect through an SRC paradigm, in which compatible stimulus-response pairings produced faster response times than incompatible stimulus-response pairings (Brass et al., 2000; Longo et al., 2008). This study found a similar effect of congruency to the present paper, in which the results showed a significant RTs difference between congruent trials than incongruent trials. These outcomes then provide general support for our hypothesis that there was an automatic imitation effect in the SRC task, reflected by a faster RTs in the congruent trials (compatible stimulus-response pairing) than in the incongruent trials (incompatible stimulus-response pairing).

A similar trend in the existing literature using the SRC task was that the spatial compatibility effects (where stimuli and response were spatially compatible to participants) could have passed the automatic imitation effect (Brass et al., 2000; Pan & Hamilton, 2015). The SRC task in our study was structured to control or at least minimize the compatibility effects through orthogonal presentation of video stimuli, allowing for a clear effect of automatic imitation. These results are in agreement with those obtained by Jiménez et al. (2012)

who also reported an automatic imitation effect after controlling for spatial compatibility effects using orthogonal stimuli arrangements. Overall, our findings successfully replicated the automatic imitation effect found in previous studies with the familiar SRC key-pressing task, as well as demonstrating this effect while controlling for spatial compatibility in a manner coherent with the study by Jiménez et al. (2012).

Another important finding was that RTs were significantly faster for jitter of 740 ms than 500 ms, when the prompt onset was displayed later relative to the video distractor. This result seems to be partially consistent with the other study on automatic imitation of vowels using the SRC paradigm (Adank et al., 2018). This study used a broader set of stimulus onset asynchrony (SOAs); with four SOAs, and found a greater automatic imitation in both experiments at the later SOA, i.e., SOA3. In this study, since there was no interaction between congruency and jitter, we could not conclude that automatic imitation was largest for a later jitter/SOA, hence the main effect of jitter explained that participants were better at discriminating incongruent trials when the prompt appeared later in a trial at 740 ms.

Prospect of reward does not influence automatic imitation

We hypothesized that there would be considerable difference in overall RTs between the reward and no-reward group. Our results support this prediction since there was a significant main effect of condition predicting RTs in the SRC task. The reward manipulation worked as intended, as the reward group who had the false belief that they were rewarded more, performed better. The results support a previous finding by Yamaguchi & Nishimura (2019), who investigated the influence of reward anticipation on a flanker task performance in which participants responded to the targeted colours stimuli while ignoring the accompanied flankers with identical colours. They found faster RTs were associated with the anticipation of performance-dependent rewards, which aligns with the present study where participants were instructed that the monetary reward would be granted for reaching performance levels in the top 10% of the study. Bianco et al. (2021), in their study of participants' engagement in online auditory tasks, also found an improvement in listening performance when a monetary bonus was granted, which is interestingly similar to participants' performance in a lab setting. Consistent with the literature, this research found that the participants in the reward group could have paid more attention to the task than the participants in the no-reward group. They also caution that, because online participants are unsupervised, anonymous, and motivated by monetary incentives, improving the quality of an online study is necessary to improve its reliability. The author also mentioned that the monetary incentives may have no influence if the task is too simple, or if the return on effort is low, i.e., it is difficult to improve performance (Bianco et al., 2021).

Nevertheless, the main focus of our study was to test the hypothesis that the automatic imitation effect was stronger for the reward group than the no-reward group. Our results, however, do not provide evidence for this prediction as the interaction between congruency and condition was not statistically significant. Previous studies have demonstrated mixed results with regards the effect of rewards on cognitive control. Padmala & Pessoa (2011) showed that trials which were cued with reward had a faster response speed, implying an increase in proactive control, which is consistent with previous findings that reward incentives improve attentional control and reduce conflict. This appears to not be the case in our study in which such effects were not detected.

Furthermore, we could argue that the automatic imitation is truly automatic, and not controlled regardless of reward manipulations. Research by Michałowski et al., (2017) adapted a go/no-go task, where participants had to quickly respond to a go signal and inhibit a no-go signal, with the addition of monetary incentives (rewards) in the instruction at the beginning of each trial, i.e., high, medium, or no rewards. They observed that the performance in the inhibitory go/no -go task did not differ regardless of reward magnitudes. As such, the effect of monetary incentives in the current study could have influenced performance in the SRC key-pressing task, but not significantly when the congruency effect was accounted for. In other words, automatic imitation in the SRC task may be independent of monetary incentives.

Limitations and future research

The first limitation of the current paper is the online experimental setup which might have failed to capture the effect of monetary incentives on automatic imitation. Indeed, several previous studies which explored the effect of reward on improvement of cognitive control were carried out in a lab setting (Prével et al., 2021; Yamaguchi & Nishimura, 2019). Perhaps, if the study was carried out in a lab, the results would yield a significant effect of reward on automatic imitation. The reason being was that in a lab setting, experimenters could monitor participants' engagement on the task, which otherwise was restricted in an online study. On that note, future online studies should check participants' performance across blocks of trials, whether there could be a diminishing performance due to fatigue effect from online studies. This might not be the case in this study as the experiment lasted for ~20 minutes on average. Despite the limitation, our study succeeded at capturing the automatic imitation effect, as shown by faster RTs in the congruent trials than the incongruent trials.

Moreover, the study only mentioned the monetary incentives once before the first block of trials in the treatment group. Krebs & Woldorff (2017), suggested an alternative reward paradigm in a form of feedback after each trial or block to participants. The distinction between reward prospect and reward granted context could heavily influence the performance in the SRC task. Thus, future studies should explore influences on automatic imitation under different aspects of rewards manipulations, such as high vs low reward magnitudes, punishment of incorrect response, and performance change when reward is granted for correct trials.

In this regard, potential future directions of research in this area were proposed. Future studies should first discern the automaticity of a behaviour, and what role awareness has in the cognitive processing of this behaviour. One of the main considerations of covert imitation of action is to discern the automaticity of behaviours, i.e., how does intentional control or unconscious process modulate an imitation and if it is possible for this to be understood through neuroimaging and event-related potentials based methods. Behaviourally, a similar experimental setup with multiple SOAs predicting RT in an SRC task could lead to a clearer distinction between automatic and deliberate cognitive control. Nonetheless, while these outcomes are an important step toward developing an integrated automatic imitation theory, they are currently largely unspecified at the process level. Furthering the current reward paradigm, by incorporating a broader and more diverse set of motivational factors, holds great promise for moving the field forward.

Conclusion

This study explored the influence of monetary incentives on automatic imitation. The results showed an overall automatic imitation effect, and an overall effect of monetary incentive on task performance. However, we found that the automatic imitation effect was not modulated by the monetary incentive. Further studies could explore the use of alternative reward manipulations to further explore how the mental processes involved in the SRC task differ from other tasks, such as the Stroop task and flanker task.

References

Adank, P., Nuttall, H., Bekkering, H., & Maegherman, G. (2018). Effects of stimulus response compatibility on covert imitation of vowels. *Attention, Perception, and Psychophysics, 80*(5), 1290–1299. https://doi.org/10.3758/s13414-018-1501-3

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 10.1016/j.jml.2012.11.001. https://doi.org/10.1016/j.jml.2012.11.001

Bianco, R., Mills, G., de Kerangal, M., Rosen, S., & Chait, M. (2021). Reward enhances online participants' engagement with a demanding auditory task. *Trends in Hearing, 25,* 23312165211025940. https://doi.org/10.1177/23312165211025941

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition, 44*(2), 124–143. https://doi.org/10.1006/brcg.2000.1225

Butler, E. E., Ward, R., & Ramsey, R. (2016). The influence of facial signals on the automatic imitation of hand actions. *Frontiers in Psychology, 7.* https://www.frontiersin.org/article/10.3389/fpsyg.2016.01653

Catmur, C. (2016). Automatic imitation? Imitative compatibility affects responses at high perceptual load. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 530–539. https://doi.org/10.1037/xhp0000166

Chartrand, T. L., & Bargh, J. A. (2002). Nonconscious motivations: Their activation, operation, and consequences. In A. Tesser, D. A. Stapel, & J. V. Wood (Eds.), *Self and motivation: Emerging psychological perspectives* (pp. 13–41). American Psychological Association. https://doi.org/10.1037/10448-001

Genschow, O., Westfal, M., Cracco, E., & Crusius, J. (2022). Group membership does not modulate automatic imitation. *Psychological Research, 86*(3), 780-791. https://doi.org/10.1007/s00426-021-01526-1

Heyes C. (2011). Automatic imitation. *Psychological bulletin, 137*(3), 463–483. https://doi.org/10.1037/a0022288

Heyes, C., & Catmur, C. (2022). What happened to mirror neurons? *Perspectives on Psychological Science, 17*(1), 153–168. https://doi.org/10.1177/1745691621990638

Jiménez, L., Recio, S., Méndez, A., Lorda, M. J., Permuy, B., & Méndez, C. (2012). Automatic imitation and spatial compatibility in a key-pressing task. *Acta Psychologica, 141*(1), 96–103. https://doi.org/10.1016/j.actpsy.2012.07.007

Krebs, R. M., Boehler, C. N., Appelbaum, L. G., & Woldorff, M. G. (2013). Reward associations reduce behavioral interference by changing the temporal dynamics of conflict processing. *PLoS ONE, 8*(1). https://doi.org/10.1371/journal.pone.0053894

Krebs, R. M., Boehler, C. N., & Woldorff, M. G. (2010). The influence of reward associations on conflict processing in the Stroop task. *Cognition, 117*(3), 341–347. https://doi.org/10.1016/j.cognition.2010.08.018

Krebs, R. M., & Woldorff, M. G. (2017a). The Wiley handbook of cognitive control—2017—Egner—cognitive control and reward. *The Wiley Handbook of Cognitive Control,* 442–439. https://doi.org/10.1002/9781118920497.ch24

Krebs, R. M., & Woldorff, M. G. (2017b). Cognitive control and reward. In T. Egner (Ed.), *The Wiley Handbook of Cognitive Control* (pp. 422–439). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118920497.ch24

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Longo, M. R., Kosobud, A., & Bertenthal, B. (2008). *BIROn-Birkbeck Institutional Research Online.* https://eprints.bbk.ac.uk/policies.html

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences, 22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Michałowski, J. M., Koziejowski, W., Droździel, D., Harciarek, M., & Wypych, M. (2017). Error processing deficits in academic procrastinators anticipating monetary punishment in a go/no-go study. *Personality and Individual Differences, 117,* 198–204. https://doi.org/10.1016/j.paid.2017.06.010

Mondillon, L., Niedenthal, P. M., Gil, S., & Droit-Volet, S. (2007). Imitation of in-group versus out-group members' facial expressions of anger: A test with a time perception task. *Social Neuroscience, 2*(3–4), 223–237. https://doi.org/10.1080/17470910701376894

Moors, A., & de Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297–

326. https://doi.org/10.1037/0033-2909.132.2.297

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience, 23*(11), 3419–3432. https://doi.org/10.1162/jocn_a_00011

Pan, X., & Hamilton, A. F. de C. (2015). Automatic imitation in a rich social context with virtual characters. *Frontiers in Psychology, 6*(JUN). https://doi.org/10.3389/fpsyg.2015.00790

Phillips, N. D. (2017). *YaRrr! The Pirate's Guide to R.* Retrieved from https://bookdown.org/ndphillips/YaRrr/

Prével, A., Hoofs, V., & Krebs, R. M. (2021). Effect of non-instructed instrumental contingency of monetary reward and positive affect in a cognitive control task. *Royal Society Open Science, 8*(8). https://doi.org/10.1098/rsos.202002

Reiss, S. (2012). Intrinsic and extrinsic motivation. *Teaching of Psychology, 39*(2), 152–156. https://doi.org/10.1177/0098628312437704

Rouder, J. N., Engelhardt, C. R., McCabe, S., & Morey, R. D. (2016). Model comparison in ANOVA. *Psychonomic Bulletin and Review, 23*(6), 1779–1786. https://doi.org/10.3758/s13423-016-1026-5

Scerrati, E., Lugli, L., Nicoletti, R., & Umiltà, C. (2017). Comparing Stroop-like and Simon effects on perceptual features. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-18185-1

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Soutschek, A., Strobach, T., & Schubert, T. (2014). Motivational and cognitive determinants of control during conflict processing. *Cognition and Emotion, 28*(6), 1076–1089. https://doi.org/10.1080/02699931.2013.870134

van den Berg, B., Krebs, R. M., Lorist, M. M., & Woldorff, M. G. (2014). Utilization of reward-prospect enhances preparatory attention and reduces stimulus conflict. *Cognitive, Affective and Behavioral Neuroscience, 14*(2), 561–577. https://doi.org/10.3758/s13415-014-0281-z

Yamaguchi, M., & Nishimura, A. (2019). Modulating proactive cognitive control by reward: Differential anticipatory effects of performance-contingent and non-contingent rewards. *Psychological Research, 83*(2), 258–274. https://doi.org/10.1007/s00426-018-1027-2

Hexad
The cognitive science hexagon
And the wonderful theories that spawn
  Vital research, it's true
  But from just me to you
Reading scholarly journal's a yawn

Head (Part 4)

Angel Martin

A Semiospheric Interpretation of Pragmatic Difficulties in Autism Spectrum Conditions (ASC)

Kenny J. Fisker

Aarhus University

Department of Linguistics, Cognitive Science and Semiotics

Abstract

In recent years, the neurodiversity concept and its coherence with contemporary theories on brain function have reframed perspectives on pragmatic difficulties in Autism Spectrum Conditions (ASC), consequently acknowledging the intersubjective space as a prominent part of the puzzle. In order to concretize the paradigm shift, this thesis takes the subdivided definition of "high-functioning" autism as a paradigm example, reviewing old and new cognitive theories through the neurodiversity paradigm, and examining possible cognitive bases for the pragmatic difficulties which arise in cross-neurotype communication. A baseline of contemporary cognitive theory is established through the predictive-coding hypothesis, which postulates that sensory processing and meaning-making processes of the mind are conducted through cortical means of Bayesian inference. In its framework and contrasted with a neurotypical counterpart, the autistic brain is defined as developing its internal model of the world on a neurologically different foundation of sensory and predictive processing which leaves it in large degrees unable to intuitively adapt to and differentiate between subjectively important sources of stimuli in noisy environments—those self-same environments in which the neurotypical brain thrives in, by virtue of its predictive style being predicated on generalizing sensory information. This foundation is perceived through the notion of the semiosphere, which posits that these differences in mind-world relations permit different systematizations of semiotic information in the semiosphere. From this interpretation, the autistic brain is defined as immersing itself in a semiotic continuum which in large parts is not made for it, thereby being caught in a tug-of-war between constant background acts of translation and more effortful searches for relevance.

Keywords: **Pragmatic difficulties, Autism Spectrum Conditions, neurodiversity paradigm, cross-neurotype communication, predictive-coding hypothesis, the semiosphere .**

Introduction

In a world of spheres, of biospheres and noospheres, of asymmetry, diversity, and complexity, the biocultural niche of language has managed to constitute the minds of humankind into a sphere of semiosis, where meaning exists beyond the individual mind and subsists through patterns of self-organization in communication (Lotman, 1984; Schweitzer & Zimmermann, 2001; Sinha, 2014). Through this, the creation of a symbolically mediated biocultural complex of humankind, with the principles inherent in complex adaptive systems dictating the creation of niches within the semiosphere, has allowed languages and cultures to evolve into the constituent, heterogeneous systems and patterns of meaning that we now know today (Semenko, 2012; Sinha, 2015). It is in this world of translatable values, of semiotic[1], cultural, and neurocognitive diversity, that fluency is a term that denotes both boundaries of communication and efficacy of expression; however, it is a world where the autistic condition and the core tenet of neurodiversity has yet to frame itself and yet to find its place within the larger constitution of semiotics and pragmatics in translation, its place in the foundation of shared meanings, and thus its place in the semiosphere.

The notion that autism denotes a spectrum of conditions is a relatively novel concept. A hundred years ago, the word "autism" was coined in reference to the self-absorbed and introspective symptoms of schizophrenia, in which the inner self dominated over reality (Frith, 2008). It was not until the case studies of Leo Kanner (1943)

---

[1] **Semiotic :** relating to signs and symbols, which themselves signify a meaning. These signs and symbols remain as vessels or constitutive elements of a meaning, i.e., only conveying the meaning through arbitrary representation (consider how a word signifies a meaning, rather than itself). The sign is composed of the signifier (the form constituting the sign) and the signified (the real-world object represented by the signifier).

and Hans Asperger (1944) on “autistic disturbances” and “autistic psychopathy” in children that the term was used to denote an independent developmental disorder, marked by impairments to social and motoric functions. But even with this description of abnormal behaviour, several radical shifts in concept would occur with the emergence of epidemiological studies, categorizing features of autism into syndromes through deficit-based accounts. Throughout the decades leading up to the 21st century, the hypotheses of Baron-Cohen (1997) and Frith (2008) would classify the root causes of communication troubles and introspective traits to arise from a general mindblindness and an inability to put information into context. In recent years, autism research has encountered a significant paradigm shift, critiquing previous theories of pathology (Davis & Crompton, 2021; Fletcher-Watson, 2019; Nicolaidis et al., 2019). With its foray into popular literature, and consequently larger masses of people and researchers alike, the scientific literature around autism research has begun to redefine itself through the neurodiversity paradigm (Armstrong, 2010; Rosqvist et al., 2020; Silberman, 2015)—several authors now propose distancing the field of terminology from deficit -based “patient”-versus-“healthy”-dichotomous descriptions of autism by replacing “disorders” with “conditions” (Bolis et al., 2017).

From a linguistic standpoint, this paradigm shift presents an opportunity to study the larger narrative of pragmatics in cross-neurotype interactions, through angles which might not have been possible in a previously more restricted paradigm. Here, the assumption is that the breakdown of mutual pragmatic understanding can be shown to originate from a variable number of differences between relevant interlocutors, namely through their neurological and cognitive differences. Based on the subject matter of the neurodiversity paradigm and its relevant studies, this paper aims to contextualize these pragmatic difficulties in cross-neurotype communication through the concept of the "semiosphere" to consider how mind-world relational aspects of semiosis and immersion into a semiotic continuum affects a neurodiversity of minds. Autism Spectrum Conditions (ASC) are chosen as the paradigm example to analyze and discuss the literature around neurodiversity, as well as establish contemporary theories on cognition that directly correlate with notions of autistic cognitive function. To alleviate any concerns of romanticizing actual disorders of the mind, this paper chooses to predominantly focus on the high-functioning side of the autistic spectrum, of autistic people who exhibit no intellectual disabilities and whose main symptoms are related to neuronormative attributions of deficits in social communication and emotion recognition, and who have been disenfranchised by a seemingly rampant culture of disability.

Establishing a Neurodiversity Paradigm: Cognitive Theories of Autism

In accordance with the diagnostic requirements described in the World Health Organization's eleventh revision of the International Classification of Diseases (ICD-11) (the equivalent of the DSM-5-TR), the core features of Autism Spectrum Conditions (ASC) consist of life-long persistent abnormalities in areas of social communication, sensory experience, and patterns of behaviour—the latter of which involves inflexible or repetitive behaviours atypical or excessive to the sociocultural context (World Health Organization, 2019). These are measured from the expected range of neurotypical functioning in the specific cultural context, and any additional features which might involve a disorder of intellectual development, a degree of functional language impairment, or speech that lacks substantial prosody and emotional tone. As a classification of diseases, in its effort to characterize mental, behavioural, and neurodevelopmental disorders, the ICD-11 orients itself on notions of normality; it is neuronormative by design, so as to ensure consistent identification of conditions which limit capacities for sustaining and initiating neurotypical forms of social interaction and social communication—limitations which can prove problematic for its person and the people who surround them, by virtue of their aberrancy.

Relative to their respective time periods, the conception and development of cognitive theories of autism have always followed a through-line of representing the then-current ideals and knowledge of human cognitive function. As the most recent iteration, the ICD-11 takes care to mention the cultural variation of norms, and how the social abnormalities of ASC can border on normality in certain cultures. Yet, despite its apparent congruence with the neurodiversity paradigm on this aspect, and by virtue of its statistically-determined design affirming contemporary cognitive theories, the language through which it describes autistic people as the source of pragmatic failure lends credence to the deficit view of autism which had domineered past paradigms of autism research. This is due in part to Leo Kanner's influential account of early infantile autism, which led autism definition and research down a path in which the main focus was on ensuring early identification and treatment (Kanner, 1943); it therefore stereotyped autism, and its eventual extension to a spectrum, to early childhood behaviour. This meant that the deficit-based foundational studies were largely non-inclusive and limited to a neuronormative perspective of neurocognitive diversity, which eventually cascaded into rampant deficit-focused studies, where the empirical failures of the previous paradigm were owed to an absence of restraint with regards to conjecture formed from a solely neurotypical frame of reference. As pragmatic difficulties would belong to the autistic person, with the non-autistic person not beholden to any misunderstandings which arose from intersubjective communicative gaps with mentally disordered individuals, autism research of the previous paradigm found itself concluding on matters such as mindblindness (a lack of empathy and inability to mentalize) and weak central coherence (impaired holistic processing) as core deficiencies of autistic cognition itself (Baron-Cohen, 1997; Frith, 2008). Which, despite ongoing criticism and rejection on the grounds of non-replicability and methodological error sources (Gernsbacher & Yergeau, 2019; López & Leekam, 2003), these paradigms remained influential sources of publicly available information on account of confirming the stereotypical assumptions of autistic behaviour.

The Theory of Mind (ToM) deficit hypothesis assumes that coherent empathic behaviour in social interaction is reliant on the ability to make inferences about the unobservable mental states of those around us, and that autistic people are delayed in developing this system of putting oneself into another's shoes. This description of the matter ostensibly refers to the "mindreading" system as being orchestrated through four pre-packaged brain modules, each mechanism in tandem with each quadrant producing the detection and envisioning of another person's mental state: one involving a mechanism for detecting intentionality, another for eye direction, a third for the detection of shared attention, and lastly one mechanism responsible for unifying relevant possibilities into a coherent theory of mind (Baron-Cohen, 1997). While this review and investigation of autistic children adopts the perspective of autistic cognitive development being markedly different from that of normal children, it extrapolates this mind-blind condition of the autistic brain from their consistent failure to display these mechanisms either at all or in tandem with each other, from experimental set-ups designed to instigate ToM-relevant behaviours. In these experiments, autistic children are shown to have a limited or delayed capacity for engaging with joint attention, pretend play, and emotion recognition through various neurotypical socio-emotional loci.

Uta Frith, in her theory of a Weak Central Coherence (WCC), extrapolates from observations and descriptions of autistic sensory-perceptual tendencies that there is an inherent inability to orchestrate "big picture" coherence during information-processing, and consequently pragmatic context in social interaction (Frith, 2008). While the neurotypical brain is understood as being intuitively excellent at generalizing and removing unimportant details in the processing of directly visual or socially pragmatic information (i.e., seeing the forest but not necessarily noticing the trees), the autistic brain seems to be detail-oriented in its mind-world relation (i.e., seeing the trees, the leaves, the bark, the bushes, and then, if ever, noticing the forest). Although this notion still runs throughout the current neurodiversity paradigm, it has been redefined as a cognitive style, rather than a deficit—a recognition of effort being part of the equation for capacity, as a generalizing mind-world relation will be biased to its default approach, while a detail-oriented one would be equally biased to its preferred style, but both can be prompted to overcome its bias and diverge to a reversely-focused approach (Gernsbacher & Yergeau, 2019; López & Leekam, 2003).

As mentioned in the introduction, contemporary research posits less so an image of autism constituting a disability, but rather a focus on the two-way mismatch of salience between people on the autism spectrum and those of typical neurological functioning (Bolis et al., 2017). From the retrospective position of recent findings in autism research, the ToM deficit hypothesis and its constituent readings of autistic phenomena can be disputed as fragmented accounts of a greater whole: The notion of a Double-Empathy Problem (DEP) relates the bidirectional nature of empathy to a mutual mindblindness rather than belonging to one or the other. This, in turn, clarifies the pathologization of autism as a consequence of neurotypical dominance (DeThorne, 2020; Milton, 2012). DEP is the result of shifting focus from the individual to the intersubjective space, recognizing certain aspects of social dysfunction as being harboured between interlocutors rather than belonging to one or the other. In this, DEP mainly deals with the topic of empathy and the incongruence of shared socio-emotional information; this occurs when two or more people in interaction do not share the same parameters for emotion recognition, as both parts of the interaction have trouble understanding the motives and experiences of the other. This is also no less relevant to the suggestions of the WCC theory, as it could be considered a narrow description of a perceptual-cognitive style, acting as a forerunner to broader interpretations of autistic brain function: namely, the predictive-coding theory of brain function.

The Brain as a Probabilistic Prediction Machine

Predictive-coding, or the Bayesian brain hypothesis, is a neuroscientific theory of cognitive function and a philosophical conjecture of mind-world relation, which depicts the brain as a probabilistic prediction machine continuously organizing and maintaining an internal generative model of the outside world (Clark, 2013). The notions of predictive processes are ascribed to this theory through the inherent imperative to compensate for the skull-bound brain being unable to interpret its exterior environment, other than through various means of sensory input (Paton et al., 2013). To minimize the amount of information that needs explicit processing, the internal model is updated on the principle of predictive error minimization (PEM). Through perceivable regularities in its data, the brain is able to implicitly predict upcoming sensory input and maintain the model through that lens, all while updating the model on the basis of prediction errors, which occur when there is a dissonance between predicted and actual sensory input. This process of prediction error minimization is considered to be "precision-weighted", where prediction errors are hierarchically inferred on account of their precision or certainty: On a spectrum of evaluated precision of input, lower values are more likely to be regarded as statistical noise and thus ignored, while higher values denote importance and therefore demand a correction of the internal model (Arnaud, 2020; Clark, 2013).

Now entering an almost decade-long association with autism spectrum conditions, the predictive-coding theory has acted as an umbrella body for a cluster of symptoms. Certain perceptual and sensorimotoric abnormalities of autism, together with their unique perceptual experience ostensibly described as ToM deficits or WCC, are addressed through the PEM-centric implication that the autistic brain, in its predictive framework, carries an inherent inability to differentiate between high and low values of precision (Pellicano & Burr, 2012; Van de Cruys et al., 2014). This aberrant precision account of autism implies predictive processes of autistic minds essentially functioning without a "noise-filter" (Finnemann, 2019).

These combined theories of autistic perception possess the explanatory power to address several key details of autism definition: their hyper- or hypo-sensitivity to visual and auditory stimuli, the tendency (or preference) for locally-oriented perception and detailed local processing, the lack of coherence in certain contexts of speech and perception, the inflexible behaviour in socio-emotional environments, and the tendency towards developing special interests (Asperger, 1944; Baron-Cohen & Bolton, 1993; Kanner, 1943; Bogdashina, 2004; Van de Cruys, 2014). These notions of a significant cognitive difference have inspired the integration of individual and collective levels of analysis, through the recognition of psychopathology encompassing not only the disordered function of individuals, but the interpersonal dynamics of autistic-neurotypical dyads and their respective styles of interaction (Bolis et al., 2017). The Dialectical Misattunement Hypothesis (DMH) posits that larger differences in individual predictive processes and interaction styles cause communication misalignments and weak interpersonal coupling in social interactions. This interpretation of a dialectical misattunement between interlocutors, coupled with emergent notions of cross-neurotype communication and neurodivergent well-being in neurodiversity studies, has pioneered intersubjectivity as a crucial component for substantiating accountability of neurotypical populations (Rosqvist et al., 2020; Bolis et al., 2017).

The Neurodiversity Paradigm

The neurodiversity paradigm is an inherently reformative perspective of existing conditions of human cognitive normativity, which takes the notion of diversity being an undeniable fact of nature and describes the continuum of human neurocognitive variations as being part of a naturally-induced diversity of the human brain (Armstrong, 2010). Emerging through social movements of autism awareness and autistic self-advocacy, the introduction of the term neurodiversity and its terminological underpinnings in the 1990s sought to de-pathologize and re-theorize the medical diagnostic models of neurodevelopmental disorders: most notably amongst them, autism spectrum conditions and the realm of comorbid conditions associated with it.

As the new decade of the 21st century sets in, and the paradigm has shifted in accordance with an increasingly inclusive field of autism research, further developments have accompanied the aforementioned cognitive theories on the two-way nature of interaction, prompting the emergence of a new field of inquiry: neurodiversity studies. In their introduction to their book of the same name, Rosqvist et al. (2020) describe the domain as a “new theorization of conditions that are understood as impacting on the individual’s sense of identity” (p. 2) that has “aims to problematize neurotypical domination of the institutions and practices of academic knowledge, by questioning the boundaries between the predominant neurotypes and their others” (p. 2).

The subdivisions of neurotypes, and what constitutes a neurodivergence and neurotypicality, has yet to be definitively regulated in its terminology. In the case of this paper, the terminology will be a conjecture of the subject matter of Rosqvist et al. (2020), specifically derived from Chapter 6: Neurodiversity and cross-cultural communication by Hillary (2020), where neurotype is described as a “cluster of similar neurological and cognitive ways of being” (Hillary, 2020, p. 92), which also recognizes that there can be overlaps between people of different neurotypes. Neurodiversity encompasses all neurotypes, while neurodivergence is the term used for notable divisions from the predominant “typical” neurotype.

It is under these conditions that the neurodiversity paradigm has inclined towards the disposition of regarding neurodivergent individuals as being minorities in a neuronormative [2] organization of society which favours the majority neurotype, designated as the neurotypical population (Armstrong, 2010; Rosqvist et al., 2020). On a mental health account, its reinterpretation of several disorders of the mind being “components on a broader continuum of sensory, affectual, and cognitive processing” (Crompton et al., 2020a, p. 1446) has provided new possible frameworks for inquiry, which help establish

---

[2] **Neuronormative :** of, relating to, or based on the attitude that neurotypicality is the only normal and natural mode of brain functioning (based on heteronormative).

neurotype inequality and minority stressors as main factors of neurodivergent difficulty and ill-being (Rosqvist et al., 2020).

A Social-Emotional Salience Account

Although recent findings of contemporary autism research and neurodiversity studies have taken to depositing unprecedented value towards the recognition of a misalignment in the two-way nature of social interaction, there is a need to focus on what exactly is divergent in the neurodivergent condition of autism in order to properly navigate the concept of the semiosphere.

Several accounts of autistic behaviour imply not so much a mind-blindness or a lack of sociality, but rather a need to explicitly and consciously process social and emotional stimuli in the ocean of statistical noise that the typical social environment is imbued with (Arnaud, 2020).

[Routine formulae] are expressions whose occurrence is tied to particular, highly predictable situations, whose meaning is pragmatically conditioned and whose usage is motivated by the relevant characteristics of social situations. (House, 1996, p. 225)

The specialized niche of the neurotypical social environment, expressed by routine formulae, highly rewards tuned down precision due to how a multiplicity of accidental properties are expected to be regarded as uninformative and of low precision, by virtue of meta-learned social rules that determine relevance and irrelevance. These neurotypical social interactions are most often reliant on exchanges of high-level predictions, reducing complexity through a salience for social cues, which dictates the neurotypical social environment towards an intuitive preference for social and emotional information in focused areas of attention: "By being salient, social and emotional stimuli will become more fluent and familiar for neurotypical people" (Arnaud, 2020, p. 12).

With the recognition of a highly mediated and specific complexity of neurotypical social environments, the social-emotional salience account of autism addresses the seemingly unsociable and dismissive behaviour of autistic individuals through the aberrant precision account; because the autistic mind processes virtually everything in subequal amounts by default, due to an aberrant encoding of precision, social-emotional salience is evidently hard to develop and is therefore instinctively absent in most cases. Thus, the autistic mind has to process the neurotypical social sphere of information as they would with any novel input (Arnaud, 2020). This consequently means that there is not so much an integration into the experience of social interaction as much as there is an alienation from it—without implicit processing of social cues, fluency and participation in a neurotypical social world is inhibited by the need to consciously detect and process relevant social information.

In the social environment of neurotypical majorities, the framework of autistic predictive processing is constantly presented a dilemma of either deviating from their neutral cognitive mode and hypothesizing what their fellow humans deem salient, or be socially alienated; they have to engage in a different mind-world relation, acquiring fluency at the cost of their prediction and interaction style (Livingston et al., 2019). The existence of autistic fatigue (or burnout), which is where autistic people experience chronic exhaustion, loss of skills, and reduced tolerance to stimulus, implies that constantly deviating from the thresholds of their interaction style, and circumventing the principles of minimalization in their prediction style through intellect, exhausts their neural energy supply (Raymaker et al., 2020; Wang et al., 2017). Furthermore, adaptation via intellect to neurotypical spheres of communication can occur throughout different temporal scales and multiple exposures; e.g., repeated observations and experiences and consciously learned experience can be achieved through the span of a dialogue encounter or the lifespans of individual relationships, but these never constitute a complete fluency as it is effectively the development and maintenance of a "mask" (Bolis et al., 2017; Hull et al., 2017; Pearson & Rose, 2021).

An equally significant aspect of neurodivergence relates to the contemporary notion of a two-fold social disfluency, where non-autistic people have trouble understanding autistic people. This is the basis for the Double-Empathy Problem (DEP) which, in line with social-emotional salience, denotes a bidirectional misunderstanding of feelings and perspectives (DeThorne, 2020). This notion of mismatched salience in cross-neurotype social interaction has challenged the traditional ToM-deficit view of autism by conferring breakdown in pragmatic understanding as a symptom of cognitive and perceptual differences being at odds with each other, causing a communicative gap.

