ISSUE 11/ 2023

KOÇ UNIVERSITY UNDERGRADUATE PSYCHOLOGY JOURNAL

For further information: visit kuupj.ku.edu.tr or email psychologyjournal@ku.edu.tr

TABLE of CONTENTS

EDITOR'S NOTE

I am excited and proud to present you with the 11th issue of KUUPJ. Our 11th issue consists of the efforts of our authors, previous editors, and advisors who worked truly hard for this issue to be published. When I looked at the outcome of this enormous work, I felt grateful to everyone contributing to this issue. Furthermore, I am especially enthusiastic about this issue since it marks my first work as the editor-in-chief and it marks the return of KUUPJ to its regular pace. With this issue, I hope KUUPJ will continue to grow and realize its goal of being a medium for amazing works of undergraduate psychology students in Turkey.

In this issue, you will find more papers than there usually were. We wanted to celebrate our return with all these wonderful papers from Boğaziçi University, Istanbul Bilgi University, and Koç University; by publishing seven papers in one issue. You can find three experimental articles and four review articles on diverse topics, from language to emotion regulation and its clinical implications. We are proud and hopeful as we see the hard work of our fellow undergraduate students. Also, it is our delight to present their works to you.

I would like to thank all authors and editors for the great work that they put forth. Also, I would like to thank our advisor, Assoc. Prof. Tilbe Göksun for her firm faith and endless support in us, and our dean Prof. Aylin Küntay for her support. In addition, I would like to thank Şeref Can Esmer, a former editor-in-chief and a current graduate advisor, for his enormous contribution and guidance through this issue. I also would like to thank our previous editors-in-chief Irem Tuncer and Onur Akın, and their editorial team for contributing to this issue. As I stated, many people put great effort into this issue. I am grateful to all of them.

I hope you enjoy reading the 11th issue!

Editor-in-chief
Taygun Yorulmaz

Working Memory: How Being Watched and Self-Awareness Affect Working Memory Performance

Barış Peçen, Derin Derelioğlu, Simge İlseven, and Su Ayhan
Koç University

Video-mediated communication has become increasingly popular for both social and professional purposes since the start of the COVID-19 pandemic. Some outcomes of this are increased exposure to one's camera view, as well as constant gaze from others at all times. The current study aims to examine the effects of being watched and self-awareness on working memory, in the context of video conferencing. As a design we created three different conditions, each participant was randomly allocated to one of these: 1) both participant's and researcher's cameras are open 2) only the researcher's camera is open 3) only the participant's camera is open. Our sample was gathered through social media and students from Koç University (n=90); and each participant solved the verbal n-back task in the Zoom application. The results demonstrated that only people who were manipulated to have heightened self-awareness (condition 3) showed a decreased performance compared to the group which had their video hidden. The negative effects of seeing one's video on working memory should be considered especially for educational uses of video-mediated communication platforms.

Keywords: being watched effect, video mediated communication (VMC), working memory

Since the COVID-19 pandemic started in March 2020 many activities have moved to digital platforms and started to be conducted online, including work meetings and school lectures. One difficulty people experience in this process is the lack of face-to-face interaction. Many lecturers prefer that students turn on their cameras during online lectures to increase participation and have the feeling that they are speaking to a person (Day & Verbier, 2021). It is also common for students to be hesitant in participating in classes with their cameras on (Day & Verbier, 2021). It remains largely uninvestigated how such preferences affect individuals' learning and engagement. More specifically, by manipulating whether one is being watched and whether one sees themselves, individuals' interaction with the presented material can be affected (Yamada & Akahori, 2009). Previous studies show that being watched by other people can hinder working memory performance (Bellettier & Camos, 2018). Similarly, seeing one's image can induce a state of objective self-awareness, and lead to poorer working memory performance (Ferrari, 2001). However, how these effects may present themselves on video-mediated platforms remains unclear. Hence, in this study, we will investigate the effects of self-awareness and being observed on working memory during video calls. Answering this question can give us a better understanding of the effects of communicating on video call platforms on social and cognitive processes. Furthermore, the results of this study can give insight into possible solutions and interventions for attentional problems people may experience during video call meetings.

Being Observed

Bellettier and Camos (2018) tested whether there is an effect of attentional catch on working memory due to the experimenter's presence to investigate the effect of being watched on working memory. They suggested that the social presence of the experimenter would capture attention which results in cognitive resources being taken away from executive functions. This is called the *distraction-conflict hypothesis*, where performance is expected to be hindered, especially for difficult tasks that require executive control (Belletier et al., 2015). In the first part of the experiment by Belletier and Camos (2018), the participants were asked to memorize random letters, with one group of participants being tested alone and another being tested with experimenters present. The reaction times were slower in the alone condition than with the experimenter, suggesting that the presence of the experimenter indeed decreased participants' recall performance. The study provides support for the *distractor-conflict hypothesis*, showing that the presence of others can affect working memory for difficult tasks. This suggests that a similar effect could be observed while being observed by others during video calls.

The effect of being watched on executive control of working memory (particularly, impulse inhibition) has been documented before (Belletier et al., 2015). The justification for these processes is grounded on the idea that people feel pressure under the conditions in which they are being observed and they worry about their performances (Belletier et al., 2015). In the study of Belletier et al. (2015), executive control tasks, (conflict task and Simon task) where participants needed to respond to the color of the stimuli while ignoring their spatial location, revealed different results in three different conditions which were "alone", "peer presence" and "experimenter presence" conditions (pp. 1412-1413). The experimenter's presence yielded the worst performance with the biggest interference for executive attention (i.e., the ability to regulate responses) and the effect was most prominent in people with the highest working memory capacity. This is because people with high working memory capacity can attend both to the task and