Although the predictive-coding account of autism, together with notions of neurodiversity, has helped to further induce a paradigm shift, the notion of autistic individuals relying on intellect to progress socially is not so much a novel concept as it has been one of the main identified consequences of the condition. It goes as far back as to the conception of the autistic condition by Hans Asperger:

> Normal children acquire the necessary social habits without being consciously aware of them, they learn instinctively. It is these instinctive relations that are disturbed in autistic children. To put it bluntly, these individuals are intelligent automata. Social adaptation has to proceed via the intellect. In fact, they have to learn everything via the intellect. (Asperger, 1944, p. 58)

Of course, certain aspects of this reading have been revised throughout the decades, such as with the notion of an intelligent automata: The autistic condition is no longer psychopathic in its definition, as signified by numerous accounts of there being a full range of emotions and the presence of cognitive, as well as affective, empathy—albeit with inhibited or different forms of expression (Berthoz & Hill, 2005; Brewer et al., 2016). Emotion recognition is present in autistic cognition, albeit functionalized by explicitly controlling for details rather than their predictive processes intuitively accounting for it (DeThorne, 2020), which further manifests itself as inflexibility in volatile socio-emotional environments, as their predictive style needs a higher level of certainty before proceeding to a stable processing stage (Latinus et al., 2019). This can be recognized as a paradigm of social interaction in and of itself, as it constitutes the autistic sphere of social-emotional recognition and processing. As evidenced by studies on autistic peer-to-peer information transfer and friendship quality, without the ramifications of neurotypical social dominance, autistic groups of people are able to essentially reach a level of social comfortability and identity not otherwise possible as when they are in neurotypical social spheres (Crompton et al., 2020a; 2020b).

Defining a Sphere: Neurodiversity, Cultures & Shared Meaning

It is through the introduction of separable niches of relative informational efficiency and social comfortability on the basis of their neurotype that the analysis of human social communication and experience is allowed a neurological dimension. While group membership and social identity is variable on the basis of many different factors, the key principle of similarity-based interpersonal attraction has been widely regarded as a pertinent factor in the formation and cohesiveness of a social relationship or group (Hogg et al., 1995; Triandis, 1960). It is within this similarity hypothesis that the predictive-coding perspective of autism can be utilized to regard dissimilarities in neurologically grounded mind-world relations as key catalysts for communicative gaps in social communication.

> To make meaning out of the world and successfully engage with the environment requires the ability to perceive sensory information from the world and integrate that information in meaningful ways into states of consciousness which then thrusts the individual into the world with appropriate actions.(Mueller & Tronick, 2020, p. 255)

The paradigm shift has solidified the notion of empathic difficulty being present on both sides due to a mismatch of mind-world relations, as well as different predictive styles garnering different interaction styles, certain existent meanings in a semiotic niche might not be salient in another interlocutor's world of meaning, per the social-emotional salience account. In relation to autism, these fundamental differences in sensory processing and meaning-making compound into atypical meanings and deviant experiences of events (Mueller & Tronick, 2020).

As such, with pragmatic language being reliant on an equivalent perception of context between interlocutors (Horton, 2012)—in such a way that their predictive processes account for the same details in a semiotic niche, the dearth of reliable empathy, and the jarring effects of incongruous interaction styles between autistic and non-autistic people, presents a higher likelihood of a breakdown in social communication (Williams et al., 2021). It is within this theme of dialectical misattunement on the basis of neurocognitive differences, that this main body of the paper intends to clarify the relevant communicative aspects of the neurodiversity paradigm, and further yet, contextualize the mind-world relational implications of divergent minds creating divergent environments through the comparison to the principles of culture and the semiosphere.

Cross-Neurotype Communication and its Implications

Diversity begets complexity, and in the case of human societies and its constituent phenomena, inter alia, communication, and meaning-making, the interacting adaptive entities (human beings) produce patterns and structures of behaviour, upon which principles of self-organization effectuates the development of social niches (Ashby, 1962; Graham, 2015; Page, 2010). It is in this ordering of types through their circumstances that it becomes relevant to compare the separate elements of aggregate ways of being; so far, this paper has only concerned itself with neurocognitive diversity in relation to the ways in which pragmatic information acquires its context in social communication.

It is through the neurodiversity paradigm that it is possible to infer that since neurotypes in their diversity can constitute wholly different spheres of experiences of the world and its meanings, it allows for a comparison to cultural diversity, upon which the ordering of effects in cross-neurotype communication would follow along the same principles as cross-cultural communication (Rosqvist et al., 2020). In both circumstances of communicating across significant differences, whether on the basis of neurocognitive or cultural diversity, pragmatic difficulties occur because of context-specific misalignments of mind-world relations, whether it is in their use of language, the values and morals they ascribe to themselves as well as the world, or the customs of life upon which their interaction styles are predicated (Hillary, 2020). What the subject matter of DEP and DMH suggests in this correlation is that while the neurodiversity paradigm has helped contextualize theories on cognition and social communi-

---

[3] In reference to a lesson on subtlety in cross-cultural communication, in which the examples listed by the teacher were presumed to only be applicable to cultural differences, but were equally applicable to examples of cross-neurotype communication, yet were not recognized as such because of their neuronormative orientation.

cation to core characteristics of autism spectrum conditions, the frames of understanding which acknowledge the existence and prevalence of not only functionally different minds, but also fundamentally distinct styles of social interaction, is a much more accepted aspect of cross-cultural communication versus cross-neurotype communication (Crompton et al., 2020c).

It was the same meaning – almost the same words. It was the same bluntness – even the same confusion. I claim a cultural difference: Autistic and Neuronormative, [which is] denied. They claim a cultural difference: American and Chinese, [which is] a known issue. (Hillary, 2020, p. 91)[3]

While there is a different conversation to be had on the status of neurobigotry, the more important implication of this observation is: The circumstances of interactions in which neurotype difference is the main precedent for pragmatic and empathic failure most often lack the basis for bridging communicative gaps through stories and context, in analogous reference to cross-cultural communication; and cultures have historical context, upon which the differences between interlocutors can be attributed to the external source of their cultural-societal roots (Hillary, 2020; Bolis et al., 2017). In the case of cross-neurotype communication, the communicative gap is most often owed to the assumed disability of the individual, such as an autistic condition, as an invasive insistence, as the non-autistic person would be used to most interactions going as according to neuronormative expectations (Milton, 2012).

The Notion of the Semiosphere

Boundaries seem to shape around neurotypes and cultures, wherein meanings, whether constituted by the formal systems of signs in their

---

[4] Or meaning-world, as per the Danish interpretation of the semiosphere : ”betydningsverden” (meaning + world).

language or by shared understanding of pragmatic meanings, seem to be apprehensible only to constituent members of the relative niche. Any attempt to communicate across mind-world relational differences is subject to potentially turbulent exchanges of half-meanings, which are only mitigated by similarities in other spheres of meaning; which is to say that interaction tuning is possible, only as long as there is some shared baseline that allows for translation or explanation of misunderstood facets (Bolis et al., 2017).

The great assumption of neurodiversity is, despite a neuroplasticity of the adaptive brain allowing adjustments to new environments through changes in its neurocognitive architecture, there are some forms of social communication which seem to depend on the foundational type and anatomy of the brain—neurodiverse conditions are more potentially vulnerable to this definition as socio-cognitive inflexibility interrelates with accounts of aberrant precision and fundamental differences in neurological and cognitive functioning (Chamberlain et al., 2020; Timberlake, 2019; Voss et al., 2017). It is within this assumption, in keeping with this trend of describing concepts that have yet to obtain an unambiguous definition, that the semiosphere is a potently relevant term, as it functions as a semiotic framework upon which conversations about the existence of borders and translation in complex interactions between different worlds of meaning[4] are made possible.

Juri Lotman's concept of the semiosphere was moderately framed in 1984 as both an abstract analogy and semiocentric continuation of Vladimir Vernadsky's biosphere, denoting it as the all-encompassing semiotic continuum comprised of "multi-variant semiotic models situated at a range of hierarchical levels" (Lotman,

---

[5] Multiple worlds and truths, in reference to how the brain only perceives whatever it deems salient, and that different senses of salience would produce different answers (see: gestalt, dialectics, socratic method, ideologies).

1984, p. 206). It had the underlying function of defining the duplex nature of the individual human organism as a constituent to not only its biological niche in the larger biosphere, but also its immersion into the self-constructed symbolic dimension of the human biocultural complex (Sinha, 2014). As much as the biosphere is defined as a biological membrane of the Earth, wherein the multiplicity of living beings all exist and are constituent members of various ecosystems and ecological niches, the primary semiosphere is both a semiotic space within which the totality of all semiotic acts exist, and simultaneously the fundament upon which all interconnected semiotic systems and products of semiosis are built upon (Nöth, 2006).

It has since then been developed and reinterpreted through several domains of research, most arguably gaining definition through its influential role in the development of cultural semiotics and otherwise through constituent niches of semiotic research: inter alia, theosemiotics, and biosemiotics (Kotov & Kull, 2011; Põder, 2021). The terms of "the semiosphere", "multiple semiospheres", and "semiotic space" have since then acquired a polysemous characteristic in terminology, wherein their meanings are dependent on the angle and field of research (Kull, 2005; Torop, 2005). In the case of this paper, we are dealing with pragmatic difficulties and capabilities in communication: As such, the primary semiosphere will be most relevantly defined as the world of shared meanings, a network of sign relations, and multiple worlds of meaning[5] (Kull, 2005; Nöth, 2006). Through this, languages are definable as systems of meaning, the corollary being that cultures are systems of shared understanding or historically transmitted patterns of meaning. A semiotic space would then denote the space of meaning delineated by a semiosphere, akin to the albumen and yolk of the egg, while a semiosphere emphasizes the existence of a shell, a boundary—which brings to mind the concept of semiospheres within semiospheres, borders within borders.

The Semiotic Space in Relation to Autism

An inherent feature of the semiosphere is that it is constituted by several smaller semiotic spaces; parallel to biodiversity in a biosphere, the diversity of semiotic systems effectuates niche construction in its complexity, of which interconnected participants of the semiosphere delineate themselves into peripheral semiotic niches of the primary semiosphere (Sinha, 2014). Termed secondary semiospheres, or simply "semiospheres", these patterns of meaning are supplementary superstructures upon the core of the primary semiosphere (this core being comprised of the most logical structure of meaning: natural language), where the systems of the core "permeate almost all semiospheric levels" throughout the superstructural niches of meaning-making (Nöth, 2006; Semenko, 2012, pp. 114-115). This is where the diversities of culture reside, and at the intersection of their peripheries, cross-cultural communication occurs: the connections through which pragmatic meanings become relevant, and in turn, fluency in different kinds of communication other than natural language. In comparison to hindrances in the conveyance of meaning in cross-linguistic interactions, where poor or non-existent comprehension has more of a basis in the componential values of the primary sign (the signifier and the signified) not yet gaining coherence in the mind of the individual (McGregor, 2015), the pragmatic difficulties of cross-cultural communication involve mismatches in the respective patterns of meaning, represented by how each individual acquits himself in the semiosphere.

As it stands, pragmatic failure in the realm of cross-neurotype communication seems to deal with a more primary level of meaning, one in which understanding is much more dependent on its neurological foundation, where a specific type of mind-world relation ensures fluency in a specific semiotic space. It is here that it should be important to address the smallest unit in the equation of the semiosphere: the concept of Umwelt coined by theoretical biologist Jakob von Uexküll (1864-1944) as the self-centred world of meanings and ostensibly described as the individual semiosphere (Semenko, 2012). Here, the Umwelt is essentially congenial to the internal model of the Bayesian brain, insofar as it "includes all the meaningful aspects of the world for a particular organism" (Kull, 1998, p. 304). In reference to aforementioned theories on cognition, the Umwelt as an individual semiosphere would be an interpretative model of the "true" world, restricted and filtered by their sensory systems, through their predictive-coding framework and parameters of precision. This consequently means that unknown semiotic values, by virtue of a different sense of salience and relevance, would not be perceived, unless through extrapolation of half-known patterns and systems of meaning (McGregor, 2015; Semenko, 2012); we are in multiple worlds of the semiosphere at all times and it is only through half-meanings, the so-called peripheries of the semiotic spaces we are immersed in, that we can engage with other semiospheres (Kull, 2005).

> A man receives only what he is ready to receive, whether physically or intellectually or morally, as animals conceive at certain seasons their kind only. We hear and apprehend only what we already half know. If there is something that does not concern me, which is out of my line, which by experience or by genius my attention is not drawn to, however novel and remarkable it may be, if it is spoken, we hear it not, if it is written, we read it not, or if we read it, it does not detain us. (Thoreau, 1961, pp. 212–213)

A neurotype is, by definition, a statistical fact of human neurological functioning in which the person is observationally grouped together with people of the same neurotype. There is no inherent communion

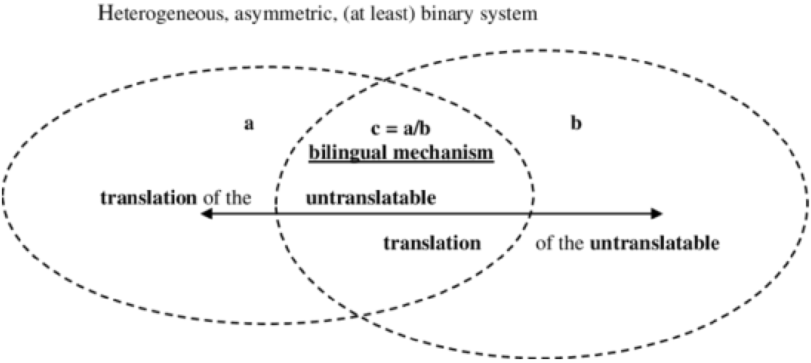


Figure 1: *Diagram of translation in cases of untranslatability (Monticelli, 2019, p. 396). The semiotic spaces of a intersects with that of b, where c represents the various peripheries of intersecting semiospheres between a and b, which in turn makes it possible to engage in the act of translation of the untranslatable.*

of minds in the diagnosis of a neurodivergent condition, wherefore a neurotype can be considered a hierarchical classification of similar Umwelten on the basis of residual signs that pertain to certain neurological and cognitive ways of being. Consequently, neurodivergent individuals seem to assimilate certain aspects of semiospheres differently from their neurotypical counterparts, with some more patently definable than others (such as with dyslexia) (Rosqvist et al., 2020), while the high-functioning region of the autistic spectrum of minds paradoxically has a more subtle difference in assimilation of core systems of the semiosphere (natural languages), yet a wholly different semiotic preference for systematic representations of meaning in a multiplicity of neurotypical-centric semiospheres (Tebartz van Elst et al., 2013). And if this interpretation is ordered on the principles of neurodiversity, where semiospheres are neurodiverse in their multiplicity, the definition would rather be one which signifies that the autistic mind has a semiotic preference for autistic systematic representations of meaning.

The intriguing character of high-functioning autism and neurodiversity as a whole on semiospheric terms is that a neurotype constitutes a mind-world relational preference for specific facets of semiotic spaces, inhabiting a specific sort of semiosphere in its engagement with the interconnected participants of the primary semiosphere. If most of the established patterns of meaning are neurotypically inclined by virtue of predominance, then autistic individuals are constantly engaging with the generic social world through "background acts of translation" and decidedly "more effortful searches for relevance" (Williams, 2020, p. 17). This is where the principal notion of the semiospheric interpretation reveals itself in how it defines cross-neurotype communication as heterogeneous, asymmetric interactions of mind-world relations.

The difference between autistic and non-autistic individuals is much less a dualism between two minds, but more so a "multiplicity of boundaries creating intersecting spaces" (Nöth, 2015, p. 20). The resulting description of a bilingual mechanism in autistic interactions with neurotypical spheres of meaning is more so a recognition of how "it is invariably autistic people who are the ones expected to function according to ways of organizing (and perceiving) concepts that do not necessarily come naturally to them" (Williams, 2020, p. 17). Despite a reciprocity in untranslatable spaces, the autistic person bears the brunt of neuronormative expectations, as the neurotypical will inevitably be accustomed to mutual experiences of semiotic spaces, consequently possessing less incentive to engage in a translational process of possible unknowns (Hillary, 2020; Davis & Crompton, 2021; Milton, 2012).

The Relevance of an Autistic Semiosphere

There is a constitutive explanatory value in establishing a neurological dimension to the semiosphere as well as the semiosphere reinforcing the contemporary idea of multiple interrelating mind-world

---

[6] See : The sorites paradox

relations in the neurodiversity of humankind. Although the potential curriculum of neurodiversity will inevitably be tied to the motifs of identity and de-pathologization, the semiospheric interpretation presents itself as a conceptual framework upon which dualistic notions of neurodiverse-neurotypical dichotomies can be contested in favour of the more fundamental concept of neurodiversity: that every human being is part of the diversity of neurological and cognitive functioning, and that the continuum of neurocognitive variations does not denote wholly different mind-world relations, but a multivariable continuance with other constituent members of the semiosphere through the intersecting spaces of our mind-world relations, with the ones denotable as neurodivergent essentially existing on the peripheries of normative systematic representations of meaning in the sharing of Umwelten. With this in mind, it is also possible to envision a future use case of the semiospheric interpretation beyond solely identifying pertinent differences between autistic and neurotypical mind-world relations, but also concretizing the terminology embedded in the neurodiversity concept by exploring where exactly the boundaries lie between what phenomenologically constitutes these categories of mind-world relations – recognizing the paradox of vagueness which exists in a phenomenological approach to pragmatic difficulties in ASC and how neurotypes would be categorized against each other.[6]

This semiospheric interpretation also challenges the notion behind humans sharing a foundation of basic psychological processes, and that what is consequently perceived as a "process commonality" (Berry, 2004) might be more nuanced in its diversity than what has been previously assumed; neurodiversity posits the idea that behavioural variability is not just a result of cultural shaping, but also variable foundations of neurological and cognitive styles of meaning-making and sensory processing.

Consequently, it also reinforces the idea that at this point in time, "we know more about autism than we've ever known, [yet] what we know is very little, and what we know is decidedly non-autistic" (Yergeau, 2018, p. 11). This is why inclusive research, wherein a research group consists of a decent amount of autistic researchers, has been so vital in properly defining the facets of cross-neurotype communication. The neurodiversity paradigm has allowed the centring of the neurodiverse sensory experience, thereby recognizing translation as a core aspect of autistic existence and allowing for more autism research from a predominantly autistic point of view, instead of relying on predominantly neurotypical descriptions of untranslatable semiotic values. Through the subject matter of neurodiversity, of DEP and DMH, and its referential interpretation through the semiosphere, the argument that translation should be more relevant, if not equally so, for the neurotypical population has been reinforced.

Nevertheless, the potential intrinsic harmfulness in an ill-considered application of neurodiversity should also be recognized when attempting to clarify the characteristics of abnormality and dysfunction. The idea of neurodiversity should not romanticize the whole field of neurominorities and mental illnesses; precautionary measures should be considered in the attempt to clarify and de-pathologize a spectrum of developmental conditions because a considerable amount of autistic people will be disabled, with the low-functioning parts of the spectrum falling under the flag of disability (Hughes 2021) as the symptoms of autism can become so severe that the person in question is unable to live without assistance (American Psychiatric Association, 2013; Armstrong, 2010; Rosqvist et al., 2020). Further yet, while neurodiversity studies and constituent efforts towards instituting the findings of DEP and DMH into diagnostic and treatment procedures of autism, the rest of the world (namely workplaces and school settings) have yet to formally adjust to the paradigm shift. If society in its neuronormative state does not alleviate possible difficulties of being of a different mind-world relation, in the levels as described in this paper, then there is still a concept of social disability to consider (Ruesch & Brodsky, 1968; Shakespeare, 2017). As such, the importance of sensitive periods and early identification of autism should not be ignored, even under the idea of a neurodiversity of humankind (Mueller & Tronick, 2020).

Conclusion

In this paper, the purpose of aligning neurodiversity and its constituent phenomenon of cross-neurotype communication with the fundamental concept of the semiosphere was to properly recognize the ties it had to cross-cultural communication, and in turn clarify the disposition of ASC, together with relevant cognitive theories of mind-world relations, in the larger narrative of a semiotic continuum. In the frame of neurodiversity, previous theories of autistic cognition are inferable as fragmented accounts of a greater whole, presenting components of these theories in a more complete manner through contemporary hypotheses of dialectical misattunement and a double-empathy problem between interlocutors. This is all made possible by the foundation of the predictive-coding hypothesis, which accentuates the autistic brain as having a different sense of salience in the semiotic spaces of the symbolic biocultural complex by virtue of a different sensory system. Through this, the semiospheric interpretation suggests that the autistic mind immerses itself in a generically neurotypical semiotic continuum which, to a large extent, is not suitable to the different systematization of meaning inherent in its predictive style. On the basis of the similarity hypothesis, this has repercussions for the membership into social niches, as pragmatic elements of a social interaction rely on the continuity of equivalent perception and mutual understanding of context. The discontinuity between autistic and non-autistic predictive and interaction styles creates a self-sustaining communicative gap, which is further worsened by a generally neuronormative orientation of society encouraged by previous deficit-based accounts of autism research. The autistic individual has to adapt, translate, and search for relevance in a mind-world relation that does not come naturally to them. All this culminates into the recognition of neurodiversity as a still-developing perspective and paradigm that has yet to substantially change the current situation of neuronormativity. But nevertheless, a paradigm in which the concept of the semiosphere can help delineate the borders between disability and interpersonal mismatches of asymmetrically equal minds. It can contribute to the expanding literature of neurodiversity in tandem with autism, such that even at this conceptual stage, it can be an informational tool for substantiating intersubjectivity and accountability of neurotypical populations in the bridging of communicative gaps.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, D.C.: American Psychiatric Publishing.

Armstrong, Thomas. (2010). Neurodiversity: discovering the extraordinary gifts of autism, ADHD, dyslexia, and other brain differences. *Cambridge: Da Capo Lifelong.*

Arnaud, Sarah. (2020). A social–emotional salience account of emotion recognition in autism: Moving beyond theory of mind. *Journal of Theoretical and Philosophical Psychology, 42(1),* 3-18.

Ashby, W. R. (1962). Principles of the self-organizing system. Principles of Self Organization: Transactions of the University of Illinois Symposium (H. Von Foerster & G. W. Zopf, Eds.). *London: Pergamon Press,* 255-278.

Asperger, Hans. (1944). "Autistiche Psychopathen" im Kindesalter. Translated and edited by Uta Frith. (1991). *Autism and Asperger syndrome*, 37-92. Cambridge: Cambridge University Press.

Baron-Cohen, S. (1997). Mindblindness: An essay on autism and theory of mind. Cambridge, MA: MIT Press.

Baron-Cohen, Simon & Bolton, Simon. (1993). Autism: The facts. Oxford: Oxford University Press.

Berry J. W. (2004). Psychology of group relations: cultural and social dimensions. *Aviation, Space, and Environmental Medicine, 75 (7),* 52-57.

Berthoz, Sylvie & Hill, Elisabeth L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. *European psychiatry: the journal of the Association of European Psychiatrists, 20(3),* 291–298.

Bogdashina, Olga. (2004). Communication Issues in Autism and Asperger Syndrome: Do We Speak the Same Language?. London: Jessica Kingsley Publishers.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond Autism: Introducing the Dialectical Misattunement Hypothesis and a Bayesian Account of Intersubjectivity. *Psychopathology, 50 (6),* 355–372.

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., & Bird, G. (2016). Can Neurotypical Individuals Read Autistic Facial Expressions? Atypical Production of Emotional Facial Expressions in Autism Spectrum Disorders. *Autism research: official journal of the International Society for Autism Research, 9(2),* 262–271.

Chamberlain, S. R., Solly, J. E., Hook, R. W., Vaghi, M. M., & Robbins, T. W.(2021). Cognitive Inflexibility in OCD and Related Disorders. Current topics in behavioral neurosciences, 49, 125–145.

Clark, Andy. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences, 36(3),* 181-204.

Crompton, C. J., Hallett, S., Ropar, D., Flynn, E., & Fletcher-Watson, S. (2020a). 'I never realised everybody felt as happy as I do when I am around autistic people': A thematic analysis of autistic adults' relationships with autistic and neurotypical friends and family. *Autism: the international journal of research and practice. [Online] 24 (6),* 1438–1448.

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher Watson, S. (2020b). Autistic peer-to-peer information transfer is highly effective. *Autism: the international journal of research and practice. [Online] 24 (7),* 1704–1712.

Crompton, C. J., Sharp, M, Axbey, H., Fletcher-Watson, S., Flynn, E. G., & Ropar, D. (2020c). Neurotype-Matching, but Not Being Autistic, Influences Self and Observer Ratings of Interpersonal Rapport. *Frontiers in psychology. [Online] 11,* 586171.

Davis, Rachael & Crompton, Catherine. (2021). What Do New Findings About Social Interaction in Autistic Adults Mean for Neurodevelopmental Research? *Perspectives on psychological science. [Online] 16 (3),* 649–653.

DeThorne, Laura. (2020). Revealing the Double Empathy Problem: It's not that autistic* people lack empathy. *Rather, their different neurotypes and experiences may make it harder for nonautisic people to understand them—and vice versa. The ASHA Leader, 25,* 58-65.

Finnemann, Johanna J. S. (2019). Investigating Sensory Prediction in Autism Spectrum Conditions. PhD thesis. University of Cambridge: Cambridge, England. Retrieved from: *https://www.repository.cam.ac.uk/handle/1810/297678.*

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Crane, L., Cusack, J., Leekam, S., Milton, D., Parr, J. R., & Pellicano, E. (2019). Making the future together: Shaping autism research through meaningful participation. *Autism: the international journal of research and practice. [Online] 23 (4),* 943–953.

Fletcher-Watson, S., Brook, K., Hallett, S., Murray, F., & Crompton, C. (2021). Inclusive Practices for Neurodevelopmental Research. *Current developmental disorders reports. [Online] 8 (2),* 88–97.

Frith, Uta. (2008). Autism: A Very Short Introduction. Oxford: Oxford University Press.

Gernsbacher, Morton A. & Yergeau, Melanie. (2019). Empirical Failures of the Claim That Autistic People Lack a Theory of Mind. *Archives of Scientific Psychology, 7(1),* 102–118.

Green, A. R., Carrillo, J. E., & Betancourt, J. R. (2002). Why the disease-based model of medicine fails our patients. *The Western journal of medicine, 176(2),* 141–143.

Hillary, Alyssa. (2020). Neurodiversity and cross-cultural communication. *Neurodiversity Studies: A New Critical Paradigm (Hanna BertilsdotterRosqvist, Nick Chown & Anna Stenning, eds.).* Milton: Taylor and Francis, 91 107.

Hogg, M. A., Hardie, E. A., & Reynolds, K. J. (1995). Prototypical similarity, self categorization, and depersonalized attraction: A perspective on group cohesiveness. *European journal of social psychology. [Online] 25(2),* 159–177.

Horton, William S. (2012). Shared knowledge, mutual understanding and meaning negotiation. Cognitive Pragmatics (Hans-Jörg Schmid, ed.). Berlin, Boston: De Gruyter Mouton, 375-404.

House, Juliane. (1996). Developing Pragmatic Fluency in English as a Foreign Language: Routines and Metapragmatic Awareness. *Studies in Second Language Acquisition, 18(2),* 225–252.

Hughes, Jonathan A. (2021). Does the heterogeneity of autism undermine the neurodiversity paradigm?. *Bioethics, 35(1),* 47– 60.

Hull, L., Petrides, K.V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M., & Mandy, W. (2017). "Putting on My Best Normal": Social Camouflaging in Adults with Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders, 47,* 2519–2534.

Kanner, Leo. (1943). Autistic disturbances of affective contact. *Nervous Child, 2,* 217-250.

Kotov, Kaie & Kull, Kalevi. (2011). Semiosphere Is the Relational Biosphere. Towards a semiotic biology: life is the action of signs (Claus Emmeche & Kalevi Kull, Eds.). London: Imperial College Press, 179-194.

Kull, Kalevi. (1998). On semiosis, Umwelt, and semiosphere. Semiotica*, 120(3-4),* 299-310.

Kull, Kalevi. (2005). Semiosphere and a dual ecology: Paradoxes of communication. *Sign Systems studies, 33(1),* 175-189.

Latinus, M., Cléry, H., Andersson, F., Bonnet-Brilhault, F., Fonlupt, P., & Gomot, M. (2019). Inflexibility in Autism Spectrum Disorder: Need for certainty and atypical emotion processing share the blame. *Brain and Cognition, 136,* 103599.

Livingston, L. A., Colvert, E., , Bolton, P., & Happé, F. (2019). Good social skills despite poor theory of mind: exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 60(1),* 102-110.

López, Beatriz & Leekam, Susan R. (2003). Do children with autism fail to process information in context?. *Journal of child psychology and psychiatry, and allied disciplines, 44(2),* 285–300.

Lotman, Juri M. (1984). On the semiosphere. Translated and edited by Wilma Clark. (2005). *Sign Systems Studies, 33(1),* 205–229.

McGregor, William B. (2015). Linguistics: An Introduction (2nd Edition). London; New York: Bloomsbury Academic.

Milton, Damian E. M. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society, 27(6),* 883-887.

Monticelli, Daniele. (2019). Borders and translation: Revisiting Juri Lotman's semiosphere. *Semiotica, 2019(230),* 389-406.

Mueller, Isabelle & Tronick, Ed. (2020). Chapter 14 – Sensory processing and meaning-making in autism spectrum disorder. Autism 360 ° (Undurti Das, Neophytos Papaneophytou & Tatyana El-Kour, Eds.). London: Academic Press, 255-267.

Nicolaidis, C., Milton, D., Sasson, N. J., Sheppard, E., & Yergeau, M. (2019). *An Expert Discussion on Autism and Empathy. Autism in Adulthood, 1 (1),* 4-11.

Nöth, Winfried. (2006). Yuri Lotman on metaphors and culture as self-referential semiospheres. *Semiotica, 2006(161),* 249-263.

Page, Scott E. (2010). Diversity and Complexity. Princeton: Princeton University Press.

Paton, B., Skewes, J., Frith, C., & Hohwy, J. (2013). Skull-bound perception and precision optimization through culture. *The Behavioral and Brain Sciences, 36(3),* 222–222.

Pearson, Amy & Rose, Kieran. (2021). A Conceptual Analysis of Autistic Masking: Understanding the Narrative of Stigma and the Illusion of Choice. *Autism in Adulthood. [Online]. 3(1),* 52-60.

Pellicano, Elizabeth & Burr, David. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16(10),* 504-10.

Põder, Thomas-Andreas. (2021). Religion in the semiosphere: Theosemiotics in dialogue with Juri Lotman. Sign, Method and the Sacred: New Directions in Semiotic Methodologies for the Study of Religion (Jason Cronbach Van Boom & Thomas -Andreas Põder, Eds.). Berlin, Boston: De Gruyter, 29-52.

Rosqvist, H. B., Chown, N., & Stenning, A. (2020). Neurodiversity Studies: A New Critical Paradigm. Milton: Taylor and Francis.

Ruesch, Jurgen & Brodsky, Carroll M. (1968). *The Concept of Social Disability. Arch Gen Psychiatry, 19(4),* 394-403.

Sack, Graham. (2015). Culture as a complex system [Powerpoint presentation]. Glocomnet: Human Complexity and Uncertainty. Available at: https://d1rz56ot08aua0.cloudfront.net/attachments/library_files/128/original-.pdf?1443800048 (Accessed: 7 February 2022).

Schweitzer, Frank & Zimmermann, Joerg. (2001). Communication and Self Organisation. Complex Systems: A Basic Approach (M. M. Fischer & J. Fröhlich, Eds.). Berling: Springer, 275-296.

Semenko, Aleksei. (2012). The Texture of Culture: An Introduction to Yuri Lotman's Semiotic Theory. New York: Palgrave Macmillan.

Shakespeare, Tom. (2017). The Social Model of Disability. The Disability Studies Reader (5th Edition). (Lennard J. Davis, Ed.). New York: Routledge, 195-218.

Silberman, Steve. (2015). NeuroTribes: The Legacy of Autism and the Future of Neurodiversity. New York: Avery Publishing.

Sinha, Chris. (2014). Niche construction and semiosis: biocultural and social dynamics. The Social Origins of Language (Daniel Dor, Chris Knight & Jerome Lewis, Eds.). Oxford: Oxford University Press, 31-46.

Sinha, Chris. (2015). Language and other artifacts: socio-cultural dynamics of niche construction. Frontiers in Psychology, 6, 1601.

Tebartz van Elst, L., Pick, M., Biscaldi-Schäfer, M., Fangmeier, T., & Riedel, A. (2013). High-functioning autism spectrum disorder as a basic disorder in adult psychiatry and psychotherapy: Psychopathological presentation, clinical relevance and therapeutic concepts. *European Archives of Psychiatry and Clinical Neuroscience, 263(2),* 189-196.

Thoreau, Henry David. (1961). The Heart of Thoreau's Journals. (Odell Shepard, Ed.). New York: Dover Publications.

Timberlake, Howard. (2019). Why there is no such thing as a 'normal' brain. BBC Future. [Online]. October 10.

Torop, Peeter. (2005). Semiosphere and/as the research object of semiotics of culture. *Sign Systems Studies. [Online]. 33(1),* 159–173.

Triandis, Harry C. (1960). Cognitive Similarity and Communication in a Dyad. *Human relations. [Online]. 13(2),* 175–183.

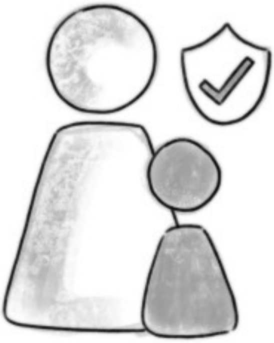
Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121,* 649-675.

Voss, P., Thomas, M. E., Cisneros-Franco, J. M., & de Villers-Sidani, É. (2017). Dynamic Brains and the Changing Rules of Neuroplasticity: Implications for Learning and Recovery. *Frontiers in Psychology, 8,* 1657.

Williams, G. L., Wharton, T., & Jagoe, C. (2021). Mutual (Mis)understanding: Reframing Autistic Pragmatic "Impairments" Using Relevance Theory. *Frontiers in Psychology, 12,* 616664.

Williams, Gemma L. (2020). Talking together at the edge of meaning: Mutual (mis)understanding between autistic and non-autistic speakers. PhD thesis. University of Brighton: Cambridge, England. Retrieved from: https://cris.brighton.ac.uk/ws/portalfiles/portal/30794840/Gemma_L_Williams-_PhD_Thesis_October_2020_WITH_APPENDICES.pdf.

World Health Organization. (2019). *International Statistical Classification of Diseases and Related Health Problems, 11th Revision. Geneva: World Health Organization.*



Effects of Parental Control on Adolescent Body Dissatisfaction

Liu Mei'En

National University of Singapore

Department of Psychology

Acelyn Aw Li Ping

National University of Singapore

Department of Psychology

Cheng Chi Lam Stephanie

National University of Singapore

Department of Psychology

Chermaine Tay Yue Ling

National University of Singapore

Department of Psychology

Faith Wong Yann Hwa

National University of Singapore

Department of Psychology

Abstract

Body dissatisfaction is most prevalent in adolescence when boys and girls experience physical changes in their bodies. The aim of the current study was to test the predictive effect of parental pressure and overt control on body dissatisfaction among Singaporean adolescents (N = 155). Self-reported measures designed to assess adolescents' body dissatisfaction, body-related parental pressure, and parental overt control on food intake were administered to adolescents aged 13–20 years (M = 17.22, SD = 1.85). The data were analyzed with a series of hierarchical regression and Pearson correlation tests. Path analyses were performed separately for males (N = 50) and females (N = 100). The results showed that only parental overt control positively predicted body dissatisfaction in adolescent girls. Neither body-related parental pressure nor overt control predicted body dissatisfaction in adolescent boys. Implications of the findings were broadly discussed.

***Keywords:* Adolescence, body dissatisfaction, parental pressure, overt control**

Introduction

Body dissatisfaction can be defined as having negative thoughts and feelings toward one's body (Myers & Crowther, 2009). Body dissatisfaction tends to be strongest during adolescence and most salient during the onset of puberty (Santrock, 2019). Similar to Western studies, Singaporean studies have found that adolescent girls were less happy with their bodies than adolescent boys (Lwin & Malik, 2012; Chng & Fassnacht, 2016). This phenomenon could be explained by the physical changes that occur during adolescence and media portrayals of ideal body shapes (Griffiths et al., 2017). During puberty, girls tend to gain more body fat, which draws them further from the thin ideal body shape portrayed in the media. In contrast, boys tend to gain more muscle, which draws them closer to the muscular ideal body shape portrayed in the media.

Studies have mostly looked at the effects of parental influence on body dissatisfaction in terms of parenting styles. Authoritarian parenting, characterized by high parental control and low warmth, was associated with higher risks of disordered eating behaviours in adolescents (Jáuregui Lobera et al., 2011). Research revealed that children with authoritarian parents experienced greater significant amounts of body dissatisfaction and had a higher likelihood of engaging in extreme weight-control behaviours (Enten & Golan, 2009; Zubatsky et al., 2014). Moreover, highly controlling authoritarian parents tended to be psychologically controlling, which can lead to disordered eating (Robertson, 2020; as cited in Reilly et al., 2016). Given that most Singaporean mothers employ authoritarian parenting practices (Cho et al., 2020), which are linked with body dissatisfaction, the present study explored the effects of parental control on body dissatisfaction in Singaporean adolescents.

Parental Control

In parenting literature, *parental control* refers to parents' involvement in making decisions for children (Grolnick & Pomerantz, 2009). There are two dimensions of parental control: *psychological control,* which regulates children's feelings and thoughts, and *behavioural control,* which regulates children's actions and behaviours (Barber, 1996; Barber et al., 2005). Studies have consistently shown that psychological control can lead to internalizing depressive symptoms (Barber, 1996) and negatively impacting the emotional well-being of children (Gray & Steinberg, 1999; Wang, et al., 2007). Based on self-determination theory, satisfying the universal need for autonomy is indispensable for psychological functioning (Deci & Ryan, 1985). Psychological control, which thwarts the need for autonomy in adolescents, has thus been found to be associated with adverse developmental outcomes, such as diminished emotional functioning, depression, and delinquency (Barber et al., 2005; Pomerantz & Wang, 2009). In contrast, behavioural control has been linked to more positive outcomes. For instance, it was found that behavioural control can reduce adolescents' exposure to maladaptive behaviours while increasing adaptive outcomes such as improved academic competence and lower delinquency (Fletcher et al., 2004; Gray & Steinberg, 1999; Grolnick & Pomerantz, 2009; Wang et al., 2007). The aspects of psychological and behavioural control are summarized in Table 1.

Psychological Control: Body-Related Parental Pressure

One aspect of psychological control is body-related parental pressure, which is conceptualized as the perceived demand to modify one's body shape or weight through parents' negative verbal commentary. Existing literature highlights three forms of body-related parental pressure that can affect body dissatisfaction levels: parental teasing, parental encouragement, and parental criticism (Helfert & Warschburger, 2011; Kluck, 2009).

Parental teasing refers to negative verbal commentary about the adolescent's diet, weight, or body shape, delivered in a playful or light -hearted manner (Helfert & Warschburger, 2011). A systematic meta-analysis has found parental teasing to be positively associated with low self-esteem and body dissatisfaction (Menzel et al., 2010). *Parental encouragement* refers to negative verbal commentary about the adolescent's diet, weight, or body shape that is intended to be encouraging (Helfert & Warschburger, 2011). In a study, parental encouragement to control weight and size was found to be the most robust predictor of body dissatisfaction (Kluck, 2010). *Parental criticism* refers to negative verbal commentary about the adolescent's diet, weight, or body shape delivered in a hostile or harmful manner (Rodgers & Chabrol, 2009). Studies have revealed that female adolescents' self-reported frequency of parental criticism is a strong predictor of body dissatisfaction and disordered eating outcomes (Rodgers & Chabrol, 2009). However, much less is known about the relationship between parental criticism and body dissatisfaction in adolescent boys.

By and large, parental pressure regarding adolescents' body shape and weight has been linked to greater body dissatisfaction (Chng & Fassnacht, 2016). While the underlying mechanism behind this link is not completely understood, research revealed that this link was partially mediated by the internalization of media ideals and appearance comparison (Rodgers, Paxton & Chabrol, 2009). Additionally, a study led by Lwin and Malik (2012) in Singapore indicates a gender difference in the impact of parental pressure, where the negative verbal commentary was more strongly linked with body dissatisfaction in boys than girls in early adolescence. Though hypothetical, the researchers believed that this finding can be attributed to the weight status in their sample, wherein there is a higher percentage of boys (39.5%) who are obese than girls (24.1%). As such, boys might have received greater parental criticism, leading to increased body dissatisfaction amongst boys.

Behavioural Control: Overt Control

Overt control can be defined as children's perception of their parent's actions to regulate their eating practices (Ogden et al., 2006). There are two facets of overt control: parental restriction and parental monitoring. *Parental restriction* comprises the ways in which children perceive that their parent(s) limit their access to certain foods; *parental monitoring* involves the degree to which children perceive that their parent(s) track, check on, and oversee their eating (Birch et al., 2001). Studies have shown that parental restriction of palatable foods was associated with children's poor self-regulation of food consumption and unhealthy eating patterns (Carper et al., 2000; Fisher & Birch, 1999). Research has also found parental monitoring of food intake to be associated with greater overeating behaviour (Kenyon et al., 2009). Taken together, these findings indicate that parental overt control may lead to poorer self-control, which has been linked with higher saturated fat intake and less strenuous exercise in adolescents (Wills et al., 2007). As a result of unhealthy eating and lifestyle, parental control can lead to overeating and obesity (Costanzo & Woody, 1985; Johnson & Birch, 1994). Given that studies have consistently found body mass index (BMI) to be strongly linked with body dissatisfaction in adolescents (Barker & Galambos, 2003; Paxton et al., 2006; Taylor et al., 2012), the negative impact of high parental control on body dissatisfaction becomes evident.

Table 1: *Summary of Psychological Control and Behavioural Control Constructs*

Construct	Definition	Example
Psychological Control		
Parental Teasing	Negative verbal commentary; delivered in a playful or light-hearted manner	"You should watch your weight; you look like a watermelon!" "Please eat more, I don't want you to get blown away by the wind."
Parental Encouragement	Negative verbal commentary; delivered in a motivating manner	"You should eat more; you look like you have lost weight recently." "Exercise will make your body look better; you should do it more often."
Parental Criticism	Negative verbal commentary; delivered in a hostile or harmful manner	"You're a nightmare! Why are you still eating so much when you have put on so much weight lately?" "You are too skinny for your age. You should exercise more and build up some muscles!"
Behavioural Control		
Parental Restriction	Perception that parent(s) limit their access to food	The child notices that their parents do not allow them to eat food high in saturated fat (e.g., fried food).
Parental Monitoring	Perception that parent(s) track, check on, and oversee their consumption	The child notices that their parent does not allow them to eat ice cream more than once per week.

The Present Study

Existing literature suggests that both the psychological aspect (i.e., body-related parental pressure) and behavioural aspect (i.e., overt control) of parental control can have undesirable outcomes on body dissatisfaction. In the context of this paper, body dissatisfaction is defined as the negative perception of one's body, weight, and appearance. While some studies suggest that overt control can lead to slightly more positive outcomes in other areas of development (Pomerantz & Wang, 2009; Grolnick & Pomerantz, 2009), there is a paucity of research on the effects of parental overt control on body dissatisfaction in adolescents. Therefore, the present study aims to fill this lacuna by investigating whether psychological and behavioural control are significant predictors of body dissatisfaction in thirteen- to twenty-year-old female and male Singaporeans. The following were hypothesized:

1. Body-related parental pressure would positively and significantly predict adolescents' body dissatisfaction above and beyond overt control;
2. Overt control would positively and significantly predict adolescents' body dissatisfaction above and beyond body-related parental pressure.

Methods

Participants

A total of 182 participants were recruited via online invitations sent through social media platforms like Whatsapp and Telegram. Convenience sampling was employed in this study due to the short data collection period lasting about 2 weeks. Participants were required to meet the following criteria: 1) must be Singaporean, 2) must currently live with at least one parent, 3) must be aged between 13 to 20 years old. 27 responses that did not meet these criteria were excluded from data analysis. The final sample consisted of 155 participants (Age: M = 17.22, SD = 1.85; BMI: M = 21.12, SD = 3.71) with 100 females, 50 males, and 5 who did not specify their gender. The ethnic composition of the sample was 91.6% Chinese, 3.2% Indian, 2.6% Malay, and 2.5% who selected "other" or did not specify.

Procedure

The study was conducted as part of the Adolescent Psychology laboratory module at the National University of Singapore, where the researchers were enrolled in. Online invitations were sent to the social media contacts of the researchers. After responding to the online invitation participants would read through a brief of the study (Appendix A) and informed consent (Appendix B). Consenting participants would then complete an e-survey containing the set of questionnaires (Appendix D). Upon completing the study, participants would read a final debrief (Appendix C).

Materials

Body-Related Parental Pressure

Body-Related Parental Pressure was operationalized as negative verbal comments about the adolescent's body weight or shape. This was measured with the 18-item Body-Related Parental Pressure Scale (BPPS), which was an original scale developed for the purpose of this study due to a lack of existing scales that met the study's criteria. The scale was comprised of three subscales, namely parental teasing, parental encouragement, and parental criticism. Participants responded on a 5-point Likert scale, ranging from 1 (*Never*) to 5 (*Every time*). The Cronbach alpha of the scale (α = .87) indicated a high internal consistency in the present sample. A composite score for each participant was derived by summing the mean scores of each subscale, with a higher composite score indicating greater body-related parental pressure.

Parental Teasing. The Parental Teasing subscale was used to measure the frequency of negative jokes about the adolescent's diet, weight, or body shape. This subscale consists of 6 items adapted from the Weight-Related Teasing subscale (Thompson et al., 1995). Similar to a previous study (Pötzsch et al.., 2018), the word "people" was replaced with "parents" in all items of the Weight-Related Teasing Subscale to specifically measure perceived parental pressure via teasing. The items have also been altered to evaluate parental teasing on body -image-related concerns in general, and not just among adolescents who are considered heavy or overweight. One such modification was changing "People made fun of you because you were heavy" to "My parent(s) make fun of me because of my body shape or weight."

Parental Encouragement. The Parental Encouragement subscale was used to measure how often the adolescent receives negative messages about one's body shape or weight from one's parents which are intended to be encouraging (Helfert & Warschburger, 2011). Due to a lack of access to relevant scales in existing literature, this subscale was self-constructed. An example item is "My parent(s) praise me when I eat more/less."

Parental Criticism. The Parental Criticism subscale was used to measure how often an adolescent receives negative verbal messages about one's body shape or weight delivered by one's parents in a harmful way (Rodgers & Chabrol, 2009). Six items measuring the frequency of negative parental messages about eating and weight were adapted from the Parental Eating and Weight Messages Survey (Rivero, 2021). An example item is "My parent(s) reprimand me for being too skinny/fat."

Overt Control. The Overt Control Scale (OCS) was adapted from the Child Feeding Questionnaire (CFQ) (Birch et al., 2001) to measure adolescent-reported, instead of parent-reported, imposed restrictions and monitoring on food intake by parents. The OCS of the present study borrowed 11 items from two subscales of CFQ: Parental Restriction and Parental Monitoring. Prior research which adapted CFQ to assess adolescents' reports achieved a high internal consistency, α = .82 (Farrow, 2012). The OCS in the present sample achieved high internal reliability (α = .90). A composite overt control score for each participant was derived by summing the mean scores of the subscales, with a larger score indicating greater overt control.

Parental Restriction assessed the extent to which adolescents perceive their parents limiting their access to food. It was comprised of 8 items which participants rated on a 5-point Likert scale, ranging from 1 (Strongly disagree) to 5 (Strongly agree). An example item is "My parent(s) intentionally keep some foods out of my reach." The internal consistency of the subscale was high in the sample (α = .88).

Parental Monitoring assessed the extent to which adolescents perceive their parents overseeing their eating. It was comprised of 8 items that participants rated on a 5-point frequency scale, ranging from 1 (Never) to 5 (Every time). An example item is, "How often does your parent(s) keep track of the high-fat foods that you eat?" The internal consistency of the subscale was satisfactory (α = .75).

Body Dissatisfaction

Body Dissatisfaction was measured with a modified version of the 23 -item Body Esteem Scale for Adolescents and Adults (BESAA), which involved positive and negative judgements of one's body weight and appearance. Prior research found that Cronbach's alphas of all three subscales were 0.90 and above (Cragun et al., 2013). However, 5 items on external attributions were excluded as they were irrelevant to the study. The original BESAA was a 5-point Likert scale, ranging from 0 (*Never*) to 4 (*Always*). To standardize the values across all the scales, the BESAA scale was modified such that the lowest score is 1 (*Never*) and the highest score is 5 (*Always*). Negatively worded items were reverse coded, with a higher score indicating a more negative body evaluation. The internal consistency of the subscale was high in the sample ($\alpha = .94$).

The BE-Weight subscale evaluated one's thoughts about one's own weight. An example item of the 6-item subscale is "I am satisfied with my weight." The internal consistency of the subscale was high ($\alpha = .91$).

The BE-Appearance subscale measured feelings about one's general appearance. An example item of the 12-item subscale is "I wish I looked better." The internal consistency of the subscale was high ($\alpha = .93$).

Data Analysis

All statistical analyses were performed on SPSS 27.0, except Confirmatory Factor Analysis (CFA), which was performed on JASP 14.0. In the preliminary analyses, CFA was conducted to examine the factorial validity of the Body-Related Parental Pressure original scale (BPPS). Goodness-of-fit was evaluated using the $\chi 2$ test of model fit, the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). Following Hu and Bentler (1999), CFI greater than 0.90 indicated a good fit to the data, and a cut-off value close to .06 for RMSEA is deemed to be a relatively good fit.

Tests for internal consistency reliability were run on SPSS to determine the reliability of the scales used in this study. The Cronbach's alphas obtained were used to determine internal consistency reliability (poor: $\alpha \leq .50$, acceptable: $\alpha \geq .70$, excellent: α

≥ .90). The two independent variables in this study, Body-Related Parental Pressure and Overt Control, are continuous and unnested within each other. Hence, a series of hierarchical multiple regression tests were conducted to determine the unique contribution of body-related parental pressure and overt control to body dissatisfaction among female and male adolescents. The first hierarchical multiple regression analysis was performed by including Overt Control in the first block and Body-Related Parental Pressure in the second. This tested if Body-Related Parental Pressure could predict Body Dissatisfaction after controlling for Overt Control, thus answering the first hypothesis. The second hierarchical multiple regression analysis was performed by including Body-Related Parental Pressure in the first block and Overt Control in the second. This would test if Overt Control could predict Body Dissatisfaction after accounting for Body-Related Parental Pressure, thus answering the second hypothesis.

Additionally, a series of independent samples t-tests were performed to determine if there were significant gender and age group (i.e., secondary versus post-secondary) differences in body dissatisfaction, body-related parental pressure, and overt control. Bonferroni corrections were not performed due to the limitations of the analysis software used. However, this would have little impact on the outcome of the study as the t-tests are additional analyses that do not directly answer the research question. If differences were to be found, follow-up analyses would include performing the hierarchical multiple regression and correlational analyses separately for male and female responses using SPSS.

Results

Preliminary Analyses

The Confirmatory Factorial Analysis (CFA) revealed that the BPPS model was significant, but it did not fit the data well (CFI = 0.775; RMSEA = 0.144). In other words, the model was not consistent with the data. After identifying and removing 4 misfit items, the adjusted model was found to be a good fit (Appendix F). Hence, all subsequent analyses were run without the misfit items. The internal consistency for the overall BPPS scale ($\alpha = 0.87$), Encouragement subscale ($\alpha = 0.90$), Criticism subscale ($\alpha = 0.90$), and Teasing subscale ($\alpha = 0.87$) remained high without the misfit items. Descriptive statistics of study variables are presented in Table 2.

Table 2: *Mean and standard deviation for parental pressure, overt control, body dissatisfaction*

Variable	Minimum Possible	Maximum Possible	Minimum Obtained	Maximum Obtained	Mean (SD)
Body-related parental	3	15	3.00	13.87	5.86 (2.27)
Overt control	2	10	2.00	10.00	4.02 (1.75)
Body	2	10	2.00	9.83	5.85 (1.73)

Main Analyses

The first hierarchical multiple regression was conducted with Overt Control entered at Stage 1 and Parental Pressure entered at Stage 2 (Table 2). The results showed that adding Parental Pressure at Stage 2 explained an additional 1.7% of the variance. However, this change was not significant, $F(1, 152) = 2.76$, $p > .05$. This showed that Paren-

***Table 3:** Hierarchical Regression Model examining the relationship between Parental Pressure and Body Dissatisfaction after Overt Control was controlled for*

Variable	*B+*	β+	*B (SEB+)*	*t*	R^2	ΔR^2
Model 1					.06	.06
Overt Control	.25	.25	.08	3.18*		
Model 2					.08	.02
Overt Control	.35	.35	.10	3.54***		
Parental Pressure	-.12	-.16	.08	-1.61		

***Note.** B+ unstandardized coefficient; β+ = standardized coefficient; SEB+ = standard error; ΔR2 = R Square Change. ***p < .001.*p < .05.*

tal Pressure did not significantly predict Body Dissatisfaction above and beyond Overt Control.

The second hierarchical multiple regression was conducted with Parental Pressure entered at Stage 1 and Overt Control at Stage 2 (Table 3). The results revealed that at Stage 1, BPP did not contribute significantly to the regression model, $F(1, 153) = 0.39$, $p > .05$, and accounted for 0.3% of the variation in BD. Introducing Overt Control at Stage 2 explained an additional 7.6% of the variance, and this change in r^2 was significant, $F(1, 152) = 12.55$, $p < .05$. This showed that Overt Control significantly predicted Body Dissatisfaction above and beyond Parental Pressure .

Independent samples t-tests revealed that there was no significant age group difference between secondary and post-secondary adolescents in any of the study variables. The only significant gender difference was body dissatisfaction, where female adolescents scored

Table 4: *Hierarchical Regression Model examining the relationship between Overt Control and Body Dissatisfaction after Parental Pressure was controlled for*

Variable	*B+*	β+	*B (SEB+)*	*t*	R^2	ΔR^2
Model 1					.003	-.004
Overt Control	.04	-.05	.06	.63		
Model 2					.08	.07
Overt Control	-.13	-.16	.08	-1.66		
Parental Pressure	.35	.35	.10	3.54***		

Note. *B+ unstandardized coefficient; β+ = standardized coefficient; SEB+ = standard error; ΔR2 = R Square Change. ***p < .001.*

higher in body dissatisfaction compared to males, t(148) = -4.24, p < .001. Therefore, further analyses were stratified by gender, but not by age group.

The first hierarchical multiple regression conducted with the female sample had Parental Pressure entered at Stage 1 and Overt Control at Stage 2. The results revealed that at Stage 1, BPP did not contribute significantly to the regression model, *F*(1, 98) = 0.44, *p* > .05, and accounted for 0.4% of the variation in body dissatisfaction. Introducing Overt Control at Stage 2 explained an additional 4.8% of the variance in female body dissatisfaction, and this change in r^2 was significant, *F*(2, 97) = 3.50, *p* < .05. Congruent with the overall analyses, this demonstrated that Overt Control significantly predicted Body Dissatisfaction above and beyond Parental Pressure.

The second hierarchical multiple regression conducted with the female sample had Overt Control entered at Stage 1 and Parental

Pressure at Stage 2. The results revealed that at Stage 1, Overt Control contributed significantly to the regression model, $F(1, 98) = 5.50$, $p < .05$, and accounted for 5.3% of the variation in female body dissatisfaction. Introducing Parental Pressure at Stage 2 explained an additional 1.4% of the variance in female body dissatisfaction. However, this change in r^2 was not significant, $F(2, 97) = 1.48$, $p > .05$. Congruent with the overall analyses, this demonstrated that Overt Control significantly predicted Body Dissatisfaction above and beyond Parental Pressure.

A two-stage hierarchical multiple regression was conducted to determine if Parental Pressure can predict Body Dissatisfaction in males when Overt Control was accounted for. Overt Control was entered at Stage 1 and controlled at Stage 2 where Parental Pressure was entered. The results (Appendix A) revealed that at Stage 1, Overt Control did not contribute significantly to the regression model, $F(1, 48) = 3.21$, $p > .01$, but accounted for 6.3% of the variation in Body Dissatisfaction. Introducing Parental Pressure explained an additional 1.5% of the variation in Body Dissatisfaction and this change in r^2 was not significant, $F(1, 47) = 0.76$, $p > .01$.

A second two-stage hierarchical multiple regression was conducted to determine if Overt Control can predict Body Dissatisfaction in males when Parental Pressure was accounted for. Parental Pressure was entered at Stage 1 and controlled for at Stage 2 where Overt Control was entered. At stage one, Parental Pressure did not contribute significantly to the regression model, $F(1, 48) = .06$, $p > .01$, and accounted for 0.1% of the variation in Body Dissatisfaction. Introducing OC explained an additional 7.6% of the variation in Body Dissatisfaction and this change in R^2 was not significant, $F(1, 47) = 3.89$, $p > .01$. This demonstrated that neither Parental Pressure nor Overt Control uniquely predicted Body Dissatisfaction for males.

The overall correlational analyses revealed that both subscales of Overt Control were significant and positively correlated with Body Dissatisfaction; only the Teasing subscale of Body-Related Parental Pressure was significantly and positively correlated with Body Dissatisfaction (Appendix G). The same pattern of results was found in female, but not male, samples (Appendix H).

Discussion

The aim of the present study was to examine whether body-related parental pressure and overt control of food intake were both unique and positive predictors of body dissatisfaction in Singaporean adolescents. Contrary to the first hypothesis, the results of this study revealed that body-related parental pressure did not significantly predict body dissatisfaction in adolescents beyond overt control. Correlational analysis showed that parental teasing was significantly and positively linked with body dissatisfaction, which is in line with previous research (e.g., Keery et al., 2005; Yoh, 2018). However, both parental encouragement and criticism were not significantly linked with body dissatisfaction, which is incongruent with the current literature (Biolcati, Mancini & Villano, 2019; Helfert & Warschburger, 2011; Lwin & Malik, 2012; Rodgers & Chabrol, 2009).

The lack of significant effect of parental encouragement and criticism on body dissatisfaction is baffling. A plausible explanation could be that the adolescents in the sample did not perceive parental encouragement negatively. As some researchers suggest, parental encouragement may not be a unidimensional construct: it can have a negative and a positive valence (Yeatts, Martin & Farren, 2021). A longitudinal study revealed that the perceived father's encouragement for physical activity is positively related to adolescents' body satisfaction a year later (Savage, M DiNallo & Downs, 2009). This indicates that it is possible for parental encouragement to be positively received by adolescents. Similarly, parental criticism may be perceived positively by adolescents. Though speculative, the lack of effect for parental criticism may be attributed to the difference between the scale used in previous research and the current study. Specifically, prior studies tended to measure parental criticism by explicitly asking participants whether their parents "criticized" their bodies (e.g., Biolcati, Mancini & Villano, 2020; Kluck, 2010; Lwin & Malik, 2012). In contrast, the scale items used here provided examples related to parental criticism, such as "My parent(s) rebuke me for eating too much or too little food." Accordingly, adolescents may view parental criticism positively or less negatively, as they might perceive parental scolding as a form of parental concern for their health or physical well-being. However, this would require future research to verify whether parental criticism and encouragement can have both positive and negative valence.

Regarding gender difference, this study found that adolescent girls reported significantly higher body dissatisfaction than adolescent boys, which is in accordance with existing Singaporean studies (e.g., Lwin & Malik, 2012; Chng & Fassnacht, 2016). Correlational analysis by gender revealed that parental teasing was significantly linked with body dissatisfaction in adolescent girls, but not in adolescent boys. This finding aligns with a recent study that found that parental teasing is significantly linked with adolescent females' body dissatisfaction, but not for males (Yoh, 2018). In other research, no significant association between parental teasing and adolescent males' body dissatisfaction was found (Helfert & Warschburger, 2011). A probable explanation is that parental teasing tends to evoke appearance anxiety in adolescent girls. A recent study has shown that females with high social appearance anxiety reported the greatest degree of body dissatisfaction (Levinson & Rodebaugh, 2014). However, it could also be that the effect of parental teasing on adolescent males' body dissatisfaction was simply not detected because of a potential gender bias in the scale. Research suggests that while girls subscribe to the thin ideals, boys tend to value muscularity over thinness (McCreary & Sasse, 2000; Jones, 2004). As muscular concerns were not incorporated into the scale, adolescent boys' body dissatisfaction might not have been fully captured in the current study.

Consistent with the second hypothesis, the overall results of this study demonstrated that overt control positively predicted body dissatisfaction, above and beyond parental pressure. This finding corroborates previous studies (e.g., Fisher & Birch, 2000; Van Den Berg et al., 2010), which suggest that parental overt control of food intake may lead to body dissatisfaction. However, parental restriction and monitoring of food intake were significantly linked with higher body dissatisfaction in adolescent girls only. One possibility could be that adolescent girls internalize both parental and societal appearance-related expectations through parental overt control of food intake. Research has shown that internalization of societal standards of attractiveness can contribute to stronger body dissatisfaction (Carlson Jones, 2004; Thompson & Stice, 2001). Another potential reason could be the lack of statistical power to detect the effect of overt control on body dissatisfaction due to the small sample size for males. Consequently, the predictive effect of overt control on adolescent boys' body dissatisfaction was not statistically significant ($p > .05$).

Implications of the Study

The contribution of this study to the existing literature is twofold. First, this research extends knowledge by examining whether psychological and behavioural parental control can positively and uniquely predict body dissatisfaction in Singaporean adolescents.

Second, the findings can provide researchers, educators, and parents with a deeper insight into the specific parenting practices that might lead to higher levels of body dissatisfaction. As studies have shown that body dissatisfaction is the strongest predictor of eating disorders in adolescents (Abbate-Daga et al., 2010; Stice et al., 2011), the present research findings can thus help guide intervention efforts or prevention programs on eating disorders.

Strengths and Limitations

Given the lack of research exploring the effects of overt control on body dissatisfaction, the present study makes an important contribution to the current body of literature. Our findings confirm previous research speculations by demonstrating that overt control can uniquely predict greater body dissatisfaction in adolescents. Although parental pressure was not found to be a significant predictor of body dissatisfaction, these findings suggest that parental pressure (i.e., encouragement and criticism) may have positive and negative valence. However, the present study findings do have a few limitations. For starters, the study employed a convenience sampling method; hence, the sample is not representative of the Singaporean population as those of Chinese ethnicity are overrepresented in this study. Consequently, the findings of the research may not be generalizable to other racial groups. Therefore, future studies should replicate this research with a nationally representative sample. Lastly, the lack of significant effect of overt control on body dissatisfaction in adolescent boys might be due to a potential gender bias in the body dissatisfaction scale. As such, muscularity-related scales should be incorporated into future studies.

Conclusion

In summary, the findings of the present research suggest that parental overt control and parental teasing of adolescents' food intake can lead to greater body dissatisfaction in adolescents. Even though these effects were only found in adolescent girls, it would be too premature to conclude that parental overt control does not affect adolescent boys. Instead, future research should replicate the current study using a larger male sample size and a more gender-appropriate scale for body dissatisfaction. Given that body dissatisfaction is linked with eating disorders, prevention programs for eating disorders should therefore address parental teasing and overt control, such as educating parents on the deleterious outcomes of these behaviours on adolescents.

References

Barber, B.K. (1996). Parental Psychological Control: Revisiting a neglected construct. *Child Development, 67(6),* 3296. https://doi.org/10.2307/1131780

Barber, B.K., Stolz, H.E., Olsen, J.A., Collins, W.A., & Burchinal, M. (2005). Parental support, psychological control, and behavioral control: Assessing relevance across time, culture, and method. *Monographs of the Society for Research in Child Development,* i–147 https://doi.org/10.1111/j.1540-5834.2005.00365.x

Barker, E.T., & Galambos, N.L. (2003). Body dissatisfaction of adolescent girls and boys: Risk and resource factors. *The Journal of Early Adolescence, 23(2),* 141–165. https://doi.org/10.1177/0272431603023002002

Biolcati, R., Mancini, G., & Villano, P. (2020). 'And yet I'm an adult now'. The influence of parental criticism on women's body satisfaction/dissatisfaction during emerging adulthood. *International Journal of Adolescence and Youth, 25(1),* 599–608. https://doi.org/10.1080/02673843.2019.1699433

Birch, L.L., Fisher, J.O., Grimm-Thomas, K., Markey, C.N., Sawyer, R., & Johnson, S.L. (2001).

Confirmatory factor analysis of the Child Feeding Questionnaire: a measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. *Appetite, 36(3),* 201–210. https://doi.org/10.1006/appe.2001.0398

Carlson Jones, D. (2004). Body image among adolescent girls and boys: a longitudinal study. *Developmental Psychology, 40(5),* 823. https://doi.org/10.1037/0012-1649.40.5.823

Carper, J.L., Fisher, J.O., & Birch, L.L. (2000). Young girls' emerging dietary restraint and disinhibition are related to parental control in child feeding. *Appetite, 35(2),* 121–129. https://doi.org/10.1006/appe.2000.0343

Chng, S.C., & Fassnacht, D.B. (2016). Parental comments: Relationship with gender, body dissatisfaction, and disordered eating in Asian young adults. *Body Image, 16,* 93–99. https://doi.org/10.1016/j.bodyim.2015.12.001

Cho, P.L.Y., Ong, A.S.E., & Cheung, H.S. (2020). Where authoritarianism is not always bad: Parenting styles and romantic relationship quality among emerging adults in Singapore. *Current Psychology*. https://doi.org/10.1007/s12144-020-00978-9

Costanzo, P.R., & Woody, E.Z. (1985). Domain-specific parenting styles and their impact on the child's development of particular deviance: the example of obesity proneness. *Journal of Social and Clinical Psychology, 3(4),* 425–445. https://doi.org/10.1521/jscp.1985.3.4.425

Cragun, D., DeBate, R.D., Ata, R.N., & Thompson, J.K. (2013). Psychometric properties of the Body Esteem Scale for Adolescents and Adults in an early adolescent sample. *Eating and Weight Disorders, 18(3),* 275–282. https://doi.org/10.1007/s40519-013-0031-1

Deci, E.L., & Ryan, R.M. (1985). Intrinsic Motivation and Self-Determination in Human Behavior. https://doi.org/10.1007/978-1-4899-2271-7

Dreher, D.E., Feldman, D.B., & Numan, R. (2014). Controlling Parents Survey. *College Student Affairs Journal, 32(1),* 97–111. https://www.researchgate.net/profile/Diane-Dreher publication/274458014_Dreher_D_E_Feldman_D_B_Numan_R_2014_Controlling_parents_survey_Measuring_the_influence_of_parental_control_on_personal_development_in_college_students/links/55a5c72408ae81aec91373ff/Dreher-D-E-Feldman-D-B-Numan-R-2014-Controlling-parents-survey-Measuring-the-influence-of-parental-control-on-personal-development-in-college-students.pdf

Enten, R.S., & Golan, M. (2009). Parenting styles and eating disorder pathology. *Appetite, 52(3),* 784–787. https://doi.org/10.1016/j.appet.2009.02.013

Farrow, C.V. (2012). Do parental feeding practices moderate the relationships between impulsivity and eating in children?. *Eating Behaviors, 13(2),* 150–153. https://doi.org/10.1016/j.eatbeh.2011.11.015

Fisher, J.O., & Birch, L.L. (2000). Parents' restrictive feeding practices are associated with young girls' negative self-evaluation of eating. *Journal of the American Dietetic Association, 100(11),* 1341–1346. https://doi.org/10.1016/S0002-8223(00)00378-3

Fisher, J.O., & Birch, L.L. (1999). Restricting access to palatable foods affects children's behavioral response, food selection, and intake. *The American Journal of Clinical Nutrition, 69(6),* 1264–1272. https://doi.org/10.1093/ajcn/69.6.1264

Fletcher, A.C., Steinberg, L., & Williams-Wheeler, M. (2004). Parental Influences on Adolescent Problem Behavior: Revisiting Stattin and Kerr. *Child Development, 75(3),* 781–796. https://doi.org/10.1111/j.1467-8624.2004.00706.x

Gray, M.R., & Steinberg, L. (1999). Unpacking Authoritative Parenting: Reassessing a Multidimensional Construct. *Journal of Marriage and the Family, 61(3),* 574–587. https://doi.org/10.2307/353561

Griffiths, S., Murray, S.B., Bentley, C., Gratwick-Sarll, K., Harrison, C., & Mond, J.M. (2017). Sex Differences in Quality of Life Impairment Associated With Body Dissatisfaction in Adolescents. *Journal of Adolescent Health, 61(1),* 77–82. https://doi.org/10.1016/j.jadohealth.2017.01.016

Grolnick, W.S., & Pomerantz, E.M. (2009). Issues and challenges in studying parental control: Toward a new conceptualization.

Child Development Perspectives, 3(3), 165–170. https://doi.org/10.1111/j.1750-8606.2009.00099.x

Helfert, S., & Warschburger, P. (2011). A prospective study on the impact of peer and parental pressure on body dissatisfaction in adolescent girls and boys. *Body Image, 8(2),* 101–109. https://doi.org/10.1016/j.bodyim.2011.01.004

Ho, T.F., Tai, B.C., Lee, E.L., Cheng, S., & Liow, P.H. (2006). Prevalence and profile of females at risk of eating disorders in Singapore. *Singapore Medical Journal, 47(6),* 499. http://www.smj.org.sg/sites/default/files/4706/4706a5.pdf

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6(1),* 1–55. https://doi.org/10.1080/10705519909540118

Jáuregui Lobera, I., Bolaños Ríos, P., & Garrido Casals, O. (2011). Parenting styles and eating disorders: Parenting styles and eating disorders. *Journal of Psychiatric and Mental Health Nursing, 18(8),* 728–735. https://doi.org/10.1111/j.1365-2850.2011.01723.x

Johnson, S.L., & Birch, L.L. (1994). Parents' and children's adiposity and eating style. *Pediatrics, 94(5),* 653–661. https://doi.org/10.1542/peds.94.5.653

Kenyon, D.B., Fulkerson, J.A., & Kaur, H. (2009). Food hiding and weight control behaviors among ethnically diverse, overweight adolescents. Associations with parental food restriction, food monitoring, and dissatisfaction with adolescent body shape. *Appetite, 52(2),* 266–272. https://doi.org/10.1016/j.appet.2008.10.004

Keery, H., Boutelle, K., Van Den Berg, P., & Thompson, J.K. (2005). The impact of appearance-related teasing by family members. *Journal of Adolescent Health, 37(2),* 120–127. https://doi.org/10.1016/j.jadohealth.2004.08.015

Kluck, A.S. (2010). Family influence on disordered eating: The role of body image dissatisfaction. *Body Image, 7(1),* 8–14. https://doi.org/10.1016/j.bodyim.2009.09.009

Levinson, C.A., & Rodebaugh, T.L. (2015). Negative social-evaluative fears produce social anxiety, food intake, and body dissatisfaction: evidence of similar mechanisms through different pathways. *Clinical Psychological Science, 3(5),* 744–757. https://doi.org/10.1177/2167702614548891

Lwin, M.O., & Malik, S. (2012). The role of media exposure, peers, and family on body dissatisfaction amongst boys and girls in Singapore. *Journal of Children and Media, 6(1),* 69–82. https://doi.org/10.1080/17482798.2011.633406

McCreary, D.R., & Sasse, D.K. (2000). An exploration of the drive for muscularity in adolescent boys and girls. *Journal of American College Health, 48(6),* 297–304. https://doi.org/10.1080/07448480009596271

Menzel, J.E., Schaefer, L.M., Burke, N.L., Mayhew, L.L., Brannick, M.T., & Thompson, J.K. (2010). Appearance-related teasing, body dissatisfaction, and disordered eating: A meta-analysis. *Body Image, 7(4),* 261–270. https://doi.org/10.1016/j.bodyim.2010.05.004

Myers, T.A., & Crowther, J.H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *Journal of Abnormal Psychology, 118(4),* 683–698. https://doi.org/10.1037/a0016763

Ogden, J., Reynolds, R., & Smith, A. (2006). Expanding the concept of parental control: a role for overt and covert control in children's snacking behaviour?. *Appetite, 47(1),* 100–106. https://doi.org/10.1016/j.appet.2006.03.330

Oh, J.Y., & Davis, C. (2020). Eating Disorders in Children and Adolescents. *National Cancer Center Singapore.* https://www.nccs.com.sg/news/medical-news-singhealth/eating-disorders-children-adolescents

Paxton, S.J., Eisenberg, M.E., & Neumark-Sztainer, D. (2006). Prospective predictors of body dissatisfaction in adolescent girls and boys: a five-year longitudinal study. *Developmental Psychology, 42(5),* 888–899. https://doi.org/10.1037/0012-1649.42.5.888

Pomerantz, E.M., & Wang, Q. (2009). The role of parental control in children's development in western and east asian countries. *Current Directions in Psychological Science, 18(5),* 285–289. https://doi.org/10.1111/j.1467-8721.2009.01653.x

Pötzsch, A., Rudolph, A., Schmidt, R., & Hilbert, A. (2018). Two sides of weight bias in adolescent binge-eating disorder: Adolescents' perceptions and maternal attitudes. *International Journal of Eating Disorders, 51(12),* 1339–1345. https://doi.org/10.1002/eat.22982

Reilly, E.E., Stey, P., & Lapsley, D.K. (2016). A new look at the links between perceived parenting, socially-prescribed perfectionism, and disordered eating. *Personality and Individual Differences, 88,* 17–20. https://doi.org/10.1016/j.paid.2015.08.038

Rivero, A. (2021). Sociocultural attitudes as a moderator of the relations between negative eating and weight messages from family members and Latinas' body image shame (Doctoral dissertation, University of Missouri--Columbia). https://doi.org/10.32469/10355/85857

Robertson, J. (2020). The Parent Behind the Eating Disorder: How Parenting Styles Affect Eating Disorders in Adolescents. *Family Perspectives, 2(1),* 3. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1031&context=familyperspectives

Rodgers, R., & Chabrol, H. (2009). Parental attitudes, body image disturbance and disordered eating amongst adolescents and young adults: A review. *European Eating Disorders Review, 17(2),* 137–151. https://doi.org/10.1002/erv.907

Rodgers, R.F., Paxton, S.J., & Chabrol, H. (2009). Effects of parental comments on body dissatisfaction and eating disturbance in young adults: A sociocultural model. *Body Image, 6(3),* 171–177. https://doi.org/10.1016/j.bodyim.2009.04.004

Savage, J.S., DiNallo, J.M., & Downs, D.S. (2009). Adolescent body satisfaction: The role of perceived parental encouragement for physical activity. *International Journal of Behavioral Nutrition and Physical Activity, 6(1),* 90. https://doi.org/10.1186/1479-5868-6-90

Stice, E., Marti, C.N., & Durant, S. (2011). Risk factors for onset of eating disorders: Evidence of multiple risk pathways from an 8-year prospective study. *Behaviour Research and Therapy, 49(10),* 622–627. https://doi.org/10.1016/j.brat.2011.06.009

Taylor, A., Wilson, C., Slater, A., & Mohr, P. (2012). Self-esteem and body dissatisfaction in young children: Associations with weight and perceived parenting style. *Clinical Psychologist, 16(1),* 25–35. https://doi.org/10.1111/j.1742-9552.2011.00038.x

Thompson, J.K., Cattarin, J., Fowler, B., & Fisher, E. (1995). The perception of teasing scale (Pots): A revision and extension of the physical appearance related teasing scale(Parts). *Journal of Personality Assessment, 65(1),* 146–157. https://doi.org/10.1207/s15327752jpa6501_11

Thompson, J.K., & Stice, E. (2001). Thin-ideal internalization: Mounting evidence for a new risk factor for body-image disturbance and eating pathology. *Current Directions in Psychological Science, 10(5),* 181–183. https://doi.org/10.1111/1467-8721.00144

Wang, Q., Pomerantz, E.M., & Chen, H. (2007). The Role of Parents? Control in Early Adolescents? Psychological Functioning: A Longitudinal Investigation in the United States and China. *Child Development, 78(5),* 1592–1610. https://doi.org/10.1111/j.1467-8624.2007.01085.x

Wills, T.A., Isasi, C.R., Mendoza, D., & Ainette, M.G. (2007). Self-control constructs related to measures of dietary intake and physical activity in adolescents. *Journal of Adolescent Health, 41(6),* 551–558. https://doi.org/10.1016/j.jadohealth.2007.06.013

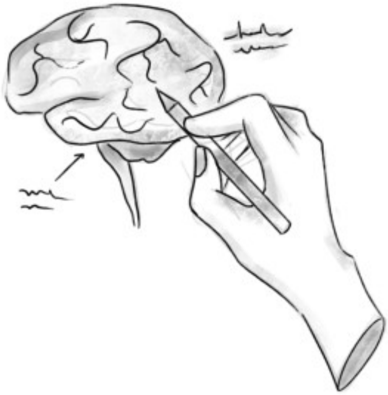
Yeatts, P.E., Martin, S.B., & Farren, G.L. (2021). Adolescents' psychological well-being and their perceptions of parental encouragement to control weight. *Journal of Family Studies, 27(4),* 607–620. https://doi.org/10.1080/13229400.2019.1674682

Yoh, E. (2018). Teasing from Parents and Siblings about Appearance Affecting Body Satisfaction and Self-Esteem of Middle School Students. Fashion, *Industry and Education, 16(2),* 17–29. https://doi.org/10.7741/fie.2018.16.2.017

Zubatsky, M., Berge, J., & Neumark-Sztainer, D. (2014). Longitudinal associations between parenting style and adolescent disordered eating behaviors. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 20(2),* 187–194. https://doi.org/10.1007/s40519-014-0154-z

Seeing Myself; Seeing Myself; Seeing Myself

Valerie Nguyen



Deinforcement Learning

Janna Hong

University of Washington

Department of Neuroscience

Varun Ananth

University of Washington

Paul G. Allen School for Computer Science

Chaytan Inman

University of Washington

Paul G. Allen School for Computer Science

Marlene Grieskamp

University of Washington

Paul G. Allen School for Computer Science

Abstract

An important challenge of human decision-making is determining via trial and error which options maximize reward and minimize punishment. In computer science, this problem is known as reinforcement learning (RL), and particular RL paradigms, such as the advantage actor-critic (A2C), have been the subject of extensive research (Niv, 2009). Current RL algorithms are insufficient representations of the brain, despite the fact that this biological analogy has historically advanced the field of computer science (Tassa et al., 2018). When mimicking dopamine pathways, RL often disregards one of the most potent biological signals: pain. The absence of a reward signal, also known as a negative signal, is frequently interpreted as being equivalent to punishment (Schultz et al., 1997). However, the biological mechanisms that interpret, transmit, and permit pain in the body contradict this assertion. We argue that people avoid unfavourable situations more rapidly if they learn through pain as opposed to through a lack of reward. Therefore, we propose that adding pain into current RL models will not only allow algorithms to converge more quickly, but also cause behaviour to become more safe, sophisticated, and generalizable. This work examines the historical connections between RL and neuroscience, synthesizes neuroscientific understandings of pain, and proposes refinements to current biologically inspired techniques for incorporating pain into RL algorithms.

Keywords: **machine learning, reinforcement learning, neuroscience, cognitive science, psychology.**

Introduction

The history of reinforcement learning (RL) is indelibly wound with the mathematical foundations of neuroscience. Neuroscience has much to offer computer science, given that the brain is one of the most flexible, adaptive, efficient, and intelligent learning paradigms. This long-standing relationship has yielded significant benefits, but more multidisciplinary effort is required to reach the full potential of applications to computer science as neuroscience advances.

Richard Sutton, who was trained in both psychology and computer science, created the 'Temporal Difference Learning' (TD Learning) algorithm in the 1980s to explain the neuronal response to violated expectations. This marked the beginning of the intertwined history of computer science and neuroscience. The discrepancy between the predicted and actual reward is the definition of prediction error. TD Learning asserts that learning occurs through prediction error. The strong similarity between dopamine firing rates and the reward prediction error (RPE) signal was then described by Wolfram Schultz and colleagues in 1997.

Schultz demonstrated that when monkeys get an unanticipated reward, their brains respond with an influx of action potentials, often known as a phasic firing pattern (Schutlz et al., 1997). Nonetheless, the phasic pattern disappears when the incentive is anticipated. These results demonstrated that TD error (the difference between the predicted and actual firing rates) was stored in the spiking patterns of dopaminergic neurons.

At the same time, developments in reinforcement learning initiated the use of TD error to update the weights of neurons in artificial neural networks. In this approach, reinforcement learning has created unique optimization algorithms that mimic human-like learning. Popular algorithms such as actor-critic reinforcement learning continue to rely on TD error. Given the significance of programmatic algorithms that replicate ancient biological brain algorithms, it is imperative that new RL paradigms continue to emerge. To date, the brain is more generalizable than any known learning algorithm; we have much to learn from studying it. Thus, we aim to improve the performance of reinforcement learning algorithms by upgrading models that learn from both pain and reward, with our implementation based on a detailed evaluation of the various neuroscience studies on pain pathways. We term the general idea of properly adding pain to RL using biological justifications "Deinforcement Learning". By comparing and contrasting Deinforcement Learning with standard RL methods, we intend to demonstrate the significance of learning from negative incentives in the building of more resilient and effective RL models.

Roadmap

We begin with a review of dopaminergic learning and reinforcement learning in the brain. We then pull concepts from psychology to differentiate between various types of reinforcement and punishment, as well as consider which ones are relevant to our ultimate theoretical implementation into RL. The pain route is then compared to the previously outlined dopaminergic learning pathway. In conclusion, we synthesize these ideas to present a high-level RL framework based on the MaxPain algorithm while incorporating a biologically-inspired pain signal, distinct from a negative reward signal.

Dopaminergic Learning

Dopamine is a neurotransmitter with a reputation for its reward and pleasure qualities in popular culture. Recent studies indicate that dopamine is also implicated in pain, which may involve the remodeling of the reward circuitry (Markovic et al., 2021). Approximately 90% of

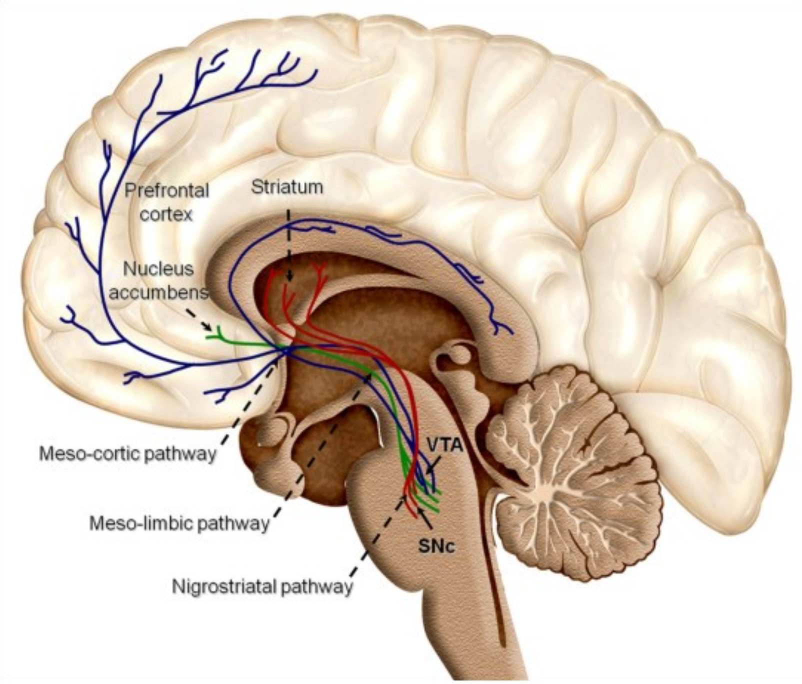


Figure 1. *The mesolimbic pathway shows the dopaminergic neurons made in the ventral tegmental area projecting to the nucleus accumbens.* (*Arias-Carrión et al., 2014*)

dopamine producing neurons are in two areas of the midbrain nuclei called the substantia nigra pars compacta (SNc) and the mesolimbic ventral tegmental area (VTA) (Arias-Carrión et. al, 2014). These neurons project to the nucleus accumbens (Naci), which is the reward-related dopamine site. Blocking the dopamine pathway to NAc deprives the rewarding effects.

In the early 1950s, scientists discovered the significance of the VTA region by observing the effects of electrical stimulation to certain regions of the brain in rats. When electrical stimulation in the VTA was followed by a certain task such as lever pressing, rats repeatedly executed that action (Olds & Milner, 1954). The rats ended up pressing on the lever 2000 times per hour when they learned that this specific behaviour reliably leads to an electrical stimulation. In this case, the action of lever pulling is followed by the electrical stimulation, thus the stimulus is the "reward". Olds and Milner elaborate on the concept of reward as follows: "In its reinforcing capacity, a stimulus increases, decreases, or leaves unchanged the frequency of preceding responses, and accordingly it is called a reward, a punishment, or a neutral stimulus" (Olds & Milner, 1954, p. 419). Recent experiments in humans undergoing deep brain stimulation (DBS) for Parkinson's show similar results. After participants learned that a certain task was followed by electrical stimulation of the SNc, an area with abundant dopaminergic neurons, they repeatedly performed that task, soon even without the stimulus (Perelman School of Medicine at the University of Pennsylvania, 2014).

Dopaminergic neurons (DA) can fire in two distinct patterns in response to varying stimuli: phasic and tonic activity. Phasic activity refers to a burst of action potentials firing in a short period of time, with a rate of up to 20 Hz. In contrast, tonic activity indicates a steady firing rate of around 5 Hz. Tonic activity recorded in monkeys implies congruence between actual and expected reward, while phasic signal indicates a component of surprise and mediates a prediction error during learning (Schultz et. al, 1997). As a result, dopaminergic reactions diminish as learning and estimation of rewards progress. Likewise, the phasic firing activity decreases following the delivery of the reward. DA might indicate the difference between the anticipated reward and the actual reward. Though the complete neurological effects of dopamine, such as tonic firing pattern effects, have not yet been integrated into reinforcement learning models, contemporary improvements to reinforcement learning are based on the understanding of the phasic dopamine response (Beeler et. al. 2010). By comprehending the phasic and tonic expressions of dopaminergic neurons, we can inform the future direction of action selection in models for temporal difference reinforcement learning.

Reinforcement Learning

Reinforcement Learning (RL) is a branch of machine learning involved with choosing optimal policies, state value estimates, or both to optimize an agent's selection of the most rewarding action in a given environment. RL has been used by companies such as DeepMind to train humanoid and non-humanoid physical models to walk, run, jump, and play games (Tassa et. al, 2018). It tends to perform best in scenarios where there are complex states and decisions are plentiful, not dissimilar to our own environment. The advantage actor-critic (A2C) RL model employs TD learning to adjust its reward predictions through time, analogous to how humans learn through classical and operant conditioning (Niv, 2009). We have chosen to study A2C because it contains the most straightforward implementation of advantage. Before exploring the relevance of the "advantage" variable and its linkages to reward prediction error (RPE) in the brain, it is crucial to differentiate between the actor and critic components.

The actor is (often) a neural network which learns policy π parameterized by ϑ, that is a function of state s. In other words, a policy determines which action should be taken in a given state and this choice is influenced by the internal values of ϑ (in this case the neural network weights). Policies can be tuned to maximize reward (as is the case with A2C), or achieve a goal parameterized as a function of that reward. Actors in A2C are stochastic by nature, meaning they output a probability distribution for taking an action in the action space based on the current state (Geron, 2019). This has high level parallels to how we as humans interact with our environment. We consider the "state" of the world around us and then "act" to maximize some reward tied to a goal. For example, when you decide to take action

"A" to walk your dog, one might think of this as applying your internal policy (π) to the state s, wherein your dog is barking at the door wanting to go for a walk. You have, in this example, tuned your internal weights ϑ so that your policy outputs the action "A" of dog walking when given the state "s" of your dog barking. In many cases, the actor is a deep neural network with input dimensions being equivalent to the dimensionality of the state, and the output dimensionality being equivalent to that of the action space. Actions selected by the actor are taken in the environment, physical or simulated, and they will receive a scalar reward for taking that action. The actor's weights are updated with the critic's weights along the "advantage" variable, which will be discussed in a later section.

The critic network in A2C is more abstract. Its purpose is to approximate the function $V(s)$. This function returns the overall value of being in a state s, given a policy π. It is equivalent to the expected return of starting in state "s" and following policy π thereafter (return in terms of discounted rewards) (Mnih et al., 2016).

Building upon the dog walking example; assume two separate states s_{leash} and s_{no_leash}. In these states, your dog is barking, and you have a leash or do not have a leash respectively. Say you do not have it and decide to walk the dog anyways (following policy π where you walk the dog if it is barking). It runs away from you and now you have no dog, which is a terrible situation to be in. If you had the leash, you would walk your dog without incident and your situation is overall better. The "value" of the state s_{leash} is greater than the "value" of the state s_{no_leash} because after following policy π in both of those states one led to ruin and the other was just fine.

$$\infty$$

$$\pi(s) = E\pi\{Rt|st = s\} = E\pi\{X\gamma r(st+k+1, at+k+1)|st = s\}$$

$$k = 0$$

In most cases, the "perfect" V function shown above is incomputable. Such is the case when we do not have access to the rewards for any given state, which we do not when we do not have a model of the environment. $V(s)$ is, however, commonly estimated by another deep neural network: in our case the critic. The input of the critic is the state, and the output is a scalar estimate of the value function. The critic is important because the value function it estimates is used in the calculation of the crucial "advantage" variable which is then used to update the weights of both itself and the actor. The actor, as discussed before, is the driving force behind the model's decision making. A better performing actor means a better performing network.

The following sections explore the relevance of extending the synergy between reinforcement learning and dopaminergic learning beyond reward-based learning.

Dopaminergic Influences on Actor-Critic Systems

Temporal Difference learning is the framework upon which actor-critic updates are based, and TD error calculations in RL are done with the use of the advantage variable; the advantage can be broken down to its constituent parts below:

$$A = r(s_t, a_t) + \gamma V(s_{t+1}) - V(s_t)$$

Assuming the agent just took an action and has moved to a new state: $r(s_t, a_t)$ is the reward given for taking that certain action in that certain state. Reward is often given by the environment after an action is taken and not model-intrinsic. The $\gamma V(s_{t+1})$ term represents the estimated future discounted rewards starting at the new state the agent has just entered. Finally, the $V(s_t)$ term is the estimated value function for the state the agent was just in before acting. Combining the first two terms, $r(s_t, a_t) + \gamma V(s_{t+1})$, we get the reward the agent received for taking an action in its previous state plus the discounted predicted value of the agent's current state. We call this the TD Target. Recall that the critic network is being trained to provide precise estimates of $V(s)$. To train the network with backpropagation, we need some ground truth to understand how well the model predicted the V function for this timestep.

The TD Target contains the ground truth in the form of $r(s_t, a_t)$. $V(s_t)$ is subtracted from the TD target, and the difference between these two is the TD Error. The term "error" refers to the discrepancy that exists between the rewards that were actually gained and those that were estimated to be obtained. Note that the $\gamma V(s_t+1)$ term is partially nullified by the tail end of the $V(s_t)$ term, since this value function contains discounted value predictions for future states as well (Watabe-Uchida et al., 2017). The worse the prediction of $V(s_t)$ by the critic, the further from zero the advantage is, and vice versa. This is why (A^2) is the critic loss in an A2C network: squaring the advantage allows us to treat negative and positive TD Error equivalently while preserving the differentiability of the function and advantage is an appropriate measure of how well the network is doing at approximating $V(s_t)$. The actor's loss is also scaled by the advantage, but the calculation of loss for a stochastic model is beyond the scope of this paper.

Since TD learning lays the groundwork for calculating advantage in actor-critic systems, it stands to reason that A2C weight updates are linked to dopaminergic learning in the brain. In the case of RPE, as more stimuli are experienced by the organism and dopaminergic learning occurs (see Dopaminergic Learning section), the RPE approaches zero. This means the organism has learned how to correctly predict the reward given its state (Schultz et al., 1997). Just as RPE lowers while an organism learns to predict reward, the advantage variable in an actor-critic system lowers as the critic learns to estimate $V(s)$.

If no reward is present but reward is predicted, then dopamine activity is heavily depressed which causes updates to the organism's "value estimation function". However, an important distinction needs to be drawn between the lack of a reward and a punishment. In RL, negative rewards stemming from a state-action pair are seen as "punishment" for a model. Initially, this seems accurate. Value updates propagate through a network, telling the agent that this state is not as valuable as it initially estimated. However, if we consider on a high level the effect of pain, we see that there is a large difference in how organisms react to pain versus negative reward. If a child puts their hand on a hot plate, they are unlikely to do it again, effective immediately. An A2C agent would perhaps place its hand on the hot plate thousands of iterations under the guise of "exploration". The result: slow model convergence and/or unsafe actions, experienced by almost all modern reinforcement learning algorithms (Ghiassian et al., 2020). Learning for these models can be made more efficient by recognizing these fundamental differences in learning mechanisms.

Operant Conditioning

In psychology, operant conditioning explores how humans learn and displays the cognitive and behavioural differences in learning from positive stimuli, negative stimuli, and lack thereof (Grison & Gazzaniga, 2019). While operant conditioning is criticized as an overly simplistic view of human learning, it serves to differentiate between negative reward and pain. In operant conditioning, *response cost punishment* describes the removal of a positive stimulus as opposed to *aversive punishment,* or the addition of a negative stimulus.

Numerous operant conditioning experiments demonstrate the differences between aversive punishment and positive reinforcement

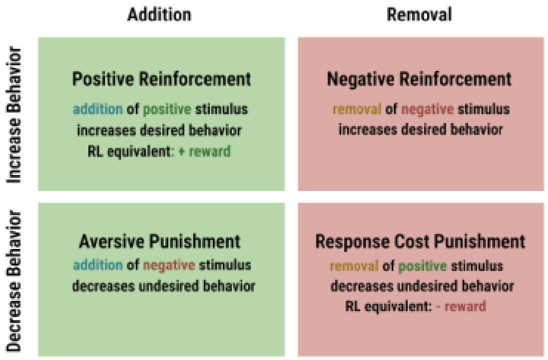
(Gershman, 2015; Kubanek et al., 2015; Steel, 2016). These findings imply that there is a behavioural, and consequently, a neurological difference in aversive punishment and reward, and also between response cost punishment and aversive punishment. Consider a child with an obsessive desire to climb trees. Their parents may warn them about the dangers of falling and perhaps take away their video game privileges as a consequence for climbing (response cost punishment). However, it is likely that the stubborn child will continue climbing until they fall and break their leg. The painful broken leg (an aversive punishment) is a much faster and stronger conditioning response than losing gaming privileges. In other scenarios, such as doing homework and getting a problem wrong, aversive punishment is far less effective than response cost, as it may deter the child from attempting the homework in the first place. It is evident from these psychological principles that both types of punishment are required for efficient learning.

As our current concepts of pain in traditional reinforcement learning are based solely on reward, we only observe the equivalent response cost punishment and positive reinforcement without any use of aversive punishment and negative reinforcement. Next sections explore the neuronal differences in these phenomena (reward and pain) in greater depth.

Neurological Pain Pathways

Neurological Pain Pathways in the MaxPain Model

In this section, we examine the neurological evidence of dissociable processes in the prediction of punishment in action systems. In light of these findings, we examine a recent technique termed "MaxPain" that uses an RL framework to strike an equilibrium between punishment and reward prediction (See Discussion: Existing Literature and



	Addition	Removal
Increase Behavior	**Positive Reinforcement** addition of positive stimulus increases desired behavior RL equivalent: + reward	**Negative Reinforcement** removal of negative stimulus increases desired behavior
Decrease Behavior	**Aversive Punishment** addition of negative stimulus decreases undesired behavior	**Response Cost Punishment** removal of positive stimulus decreases undesired behavior RL equivalent: - reward

Figure 2. *A grid showing different operant conditioning punishment/reward delineations and what they mean, along with their RL "equivalent."*

MaxPain). Several studies corroborate the MaxPain algorithm's central tenet that positive and negative reinforcement have distinct but complementary effects on learning and ultimately converge in the brain's striatum. Recent research indicates that it is capable of distinguishing between pain/punishment and penalty omission learning rates when utilizing TD-learning models to represent behaviour during an avoidance learning task (Elfwing et al., 2017). In addition, there is mounting evidence that the decision-making processes of animals incorporate separate reward and punishment systems, calling into question the fundamental validity of this approach. These results not only give a theoretical foundation for understanding punishment in the brain in both health and sickness, but they also underscore the necessity for independent punishment prediction in RL. While it is true that pain can activate regions of the brain involved in our reward circuitry, additional variables impact pain perception and should be researched to improve the current simplified model (Schmidt et al., 2002).

Basic Circuitry of Pain

The mesolimbic reward circuitry, including VTA-to-NAc dopaminergic projections, modulates pain. Dopaminergic neurons produced in the ventral tegmental region project to the nucleus accumbens through the mesolimbic route (Russo et al., 2013). In rodents, prolonged pain triggers dopamine release in the NAc and painful events can rapidly excite the dopaminergic neurons in the VTA (Schmidt et al., 2002). Furthermore, the spinal cord delivers afferent nociceptive pain signals to the brain. The thalamus communicates with the main somatosensory cortex (S1, S2), the anterior cingulate cortex (ACC), and the insula. Similarly, the basal ganglia gets signals from the amygdala (Bushnell et al., 2013). A visual representation of the distinct differences in the reward and pain pathways are summarized in Figure 3 below.

To examine the neurological basis of RL in the human brain, researchers rely on functional magnetic resonance imaging (fMRI), which permits the non-invasive monitoring of neural activity correlations. Temporal difference models reflect neurophysiological data like fMRI scans, and Pavlovian reinforcement learning works in a way that is comparable to reinforcement learning (Niv, 2009). Investigations of the neural circuit employing fMRI on avoidance revealed that action learning may be properly represented using basic temporal difference action-learning models (Sutton, 1988), with consistently identifiable prediction errors in dorsal striatal areas (Kim, 2006). Experiment results revealed an increase in activity in the medial orbitofrontal cortex, a region previously associated with the storing of the value of sensory pleasures, after individuals averted a negative outcome and were rewarded. Furthermore, as demonstrated in rats, dopamine treatments increase reward responses but not avoidance responses, indicating that the two ostensibly contradictory actions have distinct neural substrates (Fernando et al., 2013). Also, shock elicited much larger responses in the striatum than no-shock, indicating that learning may happen through punishment (Eldar et al., 2016). Overall, prediction errors that converge to the striatum are "reward-signed" in those who learn mostly from omission events and "punishment-signed" (aversive) in those who learn largely from punishment. This provides conclusive evidence that many action-value signals, including those for reward and punishment, converge on the striatum to govern behaviour.

Convergence of avoidance and reward acquisition values has been explored utilizing separate mixed reward-punishment schemes, and the outcome was independent. Studies from O'Doherty and colleagues (2004) revealed that fMRI correlates of prediction error signals may be distinguished in the dorsal and ventral striatum based on whether active decision behaviour is required to obtain a reward related to Pavlovian conditioning. During the active choice task, the reward prediction error was observed in both the ventral and dorsolateral striatum, but during the passive prediction-learning task, it was only observed in the ventral striatum. These results supported a previously proposed Actor/Critic architecture in the basal ganglia. The ventral striatum, according to this hypothesis, comprises a prediction-learning Critic, whereas the dorsal striatum has a policy-learning Actor (Joel et al., 2002). These results support the MaxPain model.

Thus, the interpretation of pain in MaxPain is reliable, but due to the subjective nature of pain and the complexity of the field, we believe there are many alternative ways to accurately model high-level concepts of pain in RL. Individual differences in pain processing add another layer of complexity to the pain pathway, which extends beyond the convergence of reward and punishment signals on the stria-

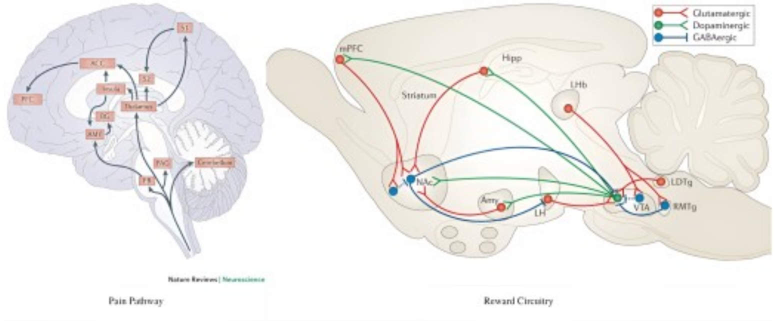


Figure 3. *A contrast of the neural pathways controlling painful and rewarding sensations respectively. Acute pain begins at nociceptors—specific receptors of somatic neurons that detect noxious stimuli apart from other stimuli. One of the first modules of the pain pathway in the brain is the thalamus; typical reward circuitry does not involve the thalamus. (Left) The spinal cord sends afferent nociceptive pain information to the brain. The pathway covers several brain regions; the primary somatosensory cortex (S1, S2), anterior cingulate cortex (ACC), and insula receives information from the thalamus. Similarly, the basal ganglia receives information from the amygdala (Bushnell et al., 2013). (Right) The mesolimbic pathway shows dopaminergic neurons made in the ventral tegmental area projecting to the nucleus accumbens (Russo et al., 2013).*

tum. In spite of many behavioural parallels between not receiving a reward and being given a punishment, these two events appear to be unique in terms of prediction learning, and the substrate for unpleasant prediction learning is still unknown. It is important to note, however, that these studies do not equate the mechanism of learning from painful stimuli to that of reward stimuli. Although reward is clearly intertwined in pain/pleasure, physiological response, learning rate, observed behaviour, and our own experience draws a clear distinction between the two. Furthermore, signaling or prediction mistakes for negative outcomes do not always appear to involve dopaminergic neurons (Mirenowicz & Schultz, 1996), despite the fact that they indicate negative prediction errors due to the absence of appetitive events (Bayer et al., 2007).

Using a unique brain mapping method, Kohoutová and colleagues located regions of the brain that show either high or low inter-individual variability in relation to pain. In addition to the anterior midcingulate cortex, the dorsolateral prefrontal cortex, and the cerebellum, twenty-one other brain areas have been shown to have a role in pain prediction as well. And contrary to common assumption, electropharmacogram analysis of brain recordings reveals that punishment prediction errors have been recorded in several brain areas, including the insula cortex, co-occurring with and with opposite sign to reward (Pessiglione et. a., 2006). Given the pre-existing psychological data described in the Operant Conditioning section of this research, the variety of the areas engaged in pain processing indicates even greater variability than the negation of reward.

Moreover, pain is highly subjective, making direct measurement difficult; instead, we must rely on self-report and, to some extent, behaviour to make sense of it. Variations in cerebral activity caused by the same painful stimuli corroborate self-reported pain differences and are predicted by brain morphology (Coghill, 2003). High levels of individual variability were found in the ventromedial prefrontal cortex, whereas lower amounts were seen in the posterior midcingulate cortex, implying that these regions' contributions to pain vary greatly among people. Analyzing the brain regions collectively as opposed to independently (i.e. multivariate analysis) yielded the same results. Individual variance was highest in the ventrolateral, vermis, and ventromedial prefrontal cortex. Individually, the posterior midcingulate cortex, the supplementary motor area, and the sensorimotor cortex were the most stable regions. Intriguingly, these findings were confirmed by tests performed with a completely new set of data. Collectively, these findings show that the relationship between brain regions and pain perception at the level of the individual is more complex than it is often portrayed at the level of the group. As a result, the fact that pain is experienced differently by different people demonstrates how subjective data may be used to refine algorithms.

Both animals and humans are able to solve in online, generalized, and sample efficient manners despite the fact that real-time neural computation is severely limited; this suggests neural mechanisms can be a source for new theoretical approaches, such as modifications to improve computational efficiency and mechanisms for interacting with constant and noisy sensory experience.

We have discussed the basics of dopaminergic learning and how it relates to the A2C model from RL literature through the concept of TD-learning. We then reviewed the difference between response cost punishment and aversive punishment in the operant conditioning subsection. The former is equivalent to negative reward but is imprecisely considered punishment in the current RL paradigm. Aversive punishment more accurately portrays pain in human learning. In the Neurological Pain Pathways section, we concluded our justification of separable and unique pain and reward pathways. This is the basis for the justification of our proposed architectures. Now, we move to the discussion where we analyze existing literature and how we can improve upon the paradigm through the incorporation of the empirical and neuronal differences of pain and reward.

Discussion

Typical RL algorithms do not incorporate learning from pain. Furthermore, learning from an action resulting in negative reward, or response cost, mirrors neither observed human behaviour in aversive punishment nor the neural circuitry involved in processing pain. This is in contrast to the rough parallels of reward based operant conditioning to reinforcement learning via reward prediction error. Therefore, we review several approaches to incorporate pain with reinforcement learning, finally proposing alternatives and expansions to form a landscape of pain in RL which we call Deinforcement Learning.

Before beginning, it should be noted that in environments where reward is uniformly distributed across all possible outcomes, such as binary right/wrong object classification, there is no purpose to learning a function to approximate pain. The purpose of pain is to learn to avoid certain states much more vehemently and faster than learning from a lack of a reward in the same situation. If misclassifications are always weighted uniformly, there is no distinction between pain and a lack of a reward.

Existing Literature and MaxPain

Existing research in the complement to our argument—understanding pain through RL—is thoroughly explored in a paper titled "Pain: A Precision Signal for Reinforcement Learning and Control", which outlines the "underlying computational architecture of the pain system" (Seymour, 2019). They formalize the concept of pain in a high-level computational model that has a basis in RL. However, it does not apply these parallels as a pain signal in RL. More concretely, another paper titled "Parallel Reward and Punishment Control in Humans and Robots: Safe Reinforcement Learning Using the MaxPain Algorithm" focuses on using pain signals in RL to avoid physical damage.

The relative "safety" of this algorithm makes it useful for robotics applications, where physical damage can occur in undesirable states and therefore the agent (robot) must learn to avoid what are perceived as "painful" states (Elfwing & Seymour, 2017). Yet another paper expands on the work of Elfwing and Seymour by implementing a similar architecture with multiple Q values and learning entirely separate value estimations for the two (Lin, 2019). They dubbed this “split-Q learning”. Both of these papers show that considering pain in some form as a control signal results in faster convergence and more complex behaviour. We argue that the MaxPain implementation can be improved upon to increase its efficacy as well as its accuracy in modeling pain.

The MaxPain paper implements pain in RL by splitting the reward scalar based on its sign. Any positive rewards remain positive. Negative rewards are inverted to be positive and are now

considered “pain”. These separate signals are then evaluated by distinct networks. To estimate these values with respect to the state of the agent, the MaxPain architecture considers two distinct state-action-pair-value estimation functions (Q-functions). One is for estimating the accumulated discounted reward for taking action “a” in state “s” and thereafter following policy π. The other estimates pain in the same way. These two resulting estimations are linearly combined into one objective through the equation shown below:

$$Q_w(s,a) = wQ_r(s,a) - (1 - w)Q_p(s,a)$$

Here, w is the weighting factor between 0 and 1. From this point onward, the goal of the MaxPain algorithm is to maximize that combination of reward minus pain through following a policy. It is implicit that finding this policy allows the agent to “solve” the problem in the environment (i.e., balancing the pole in cartpole, finding the exit to a maze, etc.). The MaxPain model saw “significantly safer exploration, as well as effective learning and near-optimal long-term performance” (Elfwing & Seymour, 2017). The average learning curves presented for a “dangerous grid world search” task indicate faster convergence to a solution as well. Next, based on a neuroscientific foundation, we compare these approaches to existing RL meth-

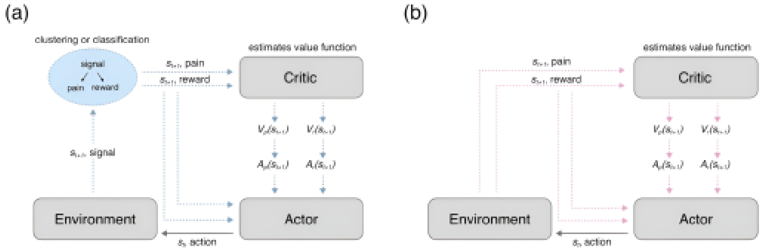


Figure 4. *Possible Implementations of Pain in A2C (a) Implementation option where the environment outputs a single feedback signal (as in traditional A2C) and this signal is split into pain and reward of various magnitudes. (b) Example where pain and reward are separate signals from the environment established through rewards-shaping.*

odologies.

Representing Pain

There appears to be two possible ways to represent pain in the context of reinforcement learning. One must distinguish between painful and non-painful states. This can either be the job of the environment, or the agent. In the former proposition, painful states may be a completely separate input, labeled as painful or not by the environment itself. For example, in the context of the game chess, the environment could send painful signals when pieces are lost, and reward when pieces are captured. This type of approach is seen in the current practice of reward shaping but is lacking a concept of pain (though the reward could have a negative sign). If we take this approach, then there is now a scalar describing the painfulness of a state and we need to augment the state to contain this new knowledge. This will allow us to describe to another network the painfulness of a state. One possibility is to use the positional encoding technique used by Vaswani et al. (2017) in “Attention is All You Need”.

But this is not how humans perceive pain as the universe does not define pain for us. Fundamentally, it is the latter approach, the one that passes a raw state to the agent and allows the agent to interpret what is reward and what is pain, that is biologically inspired. We can see this by studying the body's path of pain: nociception.

Nociception defines a distinction between cells that can receive painful input and those that do not. When you touch something, a signal is propagated along mechanoreceptors. If you touch something hard enough, pointy enough, or hot enough, the signal propagates along pain specific fibers (e.g., A-delta fibers, C-fibers) to signal an acute pain to the brain (Yam et al., 2018). As previously discussed, this pathway is disparate from that of an unpainful signal. The mechanisms that perceive pain and other stimuli fundamentally represent the state differently before interpretation in the brain. How would this look in reinforcement learning? This may take the form of a clustering algorithm whose clusters represent painful vs rewarding stimuli and various interpolations of those classes. After clustering into discrete signals, the pain and reward signals could be processed and interpreted with different learning mechanisms as they are in the brain. It may take the form of a classification neural network (or support vector machines among other algorithms), whose logit probabilities can be interpreted as dimensions along various sensory stimuli such as touch, pain, temperature etc. This leaves the state open to interpretation by multiple perception pathways; you can not only feel pain when pricked by a needle, you can also feel pressure. There are many other possibilities for representing this distinction between pain and other sensory information at the initial reception at the sensory level.

Interpreting and Learning from Pain

Next, how does one interpret pain within the RL equivalent of a brain? In our working example, the A2C method (see Reinforcement Learning section) uses the critic to evaluate how valued a state is with respect only to estimated future reward. Operant conditioning shows that reward and pain pathways trigger learning at different rates and to varying effect. Therefore, it is necessary to have different representations of $V(s)$ with respect to pain, and with respect to reward. This allows learning to be modulated according to the painfulness of experiences. One way to achieve this is to first modify the state with information from the pain classification processing mentioned above. Then this information is passed to the critic, whose weights should learn a representation of $V(s)$ which estimates and takes into account both the estimation of future pain and reward, then outputs two values of $V(s)$ with respect to both pain and reward (Figure 5). This is similar to the framework by Elfwing and Seymour (2017), but instead learns the weighting factor w of combining $Q(s,a)$ outputs. It also relies on an overparameterized critic network which converges to two nodes at some intermediate layer rather than using two separate critics.

Alternatively, the modified state information is passed to the critic, which must learn to accurately represent the state as estimations of $V(s)_{pain}$ and $V(s)_{reward}$, output these two different values, and pass these values to the actor. The actor can then learn weights to combine pain and reward into a customary single $V(s)$ rather than the two-layer output in Figure 5. This shifts the burden of estimating $V_{pain}(s)$ and $V_{reward}(s)$ to the actor, as in Figure 6. Similarly, this learns the weighting factor w to combine the different estimated values of state rather than linearly combining them as originally tested by Elfwing and Seymour (2017).

Finally, another possible implementation is to train separate critic algorithms after binary classification of pain or nonpainful stimuli. This can be likened to ensemble approaches. This approach is the least biologically faithful since states are not best represented by such

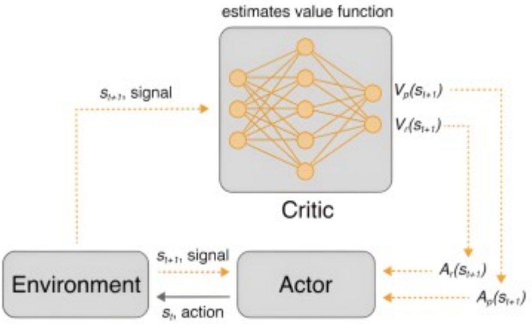


Figure 5*. Example Critic with* $V_{pain}(s)$ *and* $V_{reward}(s)$ *Outputs. Fig. 5 extends Fig. 4 with an additional example implementation of pain in A2C, wherein the output layer of the critic has two nodes in order to explicitly represent the approximated* V (s) *with respect to both pain and reward. The network's overall size could have any number of layers and nodes, but in an explicit representation of* $V_{pain}(s)$*, the output has at least two nodes in order to represent these values*

a binary classification. The complicated dopaminergic ties to pain were explored by the Neurological Pain Pathways section, who demonstrated that discretizing the pain and reward pathways by using entirely different critics is also not entirely biologically accurate. However, despite its shortcomings, any incorporation of a pain estimate is likely more accurate than none. Moreover, this approach, taken by Elfwing and Seymour (2017), as the basis of the MaxPain architecture, showed faster convergence and safer behaviour.

As shown, reward and motivation to escape is caused by painful stimuli. To mirror this type of aversion, one could use concepts like memory buffers, and algorithms may learn to associate the end of painful states with higher values, increasing the expected reward term as in TD learning.

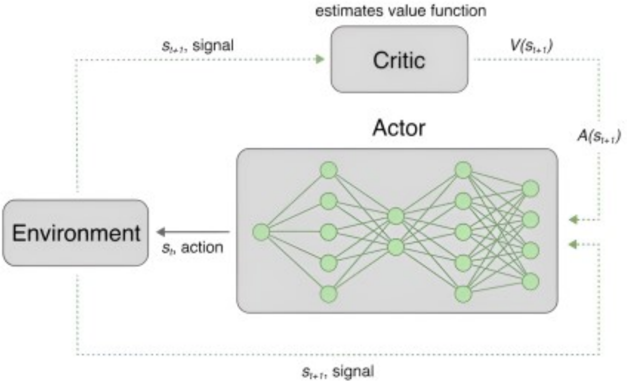


Figure 6. *Example Actor with* V $(s)_{\text{pain}}$ *and* V $(s)_{\text{reward}}$ *in Third Layer. Fig. 6 depicts an example wherein the critic only outputs one value as in the typical A2C paradigm (contrary to Fig. 5). Thus, the burden of representing* V $(s)_{\text{pain}}$ *is shifted onto the actor, which at some point should have a two-node layer as shown. There is no exact recommendation for in which layer this compression should occur, but in this example, it is depicted in the third layer.*

Valuing and Weighting Pain

In order to delve deeper into the MaxPain implementation of pain valuation in reinforcement learning. we will continue to refer to Seymour's (2019) investigation of pain. Seymour states that "it is clear that pain is constructed not only from nociceptive input, but also from a set of cortical and subcortical components that compute the effective magnitude of pain as a control signal" (Seymour, 2019, p. 1036). As discussed, MaxPain takes the latter into consideration through a fixed w (weighting) variable. The closer to 1 this w variable is, the less that pain is considered in the final Q_w result and vice versa. There is no change to how pain is weighted in the agent's "mental model" as it learns or based on state context. The hyperparameter w is set before training.

The figures in the MaxPain paper illustrate how modifying this weighting factor affects how the agent performs in the "dangerous grid world" task. In this task, crashing into the wall is considered "painful" and the frequency of that occurrence should be reduced. The authors tested various levels of pain aversion by setting w to 0.1, 0.5, and 0.9. The figures in the paper present the trade-off; models that are heavily pain-aversive will be more "careful", usually at the cost of solution efficiency (Elfwing & Seymour, 2017). In this case, fixing w and explicitly modeling different values made the results of the MaxPain algorithm more interpretable and clearer.

Alternatively, one way to modify the weighting factor is discussed in "Modular Deep Reinforcement Learning from Reward and Punishment for Robot Navigation" (Wang et al., 2021). The authors of the paper proposed a Boltzmann distribution-based selection mechanism for finding weighting factors that are applied to separate reward and pain optimization policies to find a joint policy. The weighting factor is dependent on a state-evaluation function $V(s)$ in an interesting way. There is a temperature variable τ_w that determines how evenly mixed the w^+ and w^- variables are. The way temperature and $V(s)$ affect the weighting factors are shown below:

if $\tau_w \to \infty$, $w^+(s) = w^-(s) = 0.5$ if $\tau_w \to 0$, $w^+(s) = 1$ and $w^-(s) = 0$ when $V^+(s) \geq V^-(s)$ $w^+(s) = 0$ and $w^-(s) = 1$ when $V^+(s) \leq V^-(s)$

The latter, where τ_w is zero, is called hard-max weighting. After experimentation, the authors concluded that "Deep MaxPain with hard-max weighting achieved the best overall performance" compared to fixed weights and standard DQN because it "utilized real-time assessments for weighting two sub-policies" (Wang et al., 2021, p. 125). However, pain and reward in the real world are not binaries to choose from when considering a policy to follow. There is always an influence of both future reward and pain when choosing actions in a state. As environments grow more complex, so should the considerations that affect the weighting of the policy.

We propose that a separate predictor network be used to provide the weighting variable at each timestep. Making the weighting variable dynamic and learned across timesteps may have several advantages. Firstly, it more accurately models the aforementioned "set of cortical and subcortical components that compute the effective magnitude of pain" (Seymour, 2019, p. 1036), since neurons are not fixed hyperparameters but instead dynamic and context-dependent mechanisms of learning. Moreover, it allows for more complex, higher-level behaviour where an agent has to decide if the "pain is worth the gain", depending on the context of the state.

As for the inputs of the weighting variable network, it could consider either the current state, a state memory buffer, time spent in the current training episode, time spent until the episode terminates, the level of "damage" the agent has already sustained, or any further possibilities and combinations of relevant information. The output would be a normalized scalar between 0 and 1 used to linearly combine Q_r and Q_p in the case of MaxPain, or two separate sub-policies in the case of Deep MaxPain. Regardless of this specific proposed implementation, allowing the weighting variable to be dynamic and context-dependent will lead to more balanced and adaptable behaviour in a MaxPain agent. Designing or Discovering Painful Stimuli.

Finally, in returning to general mechanisms of Deinforcement Learning, there remains the question of what is painful. Secondarily, how do we construct a state provided by the environment that might allow us to learn what is painful? This may be the most difficult component of Deinforcement Learning. For humans, this is initially partially encoded by genetics. We learn what is painful through evolutionary genetic iterations. Each iteration we approximate reward such that fitness increases. Similarly, pain can approximate behaviours to avoid such that fitness increases. Along with genetics, what is painful changes throughout a lifetime; it is learned through experiences as well as internally modulated through complex top-down modulatory pathways, beyond the scope of the neuroscience described in this paper.

The most literal machine learning analogue to the genetically encoded aspect of pain might be an actor which can reproduce or spawn new networks with its learned weights. Here we suggest applying genetic algorithms to the context of Deinforcement Learning agents, providing a method to allow networks to learn what should be considered painful. In such an implementation, the probability of reproduction in this context correlates with the problem that the network is trying to solve. The network also needs a reward heuristic, such as time alive or reproductive success. It learns as described above, making estimations of how painful or rewarding the environment might be. If the pain estimate is not correct, this negatively affects the network's predictions, making it less likely to reproduce. If an inaccurate pain estimation led to a very low performing model, the agent may be deactivated, or effectively killed. Thus, a successful, fit network should learn to define pain in a similar manner to humans—that which should be avoided for the sake of reproducing the network. These reproductive odds defined within the genetic algorithm provide a separate signal to learn from besides the immediate-term reward signal and potential pain signal.

Concretely, if the network should learn to make a stick figure walk like in OpenAI's MuJoCo framework, then reproductive success may be set as a function of time spent walking versus energy expended. Networks that avoid fatal falls or expend less energy in their movement should have a higher probability of reproducing or replicating their weights in new networks. Here, an initial sensory layer as discussed in the "Representing Pain" section would group together similar states and outcomes based on their features, modulate the given state of the environment to hold this painful information, then finally pass it to a critic. The critic estimates the value of the state with respect to potential pain and reward. Alternatively, two critics could be used similar to the MaxPain architecture. Then, if the agent were near a box that it might trip over, the current sensed pain may be 0, but the critic may weigh future states as very painful and $V_{pain}(s)$ very low. Its estimation $V_{pain}(s)$ would be tuned as it attempts to walk, and further tuned as it replicates its weights in other agents based on its walking success.

One may not find a need for literal analogues such as genetic algorithms. In the above case, one implicitly defines pain merely by defining what success is. In the MuJoCo example we did this—pain was implicitly that which must be avoided to achieve success of walking, for example, tripping. However, a model can receive a single signal (like reward in the current paradigm) and learn pain aversive behaviour. Crucially this must, at some point in processing, interpret the signal with respect to pain and pleasure, which together update the model's policy in dissimilar ways, just as the proposed algorithms in Figure 4 describe, or the algorithms discussed in "Existing Literature and MaxPain".

Those familiar with RL may now be wondering: initial clustering algorithms, modified critics, and learned hyperparameters—are these really a significant change from standard RL practices? The key difference is not only in these simple, fundamental algorithmic changes but in coupling them with environmental changes conducive to pain aversive learning. As initially described, without an environment where correctness or incorrectness is non-uniformly distributed, pain is not a useful concept. This means that to test the incorporation of pain processing with something as simple as an MNIST classifier, one would need to quantify how close each classification was to the correct classification and build that into the reward or ground truth. Ultimately, there are many possible scenarios where pain would increase convergence or safety. Anywhere where particular states must be avoided more than a typical "failure" is a good application of Deinforcement Learning.

Conclusion

We have seen that reinforcement learning and neuroscience are intricately intertwined, beginning with their overlapping uses of reward prediction error and TD learning. Continuing to draw inspiration from the brain and body to enhance modern RL algorithms is a fruitful frontier. Many areas of how humans learn from pain are yet to be investigated; the role of emotional pain and trauma was not examined in this paper. However, the growing neuroscientific body of knowledge on pain allows us to examine the phenomena as a model for novel RL algorithms. Thus, we conclude new RL models which learn pain aversive behaviour are necessary to propel the field towards more realistic, safe, efficient learning paradigms.

Acknowledgments

This work was created by the Interactive Intelligence team, a research group at the University of Washington fusing neuroscience and computer science to make machines learn like humans. We would like to thank the I2 advisor, Dr. Eric Chudler, for his support and guidance.

Thank you to everyone on the team, and everyone on and off it who helped make this paper and team possible.

References

Arias-Carrión, O., Caraza-Santiago, X., Salgado-Licona, S. et al. (2014). Orquestic regulation of neurotransmitters on reward seeking behavior. *International Archives of Medicine, 7,* 29. https://doi.org/10.1186/1755-7682-7-29

Bayer, H. M., Lau, B., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology, 98*(3), 1428–1439.

Beeler, J. A., Daw, N., Frazier, C. R., & Zhuang, X. (2010). Tonic dopamine modulates exploitation of reward learning. *Frontiers in Behavioral Neuroscience, 4,* 170. https://doi.org/10.3389/fnbeh.2010.00170

Bushnell, M., Ceko, M., & Low, L. (2013). Cognitive and emotional control of pain and its disruption inˇ chronic pain. *Nature Reviews Neuroscience, 14.* https://doi.org/10.1038/nrn3516

DosSantos, M. F., Moura, B. S. & DaSilva A. F. (2017). Reward circuitry plasticity in pain perception and modulation. *Frontiers in Pharmacology,* http://dx.doi.org/10.3389/fphar.2017.00790

Edlar, E, Hauser T. U., Dayan P., and Dolan R. J. (2016). "Striatal structure and function predict individual biases in learning to avoid pain," *Proceedings of the National Academy of Sciences, vol. 113, no. 17,* pp. 4812–4817.

Elfwing, S., Seymour, B. (2017). Parallel reward and punishment control in humans and robots: Safe reinforcement learning using the MaxPain algorithm. *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob),* 140-147, https://doi.org/10.1109/DEVLRN.2017.8329799

Fernando A., Urcelay G., Mar, A., Dickinson, A. & Robbins, T. (2013). Comparison of the conditioned reinforcing properties of a safety signal and appetitive stimulus: Effects of d-amphetamine and anxiolytics, *Psychopharmacology, 227*(2) 195–208

Geron, A. (2019). Hands-on machine learning with scikit-learn, keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly Media.

Gershman, S.J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic Bulletin & Review, 22,* 1320–1327. https://doi.org/10.3758/s13423-014-0790-3

Ghiassian, S., Patterson, A., Garg, S., Gupta, D., White, A., & White, M. (2020). Gradient temporal-difference learning with regularized corrections. *ICML,* https://doi.org/10.48550/arXiv.2007.00611

Grison, S., & Gazzaniga, M. S. (2019). *Psychology in Your Life (3rd ed.).* W.W. Norton.

Hauser, T. U., Eldar, E., & Dolan R. J. (2016). Neural mechanisms of harm-avoidance learning: A model for obsessive-compulsive disorder?" *JAMA psychiatry, 73*(11), 1196–1197.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks, 15,* 535-547.

Kim H., Shimojo S., & O'Doherty J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain, *PLoS Biology, 4*(8), 233.

Kubanek, J., Snyder, L. H., & Abrams, R. A. (2015). Reward and punishment act as distinct factors in guiding behavior. *Cognition, 139,* 154–167. https://doi.org/10.1016/j.cognition.2015.03.005

Lin, B., Cecchi, G. A., Bouneffouf, D., Reinen, J. M., & Rish, I. (2019). A story of two streams: Reinforcement learning models from human behavior and neuropsychiatry. *Adaptive Agents and Multi-Agent Systems.*

Markovic, T., Pedersen, C. E., Massaly, N., Vachez, Y. M., Ruyle, B., Murphy, C. A., Abiraman, K., Shin, J. H., Garcia, J. J., Yoon, H. J., Alvarez, V. A., Bruchas, M. R., Creed, M. C., & Morón, J. A. (2021). Pain induces adaptations in ventral tegmental area dopamine neurons to drive anhedonia-like behavior. *Nature Neuroscience, 24*(11), 1601–1613. https://doi.org/10.1038/s41593-021-00924-3

Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature, 379,* 449–451.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *Proceedings of The 33rd International Conference on Machine Learning, in Proceedings of Machine Learning Research, 48,* 1928–1937. https://proceedings.mlr.press/v48/mniha16.html

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology, 53*(3) 139–154.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38*(2), 329–337.

Perelman School of Medicine at the University of Pennsylvania. (2014, May 13). Human learning altered by electrical stimulation of dopamine neurons. ScienceDaily. Retrieved April 15, 2023 from www.sciencedaily.com/releases/2014/05/140513175006.htm

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442*(7106), 1042–1045. doi:10.1038/nature05051

Russo, S., Nestler, E. (2013). The brain reward circuitry in mood disorders. *Nature Review Neuroscience, 14,* 609–625.

https://doi.org/10.1038/nrn3381

Seymour, B. (2019). Pain: A precision signal for reinforcement learning and control. *Neuron, 101*(6), 1029–1041. https://doi.org/10.1016/j.neuron.2019.01.055

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.), 275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Steel, A. (2016). The impact of reward and punishment on skill learning depends on task demands. *Nature,* https://www.nature.com/articles/srep36056

Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., Casas, D. de L., Budden, D., Abdolmaleki, A., Merel, J., Lefrancq, A., Lillicrap, T., & Riedmiller, M. (2018). DeepMind Control Suite (Version 1). *arXiv,* https://doi.org/10.48550/ARXIV.1801.00690

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems,* 5998–6008.

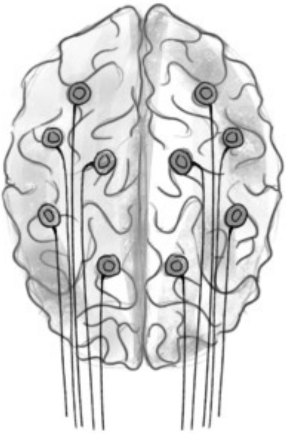
Wang. J., Elfwing, S., Uchibe, E. (2021). Modular deep reinforcement learning from reward and punishment for robot navigation. *Neural Netw, 135*(115–126). https://doi.org/10.1016/j.neunet.2020.12.001

Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual review of neuroscience, 40,* 373–394. https://doi.org/10.1146/annurev-neuro-072116-031109

Wood P. B. (2006). Mesolimbic dopaminergic mechanisms and pain control. *Pain, 120*(3), 230–234. https://doi.org/10.1016/j.pain.2005.12.014

Yam, M. F., Loh, Y. C., Tan, C. S., Khadijah Adam, S., Abdul Manan, N., & Basir, R. (2018). General pathways of pain sensation and the major neurotransmitters involved in pain regulation. *International journal of molecular sciences, 19*(8), 2164. https://doi.org/10.3390/ijms19082164

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of the dorsolateral striatum pre- serve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience, 19,* 181-189.



Towards real-time classification of EEG motor imagery with deep learning

Aleksandrs Baskakovs

Aarhus University

Department of Linguistics, Cognitive Science and Semiotics

Luke Ring

Aarhus University

Department of Linguistics, Cognitive Science and Semiotics

Abstract

Brain-Computer Interfaces (BCIs) have been widely employed to identify users' intention to control external objects by decoding motor imagery (MI) from an electroencephalogram (EEG). In recent years, the contribution of deep learning (DL) has had a phenomenal impact on MI-EEG-based BCI. Specifically, deep learning is highly attractive for MI-BCI as it requires little to no preprocessing, which results in a significant decrease in latency between a patient's intention and the execution of the command by the device, be it a prosthetic or a cursor. This study investigates the feasibility of using low-cost dry-electrode EEG recording to capture motor imagery for training neural networks to classify imagined right-hand fist clenches vs resting conditions and subsequent real-time online inference. This holds importance for many kinds of brain-computer interfaces, especially for people with impaired movement. The online aspect is optimized to minimize latency with a hard limit of 1 second from capture to classification. A complete end-to-end pipeline is provided, and although high levels of classification accuracy were not achieved, the framework sets up a clear path to implement rapid inference on consumer devices and suggests several future avenues to improve the quality and accuracy of results.

Introduction

Brain-Computer Interfaces

Brain-computer interfaces (BCIs) translate users' intentions via brain activity into external device commands used for interaction with the environment, applications, hardware devices, and prosthetics. The methods for measuring brain activity are usually divided into two categories: invasive and non-invasive. Invasive methods include electrocorticography (ECoG), where electrical potentials are recorded on the surface of the brain underneath the skull and microelectrode recordings, where electrodes thinner than the width of the human hair are inserted into the brain for the recording and stimulation of deeper brain structures (Schalk & Leuthardt, 2011). Brain activity can be measured non-invasively via electrical potentials on the scalp using electroencephalography (EEG) (Abiri et al., 2019), via magnetic field changes using magnetoencephalography (MEG) (Paek et al., 2020), and via metabolic processes related to brain function using functional magnetic resonance imaging (fMRI) (Sitaram et al., 2007), positron emission tomography (PET) (Nutt, 2002), and functional near-infrared spectroscopy (fNIRS) (Schalk, 2009).

In addition to degrees of invasiveness, the described sensor modalities can be distinguished by their temporal and spatial resolution. Temporal resolution is the capacity of a method to detect changes in activity over time, whereas spatial resolution is the degree to which a given method can detect changes in activity by location (Friston, 2009). In the context of BCIs, temporal resolution refers to how precisely a method can detect exactly when an event in a brain occurred, and spatial resolution refers to the precision of a method for detecting exactly where an event in a brain occurred. Additionally, the methods can be compared with respect to the degree of mobility they afford the subject. Mobility is highly relevant in the case of brain-computer interfaces, as the usability of real-world deployments of BCI applications would heavily rely on the ability of the user to move their body freely. Table 1 illustrates how the non-invasive neuroimaging methods (EEG, MEG, fMRI, PET, and fNIRS) compare across the three dimensions.

Table 1:

Method	Temporal resolution	Spatial Resolution	Degree of mobility
EEG	High	Low	Medium
fMRI	Low	High	Low
fNIRS	Low	Medium	High
MEG	High	Medium	Low
OPM-MEG	High	Medium	High
PET	Low	High	Low

Note. *Neuroimaging methods compared against the dimensions of spatial and temporal resolution, and the degree of mobility.*

Examples of BCI implementations include, but are not limited to: brain-to-text communication via intracortical electrode recordings which decode handwriting movement attempts from a paralyzed individual into text (Willett et al., 2021), detection of drowsiness in drivers using fNIRS (Khan & Hong, 2015), and control of various external devices such as wheelchairs (Rebsamen et al., 2007), drones (Christensen et al., 2019), and robotic limbs (Lebedev & Nicolelis, 2017) using EEG.

Currently, the most practical use of BCI systems is that of enabling communication and control of devices for individuals living with a functional impairment. However, mass adoption of such technology is hindered by the difficulty of obtaining high-quality recordings in environments outside of a lab, the processing speeds of consumer devices, and the complexity involved in building solutions that are sufficiently generalizable to be adapted to individuals or do not require expert understanding to set up (McFarland & Vaughan, 2016).

EEG-based BCIs

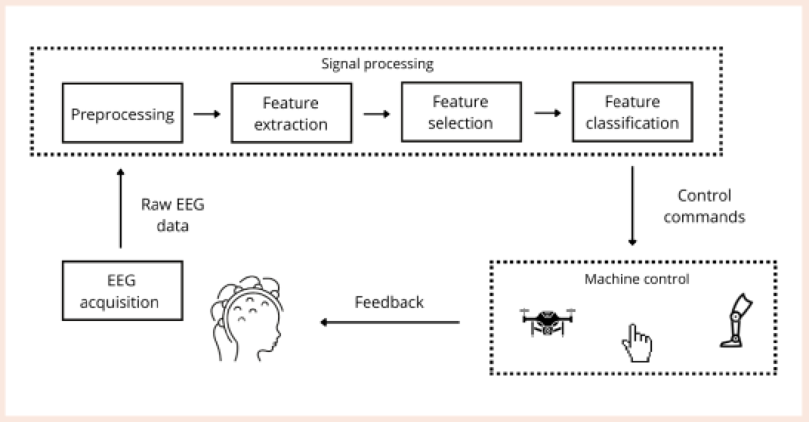
EEG is often considered the most practical choice for BCI applications due to its low cost, high portability, and high temporal resolution (Abiri et al., 2019). EEG measures neural activity using electrodes placed on a scalp via the flow of electric currents caused by synaptic excitations of neuronal dendrites (Olejniczak, 2006). It is often said that this is a "direct" measurement of brain activity, as opposed to fMRI or fNIRS which measure the activity via metabolic processes in the brain. EEG does not pick up the activity of individual neurons, but rather a synchronous activity of millions of neurons that have similar spatial locations. The type of neurons that produce the most EEG signals are pyramidal neurons, due to their alignment and tendency to fire together (Kirschstein & Köhling, 2009). The inability to pick up signals from individual or smaller numbers of neurons is one of the reasons for the weaker spatial resolution of EEG. Another reason for poor spatial resolution is that EEG records activity from the outer layer of the cerebrum, the cerebral cortex, and thus does not have access to deeper brain structures. Signals captured by EEG only account for around 5% of brain activity (Ball et al., 2009). However, given that EEG is able to directly sample regional brain activity at rates of 1 000 Hz or greater, it provides a high level of temporal resolution. While activity in the deeper brain structures is beyond the reach of an EEG device, the signals in the areas such as the primary motor cortex, primary sensory cortex, temporal lobe, or occipital lobe can be reliably detected and utilized for control of external devices (Guger et al., 2000).

The use of signals found in the cortical brain areas are often referred to as "paradigms" (Abiri et al., 2019), with each paradigm having an established protocol for the development of a BCI application. The EEG is first collected while the subject repetitively performs a specific task related to the paradigm being investigated (e.g., visual or imagery task). Typically, data is then preprocessed to improve the signal-to-noise ratio and remove invalid trials, and is also used for training and validation of the decoder. Finally, the subject can use the BCI system by performing the task again, where the BCI system translates the neural signal into commands for virtual objects or external prosthetics (Abiri et al., 2019). Figure 1 illustrates a typical EEG BCI protocol. The most common EEG-BCI paradigms are motor imagery (MI) paradigms and external stimulation paradigms.

EEG-BCI paradigms

The external stimulation paradigm is based around the purposeful

Figure 1:



Note. *A block diagram for a general BCI system* modulation of brain activity using outside signals - visual, auditory (Shangkai Gao et al., 2014), or somatosensory (Yao et al., 2017). For example, one of the more widely implemented BCI-enabled communication interfaces is a P300 Speller. P300 is an event-related potential, evoked in response to an external stimulus. P300 is thought to reflect the contextual meaningfulness of the stimulus for the subject. It is characterized by a large positive deflection peaking at approximately 300 ms after the presentation of the stimulus (Fabiani et al., 1987). Farwell and Donchin (1988) were the first to describe a reliable BCI communication system utilizing the P300 signal, called P300 Speller. The system presents a 6x6 grid with a character occupying each cell. The subject focuses their attention on the character they intend to spell, while each row and column of the matrix is highlighted in a random sequence (Figure 1.1). Whenever the highlighted row and column intersect at the subject's character of choice, a P300 response is elicited. A decoder is then able to spot the signal, recognize at which intersected character it was elicited and infer the subjects' character of choice. One patient with late-stage ALS has used the device for 4 to 6 hours a day for typing emails and other applications and has reported the P300 speller as superior to his eye-gaze-based system (Vaughan et al., 2006).

Another paradigm that makes use of external stimuli for the selection of commands is called a steady-state visual evoked potential or SSVEP (Vialatte et al., 2010). In this paradigm, multiple stimuli (each of which represents a command and can be visual, auditory, or somatosensory) are presented to the participant, each at a different frequency, with frequencies varying from low (1–5 Hz) to high (75–100 Hz). When a subject focuses their attention on a particular stimulus (e.g., a flickering light) EEG frequency observed over the visual areas of the brain is reliably correlated with the frequency of the re-

Figure 1.1:

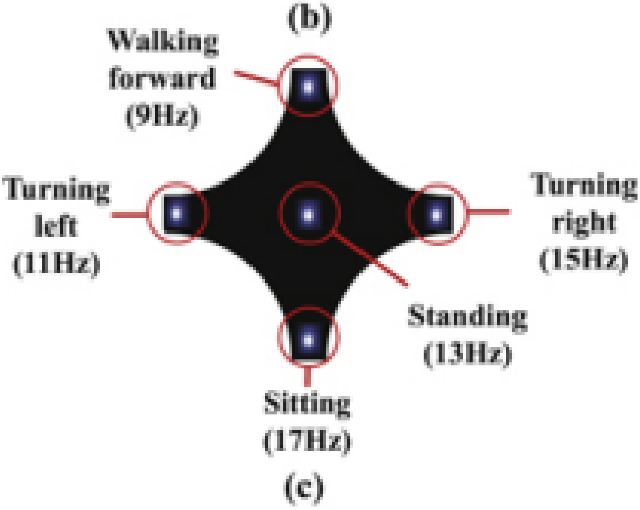
DICE (D)

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	_

Note. *The letter grid used for a typical P300 speller-based BCI. A row or column flashes for around 100 ms every 200 ms. The participant focuses his attention on one of the letters (for example, letter "D") whilst wearing an EEG cap. Whenever the row or column containing flashes (in this case, row 1 or column 4), a P300 response is elicited. The first row or the fourth column must flash multiple times (eliciting P300 response multiple times) in order for the decoder to confidently classify "D" as the participant's letter of choice. (Figure reprinted from Krusienski et al. (2008)).*

spective stimulus (Kuś et al., 2013) (Figure 1.2). Decoding which command the subject is trying to select becomes possible by matching the EEG activity pattern to the frequency of the stimulus representing the command. SSVEP is advantageous because it does not require training, can be classified more reliably than event-related potentials, and benefits from more commands due to the range of frequencies for use (Abiri et al., 2019). Applications include controlling lower limb exoskeleton (Kwak et al., 2015), orthosis (Pfurtscheller et al., 2010), and locked-in syndrome (Hwang et al., 2017).

Figure 1.2:



Note. *An example of a visual SSVEP interface panel used for the control of the exoskeleton from Kwak et al. (2015). Each of the white dots is an LED flickering at a specific frequency and represents a particular command. The subject focuses his attention on one of the lights while wearing an EEG cap, the activity recorded in the visual cortex is then correlated with the frequency of each of the lights. The command whose respective flickering frequency correlates the most with the activity is then classified as being the intent of the subject.*

Since external stimulation paradigms heavily rely on constant and repetitive presentation of stimuli to the subject, subjects may experience fatigue and may generally find their use difficult for long periods of time (Chang et al., 2014).

Motor imagery paradigms, on the other hand, do not rely on external stimuli, but rather on the wilful, internal modulation of neural activity. Using MI provides a unique way of interacting with hardware or software that has the potential to feel natural and relevant to the task being performed as well as being distinct from natural environmental stimuli (such as noises, flickering lights, etc.).

Motor Imagery

Motor imagery is a mental process of kinesthetically imagining movements without the respective physical movement occurring (Lotze & Halsband, 2006). One way to discern MI from visual imagery is to determine the “point of view” of the subject of the imaginal experience. On the one hand, mental images can be experienced from a third-person perspective, with the subject acting as a “spectator”, watching themselves perform an action. One illustrative example is the sport of climbing, where it is commonplace for individuals to imagine an avatar ascending a specific route. On the other hand, mental images can be experienced from a first-person perspective, a process involving mostly a kinesthetic representation of the movement, during which the subject feels as if they were executing the movement. This type of imagery requires representing the process of an action being executed (Jeannerod, 1995). Kinesthetic motor imagery can be experienced by remembering and/or preparing for a set of precise movements, for example a drummer preparing to play a specific, highly technical part of a musical piece. Motor imagery falls under the type of kinesthetic imagery (Stinear et al., 2006).

Motor imagery gained its prominence as a research area due to its relevance in motor learning—the process of acquiring or strengthening a skill through repetitive practice (Newell, 1991). During repetitive practice, the subject engages neural networks responsible for the movement over and over again, strengthening the connections between the neurons through the process called Hebbian learning (Hebb, 1949), or as its famous tenet goes: “Neurons that fire together—wire together.” Hebbian learning in the brain as a physical process is primarily facilitated by two neural mechanisms: long-term potentiation (LTP) (Bliss & Lømo, 1973) and long-term depression (LTD) (Artola et al., 1990). LTP and LTD influence the extent to which activity in a sending neuron leads to activation of a receiving neuron, by influencing the efficacy of synapses or junctions between neurons. LTP is a long-lasting potentiation (strengthening) of synaptic efficacy, while LTD is a long-lasting depression (weaking) of synaptic efficacy.

Whilst physical practice is most vital for the acquisition and consolidation of new motor skills (Robertson et al., 2004), motor imagery is a well-assessed complementary practice for motor learning (Schuster et al., 2011). The rationale is that imaginary movement potentiates the activation of the sensorimotor system which leads to strengthening of neuronal connections (Kraeutner et al., 2014; Pfurtscheller & Neuper, 1997). MI is used for motor learning in both healthy populations (Dickstein & Deutsch, 2007) and for motor rehabilitation in patients (Malouin et al., 2013). This willful generation of activity in the sensorimotor cortex by performing imaginary movements is the foundation of the motor imagery paradigm in BCIs.

In the sensorimotor rhythms (SMR) paradigm, a subject imagines kinesthetic movements of body parts such as the hands, the legs, the fingers, the feet, the legs, the arms, or the tongue. Imagined movement causes event-related desynchronization (ERD) and its opposite, event-related synchronization (ERS), which is observed during relaxation. ERS and ERD phenomena are found mostly in Alpha (or mu, as it is often referred to in the context of imaging sensorimotor cortex) (8–12 Hz) and Beta (16–24 Hz) frequency bands (Pfurtscheller, 2000, p. 26). Upon the imagination of a movement, a power reduction (ERS) can be observed in the mu/beta , and a rise back in power (ERS) when the movement ceases. As with physical movements, the imaginary movements in the motor imagery paradigm are unilateral, due the location and size of neural regions responsible for specific movements differing, which leads to different signal characteristics, thus making it possible to distinguish MI for left- and right-hand movements. When a movement of a unilateral limb is imagined, the recorded activity at the contralateral side of the motor-sensory cortex increases, while the activity at the ipsilateral side does not display an increase in activity. This contrast between location-specific signals as well as contextual EEG activity is able to be leveraged for successful classification of the user's intent.

BCI systems within the motor imagery paradigm make use of ERD/ERS events for an external application. Apart from allowing individuals with impaired function to control, for example, external limbs (Barsotti et al., 2015) or moving a wheelchair (Reshmi & Amal, 2013), MI-based BCIs are also often used for the purposes of neurorehabilitation (Pichiorri & Mattia, 2020).

Classification Approaches

Traditional signal processing approaches

EEG has a low signal-to-noise ratio due to the electrodes measuring brain activity at a microvolt level. This high sensitivity means that along with cortical brain activity, EEG recordings capture many biological processes including eye blinks, heartbeat, muscle movements, and respiration. The electrodes are also prone to interference caused by electronic equipment, including the recording equipment itself and proximity to electromagnetic fields like those generated electricity supply lines. In addition, there are cross-channel correlations and subject-specific patterns of activity (Altaheri et al., 2021). Hence, adequate classification of imagery movements strongly depends on the processing pipeline of raw EEG data, namely: preprocessing, feature extraction, feature selection, and feature classification (Khosla et al., 2020).

Usual preprocessing techniques include: notch (bandstop) filtering to remove power line noise at 50 or 60 Hz (region specific); high-pass filtering for the removal of baseline drift; low-pass filtering to smooth the signal; downsampling the data for quicker computation and reduced memory storage; selection of specific electrodes dependent on the performed task; referencing from specific electrodes or the signal average from all electrodes; and band-pass filtering to select frequency range(s) of interest.

Following preprocessing, EEG data go through feature extraction, feature selection and finally, feature classification where a prediction of the intended movement occurs. Feature extraction is a stage where meaningful information is extracted from the neural data. It is often achieved using time-frequency approaches—due to the non-stationary nature of EEG signals—and further improved via spatial approaches that identify and weight channels with the highest signal-to-noise ratio. The returned feature sets are of high dimensionality, and statistical techniques such as Principal Component Analysis (PCA) (Abdi & Williams, 2010) and Independent Component Analysis (ICA) (Stone, 2002) are used for dimensionality reduction and feature selection.

Often, multiple techniques are used for each phase of extraction, selection, and classification of EEG signals (Kevric & Subasi, 2017). While high-performing, if these techniques were employed for applications requiring time-critical classification, the recurring complex computations performed on the data would introduce prohibitively high latency. Additionally, they would require expert assistance to calibrate the decoder, as MI EEG data are prone to high subject specific variance (Zhang et al., 2021), which also contributes to the problem of achieving a generalizable or, at least, easy to calibrate classifier for MI EEG signals. Thus, the classical signal processing approaches are an obstacle for achieving an easily deployable BCI system that is able to perform accurate predictions of users intent in real-time.

Machine learning

Neural networks have been shown to achieve "end-to-end learning" by learning complicated and latent features from large amounts of data (Bojarski et al., 2016), bypassing the need for manual feature extraction, selection, and preprocessing. The promise of feeding raw data directly into the neural network without intensive signal preprocessing opens up both the latency and calibration bottlenecks for real-time BCI use (Craik et al., 2019). Specifically, convolutional neural networks (CNNs) have become very popular due in part to their success in image classification (Krizhevsky et al., 2012). Naturally, the high dimensionality EEG data has been tackled with deep learning, notably for previously mentioned visual-evoked responses (Cecotti & Graser, 2010), Alzheimer's classification (Morabito et al., 2016), depression (Acharya et al., 2018), epilepsy prediction (Hussein et al., 2019), and most relevantly, motor imagery classification (An et al., 2014; Tabar & Halici, 2017).

Following the promise of end-to-end learning, deep learning for motor imagery has experienced a rapid growth since 2017 (Altaheri et al., 2021). Since then, multiple approaches regarding preprocessing (or lack of it), input formulation, deep learning architectures, and performance evaluation have been tried. Preprocessing for MI EEG classification usually consists of selecting channels that contain the most distinct MI features and subsequent band-pass filtering. More than 79% of the studies reviewed by Altaheri et al. (2021), used all EEG channels. However, most of these studies were not oriented on online classification, where reduction of the volume of incoming data, process complexity, and computational time is a significant consideration. Over 91% of the studies reviewed in the aforementioned paper used a band-pass filter on the data to select for frequencies where the ERD/ERS events are observed, notably the mu (8–12 Hz) and Beta (18–26 Hz) frequency bands. The band-pass filter also effectively removes the power line noise (50 Hz) from the data.

Input formulation largely depends on the architecture of the neural network and is usually one of the following four: raw signal values, spectral images, topological maps, or extracted features. The neural network is then trained on this input and judged by its ability to correctly discern the class of the input data, which could be a binary outcome or a multi-class prediction. Classification performance of a model is usually validated by using either the hold-out approach—where a portion of the data that is not included in training is used to validate predictions—or the cross-validation approach which repeatedly splits the data in different ways for training and validation. The exact method of splitting and validating data varies by the type of cross-validation. Trained models are then selected based on a specific performance metric such as maximizing accuracy or minimizing the outcome of a loss function. The selected model can then be used on novel data for inference, meaning it is able to predict the class of data outside the original dataset. Machine learning architectures arepromising contenders as a method for effective and process-efficient classification of motor imagery signals.

Present study

Whilst many have reported high-performing MI-EEG BCIs, these systems are confined to carefully controlled, noise-free, artificial environments, and are heavily dependent on expensive research-appropriate systems that require expert preparation and configuration (McCrimmon et al., 2016). Moreover, the classification is usually applied on the complete period of performing the motor imagery movement during the experimental protocol, which ranges around four seconds. In a real-time scenario, this would result in a minimum latency of four seconds for the user, seriously inhibiting the usability of the BCI system for general tasks and making time-sensitive interactions impossible. For MI-EEG BCIs to reach a sufficient degree of practicality for everyday use, they need to be easy to set up, low-cost, robust, and most importantly, low-latency.

The main contribution of this study is the use and subsequent examination of a lightweight convolutional neural network for the real-time classification of EEG data with sub-second latency. Moreover, the EEG data is acquired in a naturalistic, interference-prone setting, using a low-cost, dry-electrode EEG device, which sheds further light on the accuracy and reliability of the BCI system for real-world applications.

Methods

Participants

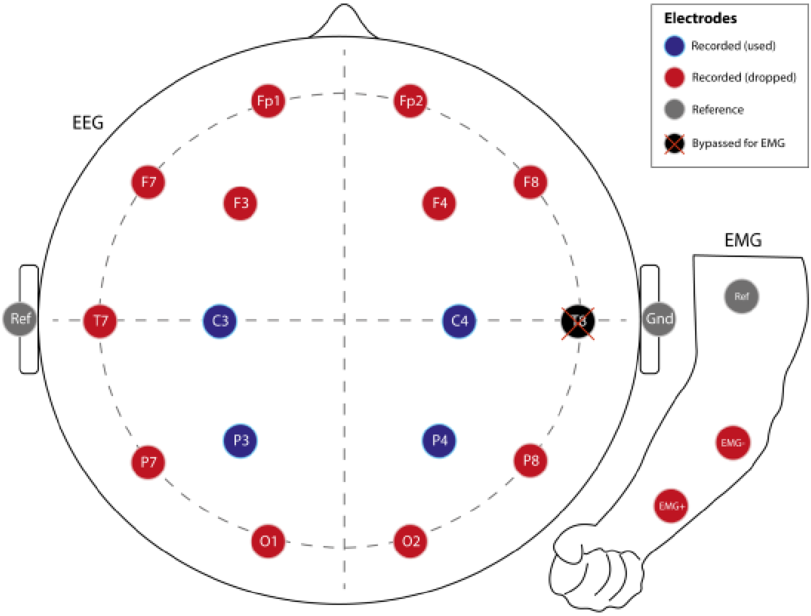
Two male volunteers (both right-handed, M = 30.0, SD = 5.0 years) participated in the experiment and gave their informed consent.

Hardware and Software

Neural activity was recorded using a low-cost, open-source/open-hardware EEG device Ultracortex Mark IV Headset[1](OpenBCI, Inc., USA) with dry electrodes, connected to the OpenBCI Cyton and Daisy biosensing boards (OpenBCI, Inc., USA). The cap followed a 10–20 international electrode placement scheme. Fifteen electrodes covering the whole head were selected (Fp1, Fp2, F3, F4, F7, F8, T7, C3, C4, P3, P4, P7, P8, O1, O2). EMG monitoring was performed using three Skintact F301 foam solid gel electrodes (two signal electrodes and one ground) plugged into the OpenBCI Daisy board. EEG Reference and ground electrodes were placed at the left and right ear lobes. Figure 2 illustrates the placement of both EEG and EMG electrodes.

Offline Raw EEG and EMG data were acquired at a sample rate of 1000 Hz via a microSD-card plugged directly into the Cyton board using a custom-built Python interface[2] using the BrainFlow API[3]. The custom interface was required to send low level commands to the board and configure pin settings for the EMG signal and grounding and in addition, a modified Cyton firmware[4] was required to fix an issue with capturing data above 250 Hz.

Figure 2:



Note. *Graphic representation of the EEG and EMG electrode placement on the scalp and arm, as well as their use in classification.*

---

[1] https://docs.openbci.com/AddOns/Headwear/MarkIV/

[2] https://github.com/zeyus/OpenBCI_Cyton_Library

[3] https://github.com/brainflow-dev/brainflow

[4] https://github.com/zeyus/BrainflowCytonEEGWrapper

[5] https://github.com/zeyus/CogNeuroExam

[6] https://github.com/onnx/onnx

Source code for the experiment, data collection, offline analysis, model training, and online classification is available on GitHub.[5]

Online classification with sub-second latency

Model training took place on an NVIDIA GTX1070 GPU with CUDA 11.6, CUDNN v8.4.0.27, and PyTorch 1.12.0.dev20220507+cu116. After model training, the model's weights were frozen and exported in Open Neural Network Exchange (ONNX) format6, subsequently the pre-trained model was loaded into a CPU-bound instance of the ONNX Runtime (ONNX Runtime developers, 2021) and real-time streaming EEG and EMG data were collected via the Cyton USB/Bluetooth dongle using the custom interface at a sample rate of 250 Hz. Data from non-selected channels were dropped, and the remaining data were resampled and sent to the model for inference Figure 3.

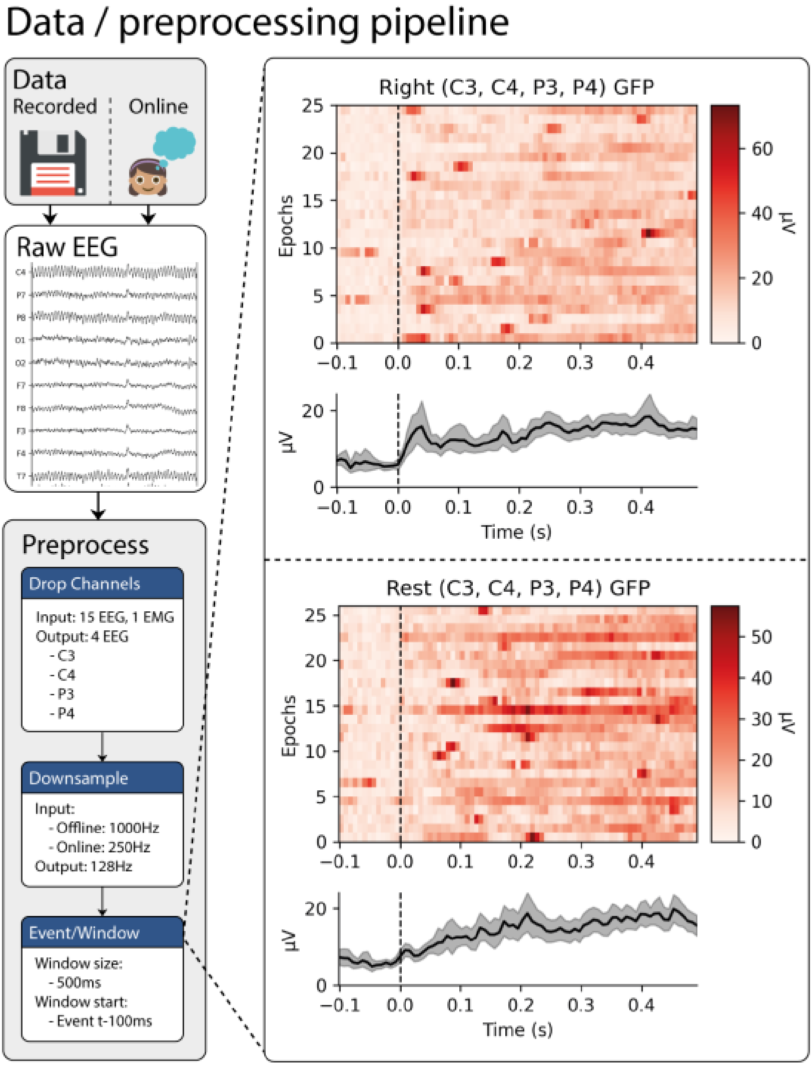
Stimuli

The python package PsychoPy (Peirce et al., 2019) was used to present the stimuli to the participants. The task was to perform imaginary movements of clenching the right hand or resting. Visual cues were presented to the participant before the beginning of a new trial to ensure attention to the task. A purple arrow pointing to the right was used as a visual cue to indicate that the participant should imagine clenching their right hand, a purple arrow pointing upwards was used as a visual cue to indicate that the participant should rest, simply focusing their gaze on the arrow.

Trials and Experimental Protocol

Experimental protocol was largely adapted from Ma et al. (2020). Each trial of the experiment began with a white circle (t = -3 s) indicating the start of the trial, followed by a red circle 2 seconds later (t

Figure 3:



Note. *Outline of data acquisition and preprocessing pipeline.*

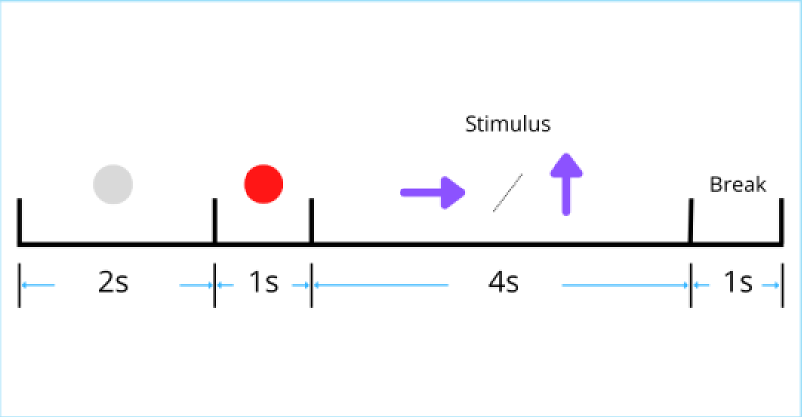
= -1 s), alerting the user about the upcoming stimuli onset. At t = 0 s, either the right or upward pointing arrow was presented indicating the requirement to either clench the right hand or to rest, for the duration of 4 seconds. The participants were told to perform the imaginary movement until the arrow disappeared from the screen. The importance of imagining single, continuous movement for the total duration of 4 seconds was stressed to the participants. At the end of the trial, a short break of 1 second followed. A schematic representation of the experimental protocol can be seen in Figure 4. The session lasted until 30 trials of each of the 3 tasks were completed, which was approximately 12 minutes. 3 sessions for Subject #1 and 4 sessions for Subject #2 were recorded over multiple days.

Procedure

The participants were seated in a comfortable chair and instructed about the protocol. The importance of performing a kinesthetic imaginary movement was stressed. To avoid motion-related artifacts, the participants were asked to keep any type of movements while going through the experiment to a minimum.

Following the instructions, the EMG sensors were placed on the participants arm, namely two sensors on the inner left side of the right forearm (flexor digitorum superficialis) and one ground sensor on the right biceps (biceps brachii). The EMG electrode locations were gelled before the placement of the sensors. The dry-electrode

Figure 4:



Note. *Schematic representation of the experimental protocol.*

EEG cap was placed on the participants head with the Cz electrode socket placed on the vertex of the head. Clip electrodes were used for the reference and ground electrodes, placed on ear lobes, which were cleaned with alcohol and gelled. A visual inspection of the quality of the EEG signals was done using the OpenBCI GUI. If any of the channels were shown to capture the signal poorly, the electrodes corresponding to those channels were gently turned until the quality of the signal was acceptable. Throughout the whole mounting procedure, the participants were repetitively asked whether they experienced any significant discomfort in relation to the procedure.

Data Pre-Processing

Assessment of muscle activity during trials

EMG activity was recorded along with EEG to monitor the amount of activity exhibited by the flexor digitorum superficialis muscle which is used for fist-clenching. Four prior experimental sessions (two per subject) were conducted which contained a physical fist-clench condition to compare the EMG signal of intentional physical movements with potential sub-activation levels of imagined movements. Visual inspection of EMG data from MI trials indicated that muscle activity was much weaker in imaginary movement trials (Figure 5).

Channel selection

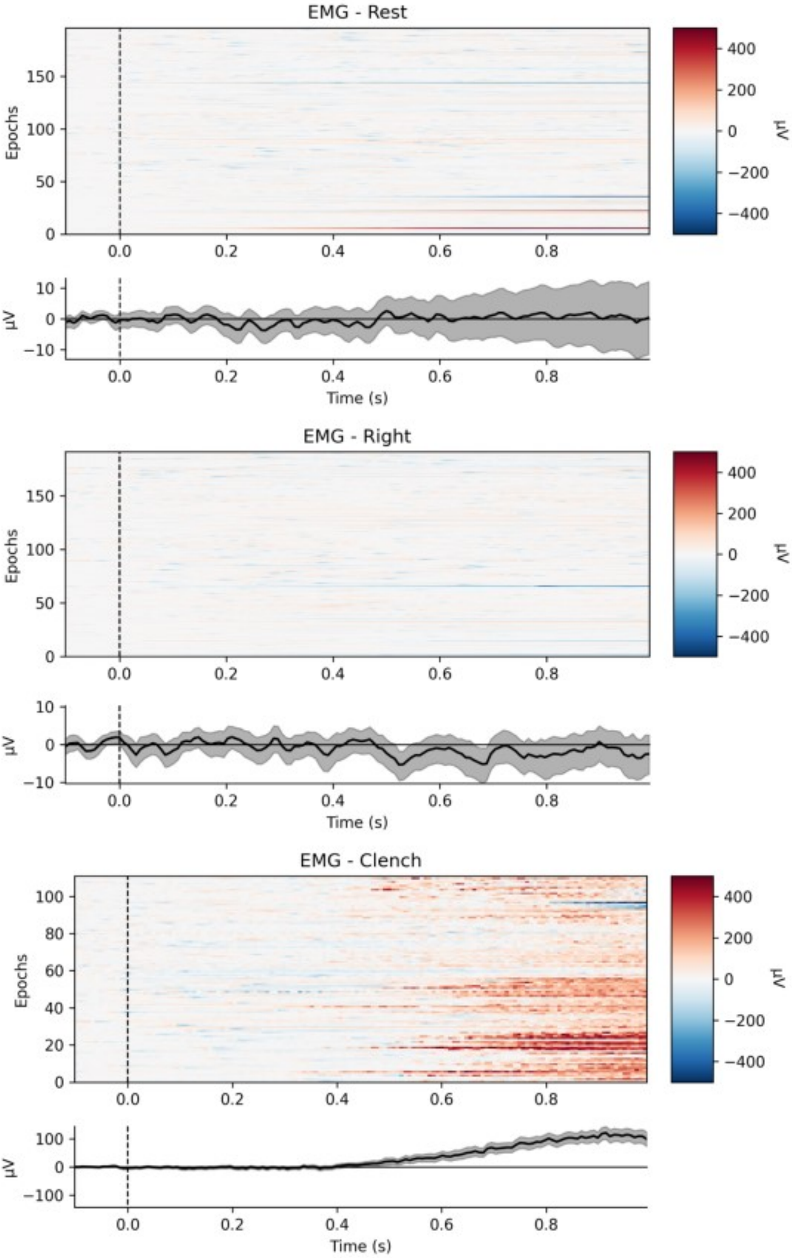
Data from electrodes covering the sensorimotor cortex, namely, C3, C4, P3, and P4 were selected for the purposes of classification of motor imagery movements. These channels were deemed most important in capturing event related desynchronization/event-related synchronization phenomena discussed previously, via the visual inspection of the recordings. Figure 6 below illustrates average activity from C3, C4, P3, and P4 channels during rest vs. right hand conditions for one participant during one of the recording sessions. The difference in activity can be distinguished most clearly in the two plots for the C3 channel, which is expected as the C3 channel covers the area of the sensorimotor cortex responsible for the movement of the right hand.

As previously noted, over 79% of studies reviewed by Hoodis used data from all electrodes for classification. However, multiple studies have achieved good accuracy using eight channels or less (Lun et al., 2020; Yang et al., 2020; Yohanandan et al., 2018). As this study heavily emphasizes the importance of sub-second latency in classifying motor imagery movements, only the essential channels are used to minimize the amount of data required for online processing. Figure 2 illustrates the placement of EEG electrodes, as well as the placement of EMG sensors mentioned above.

Band-pass filters

EEG preprocessing pipeline for MI classification often involves frequency filtering with a bandpass from 8–32 Hz or similar, the upper passband limit removes line noise from single-phase AC electricity (50 Hz or 60 Hz depending on region) and the lower passband limit helps reduce signal drifts which can be caused by changes in subject skin conductivity. Although, there is research suggesting that there is loss of relevant signal in cases of high pass (lower passband) values above 0.1 Hz (Tanner et al., 2015). In many cases, filtering can improve the signal-to-noise ratio, though the process itself can also introduce artifacts such as passband ripple, and thus there are many different approaches to identifying the important aspects of a signal, including signal decomposition methods like independent component analysis (ICA) or ML. In the case of MI classification, bandpass filtering narrows the data to the frequency where the ERD/ERS signal is observed most saliently. However, as EEGNet—which the neu-

Figure 5:



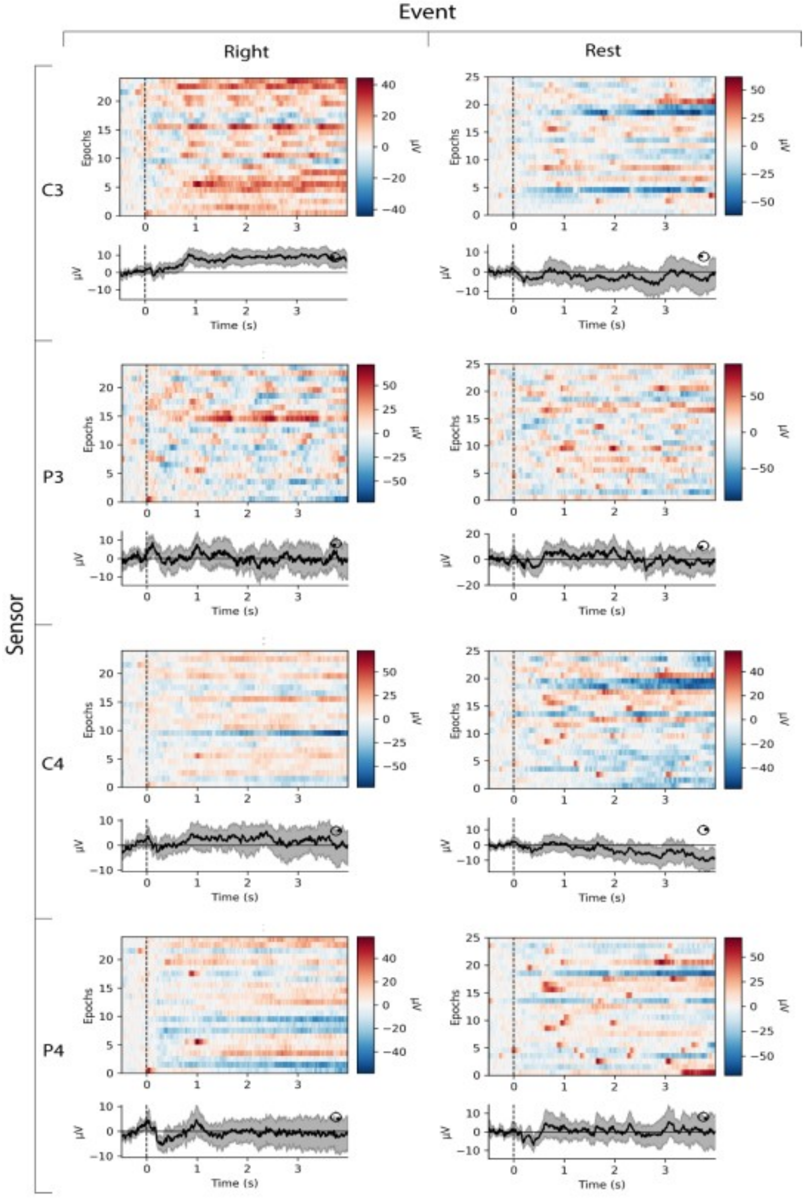
EMG - Rest

EMG - Right

EMG - Clench

Note. *Visualization of EMG data for resting, right MI, and right hand clench trials, note the clear increase in activity and scale of the signal in the clench trials.*

Figure 6:



Note. *Two MI events averaged across all trials for one session of one participant as displayed by the four sensors (C3, C4, P3, P4) used in the training of the neural network.*

ral network proposed below is based on—includes temporal and spatial filters, no frequency filters are applied to the EEG data. Moreover, for real-time, online classification, any frequency-based signal filtering is severely limited by the window size of the data that is being classified, and in our case, the upper window limit of 0.795 s (□101 samples at 128 Hz) makes the introduction of artifacts extremely likely. Reducing the amount of preprocessing also has the important effect of increasing the processing speed of classification.

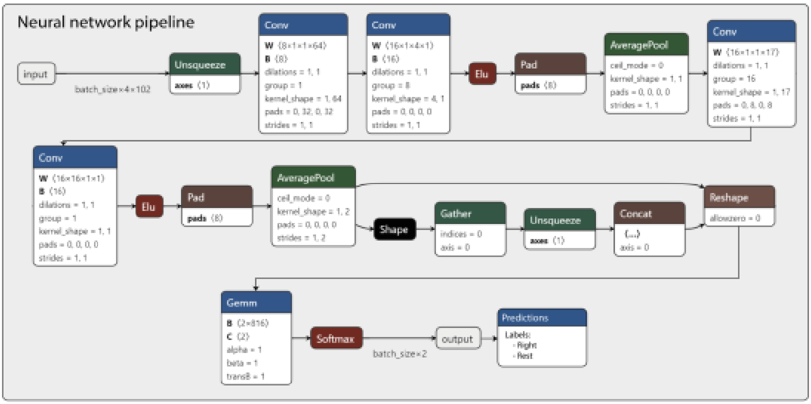
Neural network architecture

A multi-layered Convolutional Neural Network (CNN) architecture (Figure 7) was implemented based on EEGNet (Lawhern et al., 2018) from a modified version of DN3[7] (Kostas & Rudzicz, 2020b, p. 3). All model fitting was performed with Stochastic Gradient Descent (SGD) optimization with momentum, minimizing the two-category cross-entropy loss (Figure 8). A wide variation in hyperparameters were tried to find the optimal values for both within-subject and between-subject models, and while certain parameters such as increasing the number of temporal and spatial filters had a significant effect on increasing training accuracy, this was likely due to overfitting as the validation accuracy did not improve, Table 2 lists the hyperparameters used for the best performing model. Due to EEGNet's application of temporal and spatial filters to the EEG data, we found that the application of various preprocessing techniques (such as windowed FIR bandpass, notch filtering, and average signal subtraction) did not significantly improve classification accuracy. Additionally, strided EEGNet, BENDR (Kostas et al., 2021), and TIDNet (Kostas & Rudzicz, 2020a) based architectures were trained on the data to compare the performance of different networks for classification, both with and without preprocessing, with classification perfor-

---

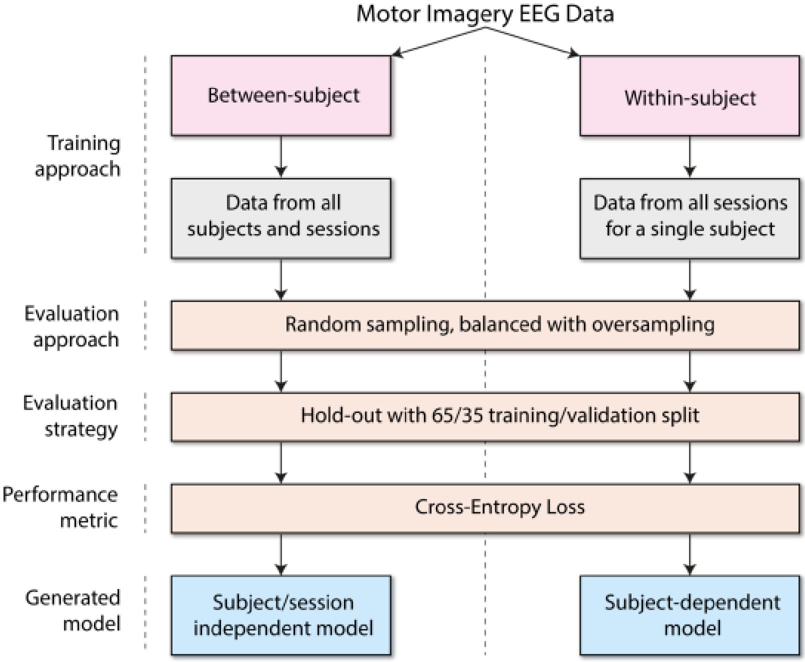
[7] https://github.com/zeyus/dn3

Figure 7:



Note. *Overview of the EEGNet based neural network pipeline used for MI classification.*

Figure 8:



Note. *Diagram describing the process of model training and evaluation.*

mance barely exceeding random chance level.

Although online classification is only viable with a high performing model, due to our goal of making a sub-second MI classifier we implemented an online real-time inference pipeline to assess latency and feasibility. The EEGNet model used resulted in 1 954 trainable parameters (in contrast the BENDR model had 58 284 091 trainable parameters and TIDNet, 608 818 trainable parameters), this reduced number of parameters along with high classification accuracy in benchmarks makes it an ideal architecture for targeting consumer devices which most likely do not have a dedicated GPU/TPU and large amounts of RAM. Minimum continuous latency of [window size] + 205 ms (latency overhead from data transmission, inference, and preprocessing) was consistently achieved with a sliding window step size of 1 sample (meaning 128 predictions/s at 128 Hz sampling rate) on an Intel Core i5-4690K CPU @ 3.50 GHz. These results indicate that online, real-time classification is possible with consumer devices, and finding the narrowest window that maintains high classification accuracy further reduces latency. In our design, the overhead meant that to achieve sub-second online inference, the maximum classification window [window size] was restricted to 0.795 s. This study investigates an even narrower, 0.5 s temporal window for the purposes of classifying motor imagery.

Validation of the network

Prior to testing the capacity of the neural network to predict motor imagery vs. rest behaviours with sub-second latency, the network's baseline prediction capability was tested on data epoched using a 2 s temporal window and compared to the performance of an existing MI-EEG BCI. Yohanandan et al. (2018) reported a mean classification accuracy of 71.09% across 7 participants using a 2 second temporal window, their neural network achieved a mean 63.34% performance across 2 participants using 2 s temporal window. Whilst the accuracy achieved here is lower, Yohanandan et al. (2018) used twice the amount of sensors and a sophisticated neurofeedback routine to improve the capability of participants to perform detectable imaginary movements. Thus, our EEGNet network's prediction capability was deemed acceptable for investigating the ability to classify motor imagery using a 0.785 s temporal window.

Results

Table 2:

Parameter	Value
Dropout Probability	0.25
L2 weight decay	0.01
Base learning rate	0.01
Batch size	16
Epochs	500
Kernel size	32
Spatial filters per temporal convolution (D)	2
Temporal filters (F1)	8
Pointwise filters (F2)	16
Pooling	2

Note. *Best performing hyperparameters for both within- and between-subject models*

Classification accuracies for a 0.5 second temporal window for the two participants were 54.8% and 57.1% respectively for subject-dependent models, and 54.1% for the subject-independent model.

Discussion

Motor imagery-based brain computer interfaces hold great promise for enabling individuals with impaired function, diagnosed with conditions such as ALS (Hosni et al., 2020), cerebral palsy (Taherian et al., 2017), or muscular dystrophy (Bao et al., 2021) to interact with the world around them. In contrast with other BCI paradigms, these systems do not rely on external stimulation, but rather on detecting neural activity produced by the user at will, making them easier to use for prolonged periods of time and easier to set-up. While classification accuracy is clearly important, a survey carried out by (Huggins et al., 2015) shows that speed and ease of use are as important for patients interested in using BCI systems.

In this work, we evaluated the possibility of using a practical EEG for real-time classification of rest vs. motor imagery movements with sub-second latency. The data were collected using a low-cost, dry-electrode Ultracortex Mark IV Headset by OpenBCI. The setup of this EEG device does not require expert assistance, making it suitable for at-home use. A lightweight, deep neural network was constructed using EEGNet, a popular compact convolutional neural network for EEG-based brain-computer interfaces. The built system was able to return predictions in real-time with less than a 1–second delay, but the capacity of the classifier to predict events with high accuracy was not achieved. Since the data were acquired in a naturalistic, interference-prone environment, this study also showcases the potential for the proposed BCI system as a robust solution.

Whilst the visual inspection of the signal hinted at the possibility of differentiating between rest vs. MI conditions using a sub-second temporal window, the neural network could not consistently distinguish the behaviours. As the plots of the signal were taken from a single session, the cause might be the variability between sessions, which was previously reported in other works (Altaheri et al., 2021).

Additional improvements to classification accuracy could be achieved via a denser sensor coverage of the sensory-motor cortex. Most MI-EEG BCI studies utilize data from over 8 electrodes for classification purposes (Altaheri et al., 2021). This study used signals recorded from C3, C4, P3, and P4 electrodes, alternatively, Cz, CP1, CP2, CP5, and CP6 electrodes could be directly added to the Ultracortex Mark IV Headset used in this study.

Limitations and future work

One obvious limitation of the study was its small sample size of two participants, which is not enough to reliably determine whether the proposed approach would generalize to larger populations. Training the neural network on more within-subject and between-subject data would give the network a broader range of activity to pick out the most relevant features of the signal and to give less weight to noise that is less consistent across trials.

Another obvious limitation that needs to be addressed is the low accuracy of the classifier. In a real-world scenario, this can lead to misinterpretation of user intentions, leading to incorrect responses from the system and a decrease in user satisfaction, as well as an increased risk of errors that could be potentially dangerous for the user.

In the section above, the possibility of adding additional sensors with the purpose of covering the sensorimotor cortex more densely was discussed. However, the 3D printed headset by OpenBCI used in the study could not facilitate the addition of the following channels: C1, C2, CP3, CP4, and CPz, which could enhance the targeted ERD/

ERS signal in the acquired data. An alternative headset could be designed and developed for the sole purpose of MI classification, which would include sockets for the sensors listed above. Such additions could increase the network classification accuracy by providing more relevant data.

An attempt at improving classification accuracy could be made via moving the temporal window forward in increments, up to around 0.4 s after the stimulus onset. As seen in Figure 5, when the experimental protocol included a clench condition, the EMG activity did not increase up until around 0.4 s after the stimulus onset, meaning it took approximately 0.4 s for participants to process the stimulus, start acting on it, and subsequently move their muscles. It can be therefore assumed there might also be a delay in the MI condition, minus the latency of signal propagation from the brain to the hand.

The session-dependent variability could be addressed via more consistent performance of the kinesthetic imagery of the required movement. Previous studies on motor imagery have utilized the KVIQ-10 questionnaire (Malouin et al., 2007) to ensure that participants are able to perform the task correctly and consistently, with experimenters explaining and showing the movements prior to the assessment. Alternatively, a neurofeedback training routine akin to that described by Hwang et al., 2009 could be implemented to help participants accurately and consistently perform imaginary movements by presenting them with real-time brain activation graphs.

Whilst this study implemented a visual diagnostic of EMG activity across the two conditions to ensure no muscle activation was present during imaginary movements, a more rigorous and automatic rejection procedure on a per-trial basis could be implemented akin to that by Peterson et al. (2022). This would further ensure that the signal in acquired data is evoked by purely imaginary behaviour, allowing for fewer false positives in the case where the user would clench their right hand.

MI classification in non-stationary settings

As noted at the beginning of the paper, for BCIs to be of value to a larger audience, they need to be usable in non-stationary scenarios. EEG has been proven effective in the classification of multiple types of signals with a high temporal resolution. Unfortunately, it suffers from a high degree of immobility. Motion-related artifacts are one of the most common causes of noise in EEG data. Two other, relatively recently developed methods, namely functional near-infrared spectroscopy (fNIRS) and optically-pumped magnetometers for magnetoencephalography (OPM-MEG) are perfectly suited to tackle this problem. Although fNIRS has a relatively low degree of immobility, its temporal resolution is rather low, meaning that high online classification speed will be difficult to achieve due to the limitations of the technology behind the method. OPM-MEG, however, has an excellent temporal resolution, and benefits from better spatial resolution than EEG, but is more costly (Boto et al., 2016). Importantly, studies have shown OPM-MEG to reliably record neural activity from moving individuals (Boto et al., 2018). Thus, research on the classification capacity of motor imagery in non-stationary settings using OPM-MEG is deemed promising.

Conclusion

This study demonstrated a successful implementation of an EEG-based BCI system able to perform real-time, binary motor imagery classification with a sub-second prediction latency. A low-cost EEG device from OpenBCI was used on data that were acquired in an interference-prone environment, indicating the deployability of the system for real-world applications. Sub-second prediction latency was achieved using a light CNN based on EEGNet, minimum preprocessing of the raw EEG data, and a temporal window of 0.785 s from the onset of the stimulus. The predictability of motor imagery vs. rest behaviours using a temporal window of 0.785 s was examined. Subject-dependent classification accuracies were 54.8% and 57.1% for the two participants, respectively. Cross-subject classification accuracy was 54.1%. Future work includes evaluating the system using data from more sensors and from participants trained in motor imagery by means of neurofeedback or expert training.

References

Abdi, H., & Williams, L.J. (2010). Principal component analysis: Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2(4),* 433–459. https://doi.org/10.1002/wics.101

Abiri, R., Borhani, S., Sellers, E.W., Jiang, Y., & Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering, 16(1),* 011001. https://doi.org/10.1088/1741-2552/aaf12e

Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H., & Subha, D.P. (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Computer Methods and Programs in Biomedicine, 161,* 103–113. https://doi.org/10.1016/j.cmpb.2018.04.012

Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S.U., Altuwaijri, G.A., Abdul, W., Bencherif, M.A., & Faisal, M. (2021). Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review. Neural Computing and Applications. https://doi.org/10.1007/s00521-021-06352-5

An, X., Kuang, D., Guo, X., Zhao, Y., & He, L. (2014). A Deep Learning Method for Classification of EEG Data Based on Motor Imagery. In D.-S. Huang, K. Han, & M. Gromiha (Eds.), Intelligent Computing in Bioinformatics (Vol. 8590, pp. 203–210). Springer International Publishing. https://doi.org/10.1007/978-3-319-09330-7_25

Artola, A., Bröcher, S., & Singer, W. (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature, 347(6288),* 69–72. https://doi.org/10.1038/347069a0

Ball, T., Kern, M., Mutschler, I., Aertsen, A., & Schulze-Bonhage, A. (2009). Signal quality of simultaneously recorded invasive and non-invasive EEG. *NeuroImage, 46(3),* 708–716. https://doi.org/10.1016/j.neuroimage.2009.02.028

Bao, S.-C., Yuan, K., Chen, C., Lau, C.C., & Tong, R.K. (2021). A Motor Imagery-based Brain-Computer Interface Scheme for a Spinal Muscular Atrophy Subject in CYBATHLON Race. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER),* 532–535. https://doi.org/10.1109/NER49283.2021.9441351

Barsotti, M., Leonardis, D., Loconsole, C., Solazzi, M., Sotgiu, E., Procopio, C., Chisari,C., Bergamasco, M., & Frisoli, A. (2015). A full upper limb robotic exoskeleton for reaching and grasping rehabilitation triggered by MI-BCI. *2015 IEEE International Conference on Rehabilitation Robotics (ICORR),* 49–54. https://doi.org/10.1109/ICORR.2015.7281174

Bliss, T.V.P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology, 232(2),* 331–356. https://doi.org/10.1113/jphysiol.1973.sp010273

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., & Zieba, K. (2016). End to End Learning for Self-Driving Cars. https://doi.org/10.48550/ARXIV.1604.07316

Boto, E., Bowtell, R., Krüger, P., Fromhold, T.M., Morris, P.G., Meyer, S.S., Barnes, G.R., & Brookes, M.J. (2016). On the Potential of a New Generation of Magnetometers for MEG: A Beamformer Simulation Study. *PLOS ONE, 11(8),* e0157655. https://doi.org/10.1371/journal.pone.0157655

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S.S., Muñoz, L.D., Mullinger, K.J., Tierney, T.M., Bestmann, S., Barnes, G.R., Bowtell, R., & Brookes, M.J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature, 555(7698),* 657–661. https://doi.org/10.1038/nature26147

Cecotti, H., & Graser, A. (2010). Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces. IEEE.

Chang, M.H., Baek, H.J., Lee, S.M., & Park, K.S. (2014). An amplitude-modulated visual stimulation for reducing eye fatigue in SSVEP-based brain–computer interfaces. *Clinical Neurophysiology, 125(7),* 1380–1391. https://doi.org/10.1016/j.clinph.2013.11.016

Christensen, S.M., Holm, N.S., & Puthusserypady, S. (2019). An Improved Five Class MI Based BCI Scheme for Drone Control Using Filter Bank CSP. *2019 7th International Winter Conference on Brain-Computer Interface (BCI),* 1–6. https://doi.org/10.1109/IWW-BCI.2019.8737263

Craik, A., He, Y., & Contreras-Vidal, J.L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering, 16(3),* 031001. https://doi.org/10.1088/1741-2552/ab0ab5

Dickstein, R., & Deutsch, J.E. (2007). Motor Imagery in Physical Therapist Practice. *Physical Therapy, 87(7),* 942–953. https://doi.org/10.2522/ptj.20060331

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in Psychophysiology, 2(S1),* 78.

Farwell, L.A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 70(6),* 510–523. https://doi.org/10.1016/0013-4694(88)90149-6

Friston, K.J. (2009). Modalities, Modes, and Models in Functional Neuroimaging. *Science, 326(5951),* 399–403. https://

doi.org/10.1126/science.1174521

Guger, C., Ramoser, H., & Pfurtscheller, G. (2000). Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI). *IEEE Transactions on Rehabilitation Engineering, 8(4),* 447–456. https://doi.org/10.1109/86.895947

Hebb, D.O. (1949). The Organization of Behaviour.

Hosni, S.M., Borgheai, S.B., McLinden, J., & Shahriari, Y. (2020). An fNIRS-Based Motor Imagery BCI for ALS: A Subject-Specific Data-Driven Approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(12),* 3063–3073. https://doi.org/10.1109/TNSRE.2020.3038717

Huggins, J.E., Moinuddin, A.A., Chiodo, A.E., & Wren, P.A. (2015). What Would Brain-Computer Interface Users Want: Opinions and Priorities of Potential Users With Spinal Cord Injury. *Archives of Physical Medicine and Rehabilitation, 96(3),* S38 -S45.e5. https://doi.org/10.1016/j.apmr.2014.05.028

Hussein, R., Palangi, H., Ward, R.K., & Wang, Z.J. (2019). Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clinical Neurophysiology, 130(1),* 25–37. https://doi.org/10.1016/j.clinph.2018.10.010

Hwang, H.-J., Han, C.-H., Lim, J.-H., Kim, Y.-W., Choi, S.-I., An, K.-O., Lee, J.-H., Cha, H.-S., Hyun Kim, S., & Im, C.-H. (2017). Clinical feasibility of brain-computer interface based on steady-state visual evoked potential in patients with locked-in syndrome: Case studies: Clinical feasibility of SSVEP-based BCI. *Psychophysiology, 54(3),* 444–451. https://doi.org/10.1111/psyp.12793

Hwang, H.-J., Kwon, K., & Im, C.-H. (2009). Neurofeedback-based motor imagery training for brain–computer interface (BCI). *Journal of Neuroscience Methods, 179(1),* 150–156. https://doi.org/10.1016/j.jneumeth.2009.01.015

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33(11),* 1419–1432. https://doi.org/10.1016/0028-3932(95)00073-C

Kevric, J., & Subasi, A. (2017). Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomedical Signal Processing and Control, 31,* 398–406. https://doi.org/10.1016/j.bspc.2016.09.007

Khan, M.J., & Hong, K.-S. (2015). Passive BCI based on drowsiness detection: An fNIRS study. *Biomedical Optics Express, 6(10),* 4063. https://doi.org/10.1364/BOE.6.004063

Khosla, A., Khandnor, P., & Chand, T. (2020). A comparative analysis of signal processing and classification methods for different applications based on EEG signals. *Biocybernetics and Biomedical Engineering, 40(2),* 649–690. https://doi.org/10.1016/j.bbe.2020.02.002

Kirschstein, T., & Köhling, R. (2009). What is the Source of the EEG? *Clinical EEG and Neuroscience, 40(3),* 146–149. https://doi.org/10.1177/155005940904000305

Kostas, D., Aroca-Ouellette, S., & Rudzicz, F. (2021). BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. *Frontiers in Human Neuroscience, 15,* 653659. https://doi.org/10.3389/fnhum.2021.653659

Kostas, D., & Rudzicz, F. (2020a). Thinker invariance: Enabling deep neural networks for BCI across more people. *Journal of Neural Engineering, 17(5),* 056008. https://doi.org/10.1088/1741-2552/abb7a7

Kostas, D., & Rudzicz, F. (2020b). DN3: An open-source Python library for large-scale raw neurophysiology data assimilation for more flexible and standardized deep learning [Preprint]. *Neuroscience.* https://doi.org/10.1101/2020.12.17.423197

Kraeutner, S., Gionfriddo, A., Bardouille, T., & Boe, S. (2014). Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations. *Brain Research, 1588,* 81–91. https://doi.org/10.1016/j.brainres.2014.09.001

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems (Vol. 25). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45 b-Paper.pdf

Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M., & Wolpaw, J.R. (2008). Toward enhanced P300 speller performance. *Journal of Neuroscience Methods, 167(1),* 15–21. https://doi.org/10.1016/j.jneumeth.2007.07.017

Kuś, R., Duszyk, A., Milanowski, P., Łabęcki, M., Bierzyńska, M., Radzikowska, Z., Michalska, M., Żygierewicz, J., Suffczyński, P., & Durka, P.J. (2013). On the Quantification of SSVEP Frequency Responses in Human EEG in Realistic BCI Conditions. *PLoS ONE, 8(10),* e77536. https://doi.org/10.1371/journal.pone.0077536

Kwak, N.-S., Müller, K.-R., & Lee, S.-W. (2015). A lower limb exoskeleton control system based on steady state visual evoked potentials. *Journal of Neural Engineering, 12(5),* 056009. https://doi.org/10.1088/1741-2560/12/5/056009

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., & Lance, B.J. (2018). EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering, 15(5),* 056013. https://doi.org/10.1088/1741-2552/aace8c

Lebedev, M.A., & Nicolelis, M.A.L. (2017). Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. *Physiological Reviews, 97(2),* 767–837. https://doi.org/10.1152/physrev.00027.2016

Lotze, M., & Halsband, U. (2006). Motor imagery. *Journal of Physiology-Paris, 99(4–6),* 386–395. https://doi.org/10.1016/j.jphysparis.2006.03.012

Lun, X., Yu, Z., Chen, T., Wang, F., & Hou, Y. (2020). A Simplified CNN Classification Method for MI-EEG via the Electrode Pairs Signals. *Frontiers in Human Neuroscience, 14,* 338. https://doi.org/10.3389/fnhum.2020.00338

Ma, X., Qiu, S., & He, H. (2020). Multi-channel EEG recording during motor imagery of different joints from the same limb. *Scientific Data, 7(1),* 191. https://doi.org/10.1038/s41597-020-0535-2

Malouin, F., Jackson, P.L., & Richards, C.L. (2013). Towards the integration of mental practice in rehabilitation programs. A critical review. *Frontiers in Human Neuroscience*, 7. https://doi.org/10.3389/fnhum.2013.00576

Malouin, F., Richards, C.L., Jackson, P.L., Lafleur, M.F., Durand, A., & Doyon, J. (2007). The Kinesthetic and Visual Imagery Questionnaire (KVIQ) for Assessing Motor Imagery in Persons with Physical Disabilities: A Reliability and Construct Validity Study. *Journal of Neurologic Physical Therapy, 31(1),* 20–29. https://doi.org/10.1097/01.NPT.0000260567.24122.64

McCrimmon, C.M., Wang, M., Lopes, L.S., Wang, P.T., Karimi-Bidhendi, A., Liu, C.Y., Heydari, P., Nenadic, Z., & Do, A.H. (2016). A small, portable, battery-powered brain-computer interface system for motor rehabilitation. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2776–2779. https://doi.org/10.1109/EMBC.2016.7591306

McFarland, D.J., & Vaughan, T.M. (2016). BCI in practice. *In Progress in Brain Research (Vol. 228, pp. 389–404).* Elsevier. https://doi.org/10.1016/bs.pbr.2016.06.005

Morabito, F.C., Campolo, M., Ieracitano, C., Ebadi, J.M., Bonanno, L., Bramanti, A., Desalvo, S., Mammone, N., & Bramanti, P. (2016). Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI),* 1–6. https://doi.org/10.1109/RTSI.2016.7740576

Newell, K.M. (1991). Motor skill acquisition. Annual Review of Psychology, 42(1), 213-237. Nutt, R. (2002). The History of Positron Emission Tomography. *Molecular Imaging & Biology, 4(1),* 11–26. https://doi.org/10.1016/S1095-0397(00)00051-0

Olejniczak, P. (2006). Neurophysiologic Basis of EEG: *Journal of Clinical Neurophysiology, 23(3),* 186–189. https://doi.org/10.1097/01.wnp.0000220079.61973.6c

ONNX Runtime developers. (2021). ONNX Runtime. https://onnxruntime.ai/

Paek, A.Y., Kilicarslan, A., Korenko, B., Gerginov, V., Knappe, S., & Contreras-Vidal, J.L. (2020). Towards a Portable Magnetoencephalography Based Brain Computer Interface with Optically-Pumped Magnetometers. 2020 *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* 3420–3423. https://doi.org/10.1109/EMBC44109.2020.9176159

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., H"ochenberger, R., Sogo, H., Lindelov, J.S. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51(1),* 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peterson, V., Galván, C., Hernández, H., Saavedra, M.P., & Spies, R. (2022). A motor imagery vs. Rest dataset with low-cost consumer grade EEG hardware. *Data in Brief, 42,* 108225. https://doi.org/10.1016/j.dib.2022.108225

Pfurtscheller, G. (2000). Chapter 26 Spatiotemporal ERD/ERS patterns during voluntary movement and motor imagery. *In Supplements to Clinical Neurophysiology (Vol. 53, pp. 196–198).* Elsevier. https://doi.org/10.1016/S1567-424X(09)70157-6

Pfurtscheller, G., & Neuper, C. (1997). Motor imagery activates primary sensorimotor area in humans. *Neuroscience Letters, 239 (2–3)*, 65–68. https://doi.org/10.1016/S0304-3940(97)00889-6

Pfurtscheller, G., Solis-Escalante, T., Ortner, R., Linortner, P., & Muller-Putz, G.R. (2010). Self-Paced Operation of an SSVEP-Based Orthosis With and Without an Imagery-Based "Brain Switch:" A Feasibility Study Towards a Hybrid BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18(4),* 409–414. https://doi.org/10.1109/TNSRE.2010.2040837

Pichiorri, F., & Mattia, D. (2020). Brain-computer interfaces in neurologic rehabilitation practice. *In Handbook of Clinical Neurology (Vol. 168, pp. 101–116*). Elsevier. https://doi.org/10.1016/B978-0-444-63934-9.00009-3

Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C.L., Zeng, Q., Laugier, C., & Ang, M.H. (2007). Controlling a Wheelchair Indoors Using Thought. *IEEE Intelligent Systems, 22(2),* 18–24. https://doi.org/10.1109/MIS.2007.26

Reshmi, G., & Amal, A. (2013). Design of a BCI System for Piloting a Wheelchair Using Five Class MI Based EEG. *2013 Third International Conference on Advances in Computing and Communications*, 25–28. https://doi.org/10.1109/ICACC.2013.12

Robertson, E.M., Pascual-Leone, A., & Miall, R.C. (2004). Current concepts in procedural consolidation. *Nature Reviews Neuroscience, 5(7),* 576–582. https://doi.org/10.1038/nrn1426

Schalk, G. (2009). Sensor Modalities for Brain-Computer Interfacing. In J. A. Jacko (Ed.), Human-Computer Interaction. *Novel Interaction Methods and Techniques (Vol. 5611, pp. 616–622).* Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-

642-02577-8_67

Schalk, G., & Leuthardt, E.C. (2011). Brain-Computer Interfaces Using Electrocorticographic Signals. *IEEE Reviews in Biomedical Engineering, 4,* 140–154. https://doi.org/10.1109/RBME.2011.2172408

Schuster, C., Hilfiker, R., Amft, O., Scheidhauer, A., Andrews, B., Butler, J., Kischka, U., & Ettlin, T. (2011). Best practice for motor imagery: A systematic literature review on motor imagery training elements in five different disciplines. *BMC Medicine, 9(1),* 75. https://doi.org/10.1186/1741-7015-9-75

Gao, S.,Shangkai Gao, Wang, Y.,Yijun Wang, Gao, X., Xiaorong Gao, & Hong, B.Bo Hong. (2014). Visual and Auditory Brain–Computer Interfaces. *IEEE Transactions on Biomedical Engineering, 61(5),* 1436–1447. https://doi.org/10.1109/TBME.2014.2300164

Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., Kuebler, A., & Birbaumer, N. (2007). fMRI Brain-Computer Interface: A Tool for Neuroscientific Research and Treatment. *Computational Intelligence and Neuroscience, 2007,* 1–10. https://doi.org/10.1155/2007/25487

Stinear, C.M., Byblow, W.D., Steyvers, M., Levin, O., & Swinnen, S.P. (2006). Kinesthetic, but not visual, motor imagery modulates corticomotor excitability. *Experimental Brain Research, 168(1–2),* 157–164. https://doi.org/10.1007/s00221-005-0078-y

Stone, J. V. (2002). Independent component analysis: An introduction. *Trends in Cognitive Sciences, 6(2),* 59–64. https://doi.org/10.1016/S1364-6613(00)01813-1

Tabar, Y.R., & Halici, U. (2017). A novel deep learning approach for classification of EEG motor imagery signals. *Journal of Neural Engineering, 14(1),* 016003. https://doi.org/10.1088/1741-2560/14/1/016003

Taherian, S., Selitskiy, D., Pau, J., & Claire Davies, T. (2017). Are we there yet? Evaluating commercial grade brain–computer interface for control of computer applications by individuals with cerebral palsy. *Disability and Rehabilitation: Assistive Technology, 12(2),* 165–174. https://doi.org/10.3109/17483107.2015.1111943

Tanner, D., Morgan-Short, K., & Luck, S.J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition: High-pass filtering and artifactual ERP effects. *Psychophysiology, 52(8),* 997–1009. https://doi.org/10.1111/psyp.12437

Vaughan, T.M., McFarland, D.J., Schalk, G., Sarnacki, W.A., Krusienski, D.J., Sellers, E.W., & Wolpaw, J.R. (2006). The wadsworth BCI research and development program: At home with BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(2),* 229–233. https://doi.org/10.1109/TNSRE.2006.875577

Vialatte, F.-B., Maurice, M., Dauwels, J., & Cichocki, A. (2010). Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology, 90(4),* 418–438. https://doi.org/10.1016/j.pneurobio.2009.11.005

Willett, F.R., Avansino, D.T., Hochberg, L.R., Henderson, J.M., & Shenoy, K.V. (2021). High-performance brain-to-text communication via handwriting. *Nature, 593(7858),* 249–254. https://doi.org/10.1038/s41586-021-03506-2

Yang, J., Ma, Z., Wang, J., & Fu, Y. (2020). A Novel Deep Learning Scheme for Motor Imagery EEG Decoding Based on Spatial Representation Fusion. *IEEE Access, 8,* 202100–202110. https://doi.org/10.1109/ACCESS.2020.3035347

Yao, L., Sheng, X., Zhang, D., Jiang, N., Farina, D., & Zhu, X. (2017). A BCI System Based on Somatosensory Attentional Orientation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(1),* 81–90. https://doi.org/10.1109/TNSRE.2016.2572226

Yohanandan, S.A. C., Kiral-Kornek, I., Tang, J., Mshford, B. S., Asif, U., & Harrer, S. (2018). A Robust Low-Cost EEG Motor Imagery-Based Brain-Computer Interface. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 5089–5092. https://doi.org/10.1109/EMBC.2018.8513429

Zhang, K., Robinson, N., Lee, S.-W., & Guan, C. (2021). Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network. *Neural Networks, 136,* 1–10. https://doi.org/10.1016/j.neunet.2020.12.013

The Net

Wen Li

The Extended Mind and Chimpanzee Consciousness: An exploration of the implications of viewing the mind as socially extendable

Mia Jacobsen

Aarhus University

Department of Linguistics, Cognitive Science and Semiotics

Abstract

In 2013, philosopher Shaun Gallagher presented a new interpretation of Clark and Chalmer's extended mind theory. He proposed that our cognition can extend not only to physical objects but to social institutions as well. This understanding of our mind rests upon the interpretation that we as conscious beings have a mind that is able to extend in a way that is observable through our behaviour. In this paper, I apply Gallagher's socially extended mind theory to chimpanzees to argue that if we accept this theory as true, then we must also accept it as true that chimpanzees are conscious. My argument builds upon instances where chimpanzee behaviour parallels human behaviour in a way that the behaviour can be interpreted as arising from a socially extended mind and thus a consciousness. I especially highlight the role of the stable social hierarchy that exists within a chimpanzee colony, and how that influences the behaviour and cognition of the chimps. I conclude the paper by arguing that research into ape consciousness so far has been too human-focused, to a point that we have forgotten what it is like to be a chimpanzee. We should instead listen to Nagel and try to take up the animals' point of view and be more open to the existence of other kinds of consciousness outside of our own.

Keywords: **extended mind, animal cognition, consciousness**

Introduction

Some four-hundred years ago, Decartes argued that the only knowledge we can be absolutely sure of is the existence of our own mind. I think, therefore I am (Decartes, 1641). This also means that the only consciousness we can be absolutely sure of is our own. Now, however true this observation may be, it does not intuitively fit with our actual experience of existing as human beings within a society. For despite my inaccessibility to your thoughts, feelings, emotions, and intentions, I still ascribe these internal states to you. Human beings, in general, automatically assume that other human beings have a conscious experience similar to their own. But despite how natural it feels to ascribe consciousness to ourselves and others, the actual nature of human consciousness still largely eludes us. Whether we approach it neurobiologically through brain scanning methods, developmentally through studies of children, or purely philosophically, there seems to be little consensus on what consciousness is and how we know we have it. And this is despite the fact that investigating human consciousness is, for the most part, the 'easy mode' of consciousness research—for we can (at least try to) communicate our conscious experiences to each other. When one wishes to branch out to other species to the question of whether non-human animals have consciousness, matters are complicated even further. For besides the obvious obstacle of communication, if the nature of our own consciousness still eludes us, how can we ever hope to understand something that is basically completely alien to us, i.e., a non-human conscious experience?

In this paper, I seek to approach the question of non-human consciousness in a slightly different—and perhaps backwards—way. I will be exploring how theories of human cognition can be applied to non-human animals' patterns of normative practices in stable social groups in order to expand our understanding of what it means to have consciousness.

Specifically, I will be investigating the implications of the socially extended mind theory, as presented by Gallagher (2013), with chimpanzees as a case study. I will be arguing that if the socially extended mind theory is accepted as true, then it must mean that chimpanzees, at least to some degree, have consciousness. The main argument of the paper is based on an analysis of different chimpanzee behaviours. I look at how they parallel human behaviour and thus exemplify the ways chimpanzee minds can be socially extendable. The paper ends with a discussion of the methodological issues of studying non-human animals, and how we can infer consciousness in other beings with inspiration drawn from Nagel's *What Is It Like to Be a Bat* (1974).

What Does it Mean to be Conscious?

To discuss whether a non-human animal can be seen as having consciousness it must first be determined what the definition of consciousness is. As with many concepts within the overlap of psychology and philosophy, it is notoriously hard to agree on a specific definition. In this paper, I define consciousness as philosopher John Searle does, by four features: 1) It's real and irreducible, 2) caused by brain processes, 3) exists in the brain, and 4) functions causally (Searle, 2009). The key takeaway from this definition is that consciousness is a biological phenomenon that is not detached from the brain but rather is a feature of the brain (Searle, 2009). Additionally, consciousness' causal functionality means that phenomena like thoughts, beliefs, and emotions arise from a being's consciousness. These internal phenomena—as they are not directly observable to outsiders—are what I define as the mind, and they can be seen as the mental contents of any conscious experience. In other words, the mind is the collection of thoughts, emotions, perceptions, and other mental states that make up an individual's conscious experience. For example, if you are experiencing the sensation of pain, your mind would be said to include the mental content of that experience: The sensation itself, your awareness of it, and any thoughts or emotions you have about it. So the mind arises from our consciousness and is an integral part of the phenomenological experience of being at any given time. Now, as over a century of thought has been spent trying to define mind and consciousness, the definitions presented here do not try to be definitive or comprehensive. To highlight just a few different accounts, a classic example is Bentham (1789), with the role of suffering in assessing consciousness. Other papers emphasize the role of certain brain connections (Lagercrantz & Changeux, 2010), attention (Posner, 1994), or an awareness and experience of time (Kent & Wittmann, 2021). Others still use behaviour to infer consciousness (Kotchoubey, 2018), or a computational framework (Dehaene, Lau, & Kouider, 2017). The same goes for the mind, where the different approaches include viewing the mind as the representation of subjective experience removed from brain function (Libet, 2006), an evolutionary developed system for information-processing (Pinker, 2005), or the attribution of thoughts, beliefs, and intention in others, also known as "theory of mind" (Sprong et al., 2007). Instead of using this paper to argue for or against any specific approach in this larger philosophical discussion of how to define consciousness and mind, I have decided to go with two simple definitions for the purpose of clarity which simultaneously emphasizes the link between the two.

Lastly, there is the concept of cognition. I will be working from an embodied view of cognition, which sees cognition as inherently tied to our body and bodily experiences. Essentially, our cognitive processes emerge from our bodily interactions with the world

(Pfeifer & Bongard, 2007). This means that cognition, the mind, and the body are not separate entities, but rather they are meshed together and shaped by our body's interaction with the world through its particular perceptual and motor systems (Thelen, 2000).

The Extended Mind

A central question is where the demarcation between our mind and the rest of the world should be. Can our mind and cognition extend beyond our brain and body? In their 1998 paper, Andy Clark and David Chalmers argue for an *active externalism* of the mind. They want to acknowledge the active role the environment plays in our cognitive processes, and how the mind and cognition can integrate the physical world (Clark & Chalmers, 1998). Their main argument for the extension of our cognitive processes, which is known as *the parity principle,* is:

> If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. (Clark & Chalmers, 1998, p. 8).

Essentially, if an external object, for example, a calculator or notebook, takes on the function of an otherwise 'inside job', e.g., the working memory or long-term memory, then that object becomes part of the cognitive process. It creates a coupled system between the person and the external object, where each of the components play an active causal role in the performance of the task at hand (Clark & Chalmers, 1998). In addition, the mind can be extended to and be formed by the environment much in the same way cognition can be extended, both by objects and other people (Clark & Chalmers, 1998). This notion of a socially extended mind has been elaborated upon by philosopher Shaun Gallagher (2013).

The Socially Extended Mind

In his 2013 paper, Gallagher presents a liberal and expanding interpretation of Clark and Chalmers' extended mind theory. In essence, Gallagher argues that cognition can be socially extensive in a way that includes not only interaction with other humans but also interactions with so-called *mental institutions* (Gallagher, 2013). The socially extended mind theory sees extended cognition as not merely including physical objects, but also social situations and institutions that in some way create the coupled system Clark and Chalmers describe. According to Gallagher, if the cognitive process somehow relies on a cognitively produced tool, rule, or institution, then there is no reason to limit mind extension to tasks which we can imagine being "done in the head" (Gallagher, 2013). A good example of this is the legal system. The legal system consists of agreements, decisions, and systems of rights that are the representation of several externalized minds. It reinforces certain behaviours, and allows us to solve certain problems by guiding our thinking about social arrangements such as property, contracts, and rights (Gallagher, 2013). Essentially, engagement with these tools and institutions created by the legal system makes simple behaviours and decisions, such as building a fence around your house, part of a larger cognitive process incorporating the cultural understanding of property and what it means to own.

Cognition and subsequently the mind are thus not only confined to the brain and body. Rather, they are co-constructed and supported by both the normative practices and institutions in our environment which we engage with and contribute to (Cash, 2013), as well as our biology which gives rise to our consciousness, mind, and cognition.

What Does it Mean for Consciousness?

With this broadened understanding of the mind and its relation to the world, the central question returns: Where should the demarcation between the mind and the world lie? If the mind and cognition extend to everything we interact with, is the whole world then potentially part of our cognition? This is known as the *cognitive bloat* worry—the critique that the extended mind theory over-extends the mind so the term becomes meaningless (Rupert, 2004). This critique has been refuted by Gallagher as he asserts the role of active engagement: The cognitive process extends as we engage with the external world in the right way (Gallagher, 2013). Specifically, when our engagement with the external object or institution defines the cognitive process in such a way that without the extension the process would be different or non-existent, then it is an extended cognitive process. One might argue, though, that this does not actually address the problem of over-extension: Maybe not all cognitive processes are extended, but those that are, do they not still potentially over-extend the mind and over-incorporate the external world? However, the nature of the extension needs to be considered as well. A socially extended cognitive process does not mean that the literal objects that are manipulated become part of our minds. Instead, the process itself is what extends to draw upon the resources in the environment that make the process a possibility at all. Take, for example, building a pasture for cattle. The wooden poles, the wire, and the actual manipulations of those objects are not part of the cognitive process, but the engagement with the local custom of solving such a problem with a fence is an extension—if the local custom had been different, the solution would have been different (Gallagher, 2013). Subsequently, it is not the brain, or any other bodily part that extends, nor physical objects that become our mind. Instead, a socially extended cognitive process should be understood as our cognition shaping and being shaped by the external environment, in such a way that the cognitive process is simultaneously occurring in the brain and extending to the resources in the outside world which inform our cognition and make the cognitive process possible.

When we then as conscious human beings engage in behaviour that is seen as extending our minds, it also becomes possible to infer consciousness based on specific behavioural patterns using the socially extended mind theory. If our behaviour arises from (extended) cognitive processes that are part of the mind, which is defined as the mental contents of consciousness, then consciousness is necessary for a mind, mind extension, and thus our behaviour. If we accept this as true, then it must also be true that if some other being engages in behaviour that is best explained as arising from mind extension, it must also arise from some sort of consciousness. And it is this exact thought process I wish to apply to chimpanzees.

Although the applicability of the social extended mind theory to other non-human animals besides chimpanzees is missing in this account, I maintain my focus on chimps for a couple of reasons. First, in order to apply a theory of human cognition to non-human animals, it seems natural to start with our closest living relative on the tree of evolution—if any animal is going to have similar internal states to humans, it should be chimps. Second, the long history of research into the mental capabilities of chimps means there is a plethora of both controlled experiments and structured observations of chimps in the wild and in captivity (see Call and Tomasello, 2008 for a review). However, this does not exclude the possibility of applying the social extended mind theory to other animals, or that this is the only way of asserting consciousness in non-human animals. Instead, the primary purpose of this paper is to assess how a theory of human cognition might be applied to investigate consciousness and internal mental states based on external behaviour in order to both argue that chimps can be ascribed consciousness, and that this methodology is a useful and meaningful way of approaching the question of consciousness in non-human animals.

The Chimpanzee Extended Mind

With this theoretical groundwork, it becomes interesting to investigate how chimpanzee behaviour can parallel human behaviour. I will use descriptions of chimpanzee behaviour as observed both in the wild and in captivity in tandem with the socially extended mind theory to argue that the behaviour exhibited by the chimps is not only a parallel to human behaviour but also evidence of a socially extended chimp mind and thus chimp consciousness.

Object Manipulation

Nearly 60 years ago, Dr. Jane Goodall reported seeing wild chimpanzees producing and using self-made tools on many different occasions (Goodall, 1964). For example, three adult chimpanzees, independently of each other, tried to open boxes of bananas by using sticks to either pry open the lid or create a hole in the box (Goodall, 1964). Since Goodall's discovery, an increasing number of long-term field studies of chimpanzees have further investigated the use of tools among chimpanzee colonies and have found a widespread use of tools and even evidence of cultural differences (Whiten et al., 1999). Specifically, it seems that chimpanzees from different colonies in different countries have different "tool kits"—a term used to describe the specific repertoire of tools a chimp colony habitually uses (McGrew, 2010). For example, in Ngogo, Uganda, tools are mainly used for courtships and hygiene, whereas in the Republic of Congo, tools are used for foraging (McGrew, 2010). These instances of cultural differences in how chimpanzees engage with the physical world parallel human behaviours which are seen as socially extended cognitive processes. After all, as mentioned earlier, the bodily manipulations of e.g., a wooden pole to build a fence is not considered an example of mind extension. However, the engagement with the particular culture and the common practices of the environment *does* make the cognitive process of solving a problem by perceiving a wooden pole as a tool an extended process (Gallagher, 2013). This means that when chimpanzees seemingly have cultural differences in tool use, their specific object manipulations stem from an engagement with their common social practices which shapes their cognition and thus their behaviour. The local custom of the colony is what guides a chimpanzee to perceive a stick as, for example, a tool to "fish" for termites (Goodall, 1964), instead of a tool for hygiene purposes. The cultural practices of each individual colony are essentially shaping the cognitive processes of the chimpanzees and the ways they perceive the objects in their environment, extending their minds to incorporate the parts of the external world that allow these processes to take place.

The Arnhem Colony

Ideally, the study of chimpanzee consciousness would be carried out on wild chimpanzees. However, a number of complications arise when attempting to do so. Most notably, the access to the apes is often unreliable—fieldworkers might not see the chimps regularly, making it impossible to notice the subtleties in their behaviour (Waal 1982). Therefore, this section will primarily build upon the work of Frans de Waal and his description of the Arnhem colony in the Burger's Zoo in the Netherlands. Specifically, his book *Chimpanzee Politics* (Waal, 1982), presents a comprehensive study of the social life among a chimpanzee colony in an open-air enclosure. With a colony consisting of around 20 chimpanzees, Waal has systematically reported and interpreted the chimps' different behavioural patterns—some of which can be argued to arise from mind extended cognitive processes.

Central for life in the chimpanzee colony is the stable hierarchy that exists between the apes, which is manifested by the so-called submissive greetings (Waal, 1982). Submissive greetings are a special form of social behaviour that confirms and reinforces the colony's hierarchy in a way that leaves no room for doubt (Waal, 1982). This non-mutual behaviour is expressed by a lower-ranked chimp greeting a higher-ranked one with a series of bows. The non-mutual nature of the behaviour means that a higher-ranking chimp will never greet a lower-ranking chimp, except in times of power take-over and instability (Waal, 1982). This essentially means that, just as our legal system is comprised of laws and contracts that are the representation of mutually agreed upon decisions, the chimp hierarchy is comprised of communicative signals that are mutually agreed upon to mean either a reinforcement or an undermining of the current power structure. It is a product of shared mental processes, i.e., a mental institution.

Quarrelling Infants

From the chimpanzees' behaviour it is evident that this complex social structure governs much of the animals' cognitive processes. Just as the legal system makes otherwise 'simple' behaviours part of a larger cognitive process, incorporating the social understanding of laws and rights, the hierarchy in the Arnhem colony has the same effect. In an example presented by Waal (1982), two mother chimpanzees, Jimmie and Tepel, are sitting in the shadow with the oldest female chimp, Mama, in between them while their infants are playfighting. The playfighting evolves into a quarrel but ends when Tepel wakes Mama, points to the infants, and Mama barks loudly at them.

On the surface, this might seem like simple chimp behaviour, but, according to Waal, in order to interpret the behaviour properly, two things about the colony are important to note: (1) Mama is the highest-ranking female, and (2) conflicts between infants regularly lead to conflicts between the mothers (Waal, 1982). To avoid a conflict with the other mother, Tepel solved the problem with the fighting infants by waking the higher-ranking female to act as an arbitrator. Mama, seemingly understanding Tepel's motivation and intention, ends the fight. From a socially extended mind perspective, this behaviour is a clear example of a chimpanzee mind being socially extended. Tepel's behaviour would have been different had the social structure and norms of the colony been different, but because the colony is constituted the way it is, it effectively shapes Tepel's cognition to incorporate the social understanding of motherhood, conflicts, and the chimps' individual rank. It creates the coupled system between her mind and the hierarchy—which as previously noted essentially functions as one of Gallagher's mental institutions. And this instance is not just Tepel being an especially clever chimpanzee. The higher-rank female, Mama, also understands what she is supposed to do when Tepel wakes her and points to the infants. Mama's behaviour is equally shaped by the social structure and norms of the colony when Tepel's pointing makes it clear that she should be the one to stop the fight in order to prevent further tensions between the two mothers.

A Social Faux Pas

Another example presented by Waal (1982) also involves Tepel. Here, Tepel silences her child, Wouter, by placing a hand over his mouth to stop him from screaming and engaging in a quarrel with another chimp. On another occasion, a different mother also silences her child to stop them from barking at and starting a conflict with a dominant male ape. According to Waal, to properly understand this behaviour, it is important to note that in the colony, when two infants are having a noisy conflict for too long, a chimp male will come to stop the conflict. However, in such cases, the child usually retreats to their mother who then becomes the receiver of the male's punishment. When looking at Tepel's behaviour from an extended mind perspective, the social norms and customs are essentially shaping how she perceives the situation. Realizing she can avoid a beating if her child stops being so noisy, she comes up with the solution of silencing him. This solution is tied to her previous experiences and would have been different had the social hierarchy and norms of the colony been different. A similar thing is true for the second instance of silencing which Waal interprets as a mother's reaction to a "social faux pas committed by the child" (Waal, 1982, p 48). This idea of a "social faux pas" in a chimpanzee colony emphasizes how complex and intricate chimpanzee social life can be – and how much of chimps' behaviour and cognition are shaped by this mental institution. It allows for what Waal describes as a "*social* application of reason and thought" (Waal, 1982, p 51), or, in other words, an example of a socially extended cognitive process.

What is it Like to Be a Chimpanzee?

Based on this evidence, it appears that chimpanzees' complex social behaviour arises from some sort of socially extended mind. And as mentioned earlier, the prerequisite for mind extension is some sort of consciousness. In this section, I will discuss what that means for our understanding of both chimpanzee and human consciousness.

Why the Research Method Matters

The current paper builds upon primarily anecdotal accounts of chimpanzee behaviour, which has been the most prevalent approach to studying chimpanzees since Dr. Goodall broke ground in the '60s.

More recently however, the field has shifted towards a predominantly experimental approach (Povinelli & Vonk, 2003). The experimental approach seeks to eliminate the problems that arise when looking at anecdotal evidence, namely, the problems of objectivity. As Waal (1982) also discusses, one of the main problems with using anecdotal evidence is that it is never absolutely certain that the interpretations are truths. Researchers must rely on intuition and learned associations between different behaviours that make up patterns of interaction between the chimps (Waal, 1982). Even though it is possible to computationally calculate which behaviours occur together frequently, it is problematic to rely on our human view and interpretation of animal behaviour as humans are notoriously known for anthropomorphizing. We ascribe even simple, geometric figures moving about a computer screen with intentions and desires (Heider & Simmel, 1944), just as we would do in a social situation with another human (Atherton & Cross, 2018). It is obvious that inferring mental states, and thus consciousness, in geometric figures on a computer screen is nonsensical. As such, it seems natural to both be wary of attributing any kind of humanization to animals (Waal, 1982), and to have a desire to move towards a more objective and neutral study method.

However, "because the most fundamental problem associated with the use of anecdotes was never widely identified, the same conceptual problem has crept, almost unnoticed, into our experiment" (Povinelli & Vonk, 2003, p. 158). What Povinelli and Vonk argue is that the same problems with inferring mental states from behaviour arise even in experimental settings, as it is often possible to explain the chimpanzees' behaviour purely based on behavioural abstractions. Instead, to avoid this problem, the experimental setups must provide a truly novel situation for the chimps (Povinelli & Vonk, 2003). Even though experimenters are trying to do so

(Tomasello et al., 2003), I will argue that experiments will still fall short of truly studying chimpanzee experiences and mental states. Because another of the main problems with both the anecdotal and the experimental approach is our undeniably human perspective. We cannot escape the fact we are humans trying to understand a different animal entirely. And just as we are unable to imagine what it is like to be a bat (Nagel, 1974), our experiments will always be tainted by the fact we are humans on the impossible task of trying to imagine what it is like to be a chimpanzee. This may sound exaggerated, but the mere fact that in studies about chimpanzees, researchers use videos of humans (Premack & Woodruff, 1978) and human experimenters (Povinelli & Vonk, 2003), shows a clear bias against taking the animals' point of view in favour of our own human point of view. To return to Nagel, this becomes problematic as we can never know what it is like to be something non-human without first taking up their point of view (Nagel, 1974). So, instead of trying to create experimental setups that remove the chimps from their natural settings, one must first take up the chimp's point of view. I will therefore argue that looking at how chimpanzees engage with their environment, and especially the other chimpanzees in their colony—as that evidently plays a central role in chimpanzee life—takes us closer to taking the chimp's point of view, and thus closer to a more accurate understanding of the chimpanzee experience. Even though anecdotal evidence has its limitations, the social nature of chimpanzee colonies makes the anecdotal approach justifiable.

Apes Are Conscious, So What?

Ascribing consciousness to chimpanzees will, for most, probably seem like the most natural thing to do. After all, they are very similar to us, sharing a whole 99% of our DNA (Gibbon, 2012). What I will discuss in this final section is why this analysis of chimps should lead us to change how we think about consciousness and how we approach the question of non-human consciousness.

As Nagel discusses in his paper, *What Is It Like to Be a Bat* (1974), when discussing whether a bat has conscious experiences, it must be argued whether "there is something that it is like to be that organism" (Nagel, 1974, p. 436). If it is like to be a bat in some way, then that something is an experience that arises from some sort of conscious awareness of the experience. So, when trying to understand whether chimpanzees are conscious, it is first a question of whether it is like to be a chimp in some way. This puts us in much the same position as an alien would be in if it tried to conceptualize the human experience (Nagel, 1974). With this perspective, an interesting thought problem arises: If aliens had the same knowledge and evidence of us humans as we have of chimpanzees, would it be justified for the aliens to *not* think of us as conscious beings with mental states? If aliens had systematic reports of humans engaging in complex social behaviour within a clear hierarchical social structure, would a dismissal of mental states as well as a sceptical account of pattern recognition and behavioural associations be welcomed? I imagine most would think it wrong of the aliens to make this conclusion, because *we know* what it is like to be us. *We know* that there is more to humans than meets the eye because we experience it every day. So, who is to say the same cannot apply to chimpanzees?

The idea of some sort of chimpanzee consciousness rests a great deal on the wording "some sort". It is impossible to know whether chimpanzee consciousness is the same as human consciousness. It is actually very likely that the two are different since humans and chimpanzees embody quite different perceptual and motor systems, and are social animals that are substantially shaped by their own environment. However, the same holds true for individual humans. I cannot possibly know that your consciousness is the same as mine, or that your conscious experience is the same as mine—again, it is probably not the same as we are both shaped by our own, and therefore different, circumstances. But that does not mean we do not both have consciousness. The main reason it is easier to infer a conscious experience in other humans than in non-human primates is first, we know what it is like to be human, and second, we have the means of communication. You can communicate your intent, emotions, and thoughts to me. Chimpanzees cannot. But disregarding animal consciousness based on this is unwarranted—non-verbal autistic people are still viewed as conscious, even if their means of communication are different. Just because chimpanzees are unable to live up to the traditional, human perception of consciousness, the conclusion that chimps are therefore not conscious is not justified. We must move beyond our own brain and body and try taking the perspective of the chimpanzees. I would argue that shifting the focus away from the individual chimp's capabilities and instead look at how behaviour in a colony arises from a complex social structure, moves us closer to the chimps' perspective and thus closer to an understanding of their internal mental states. But whether or not chimpanzees are conscious, this paper is at its center an appeal to a less human-focused perspective on consciousness and a change in the way we study the mental states of non-human animals.

References

Atherton, G., & Cross, L. (2018). Seeing More Than Human: Autism and Anthropomorphic Theory of Mind. *Frontiers in Psychology,* 9. https://doi.org/10.3389/fpsyg.2018.00528

Bentham, J. (1789). Introduction to the principles of moral and legislation. *Clarendon Press,* Oxford.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12(5),* 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Cash, M. (2013). Cognition without borders: "Third wave" socially distributed cognition and relational autonomy. *Cognitive Systems Research, 25–26,* 61–71. https://doi.org/10.1016/j.cogsys.2013.03.007

Clark, A., & Chalmers, D. (1998). The Extended Mind. Analysis, 58(1), 7–19.

Decartes, R. (1641). Mediations.

Dehaene, S., Lau, H., & Kouider S. (2017). What is consciousness, and could machines have it? *Science, 358(6362),* 486–492. https://doi.org/10.1126/science.aan8871

Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research, 25–26,* 4–12. https://doi.org/10.1016/j.cogsys.2013.03.008

Gallagher, S., & Crisafi, A. (2009). Mental Institutions. *Topoi,* 28, 45–51. https://doi.org/10.1007/s11245-008-9045-0

Gibbon, A. (2012, June 13). Bonobos Join Chimps as Closest Human Relatives. Science | AAAS. https://www.sciencemag.org/news/2012/06/bonobos-join-chimps-closest-human-relatives

Goodall, J. (1964). Tool-Using and Aimed Throwing in a Community of Free-Living Chimpanzees. *Nature, 201,* 1264–1266. https://doi.org/10.1038/2011264a0

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology, 57(2),* 243–259. *JSTOR.* https://doi.org/10.2307/1416950

Kent, L., & Wittmann, M. (2021). Time consciousness: the missing link in theories of consciousness. *Neuroscience of Consciousness,* 2. https://doi.org/10.1093/nc/niab011

Kotchoubey, B. (2018). Human Consciousness: Where is it from and what is it for. *Frontiers in Psychology, 9(567).* https://doi.org/10.3389/fpsyg.2018.00567

Lagercrantz, H. & Changeux, J.-P. (2010). Basic Consciousness of the Newborn. *Seminars in Perinatology, 34(3),* 201–206. https://doi.org/10.1053/j.semperi.2010.02.004

Libet, B. (2006). Reflections on the interaction of the mind and brain. *Progress in Neurobiology 78(3-5),* 322–326. https://doi.org/10.1016/j.pneurobio.2006.02.003.

McGrew, W. C. (2010). Chimpanzee Technology. *Science, 328(5978),* 579–580. https://doi.org/10.1126/science.1187921

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review, 83(4),* 435–450. https://doi.org/10.2307/2183914

Pfeifer, R., & Bongard, J. (2007). How the Body Shapes the Way We Think: A New View of Intelligence. https://doi.org/10.7551/mitpress/3585.001.0001

Pinker, S. (2005). So how does the mind work? Mind and Language, 20(1), 1–24. https://doi.org/10.1111/j.0268-1064.2005.00274.x

Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences, 91(16),* 7398-7403. https://doi.org/10.1073/pnas.91.16.7398

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? Trends in Cognitive Sciences, 7(4), 157–160. https://doi.org/10.1016/S1364-6613(03)00053-6

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1(4),* 515–526. https://doi.org/10.1017/S0140525X00076512

Rupert, R. D. (2004). Challenges to the Hypothesis of Extended Cognition. *The Journal of Philosophy, 101(8),* 389–428.

Searle, J. (2009). John Searle on Consciousness [Interview]. https://www.youtube.com/watch?v=WFQ0Spu50Oc&feature=emb_title

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Engeland, H.V. (2018). Theory of mind in schizophrenia. *The British Journal of Psychiatry, 191(1),* 5–13. https://doi.org/10.1192/bjp.bp.107.035899

Thelen, E. (2000). Grounded in the World: Developmental Origins of the Embodied Mind. *Infancy, 1(1),* 3–28. https://doi.org/10.1207/S15327078IN0101_02

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees versus humans: It's not that simple. *Trends in Cognitive Sciences, 7(6),* 239 –240. https://doi.org/10.1016/S1364-6613(03)00107-4

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature, 399(6737),* 682–685. https://doi.org/10.1038/21415

Waal, F. de. (1982). Chimpanzee Politics: Power and Sex among Apes (First Edition). *Jonathan Cape Ltd.*

Editorial Board

Faculty Reviewers

John Alderete

Department of Linguistics, Simon Fraser University

James Giles

Department of Psychology, Roskilde University

Margaret Grant

Department of Linguistics, Simon Fraser University

Lesley A. Schimanski

Department of Psychology, Capilano University

Matthew J. Sigal

Department of Psychology, Simon Fraser University

Kimberly Voll

Brace Yourself Games, Vancouver

Cover - Languages

Aleksandrs Baskakovs—Latvian

Andrea (Ely) Diaz—Spanish

Chi Fai Tsui—Simplified Chinese

Da Som (Clover) Kang—Korean

Ethan Dungey—French

Farrodhiyah —Malay

Hilary J. H. Tsui—ASL

Joanne Lau—Traditional Chinese

Keisei Aoki—Japanese

Kenny J. Fisker—Greenlandic

Mark Giles—Danish

Natania—Indonesian

Valeriya Zhmak—Russian

Victor Guerin—hən̓q̓əmin̓əm̓

Vishal Krishna—Tamil

Additional Support

Abhishek Parmar

Nova Islam

Julian Robbert

ISBN 978-1-77287-107-4

9 781772 871074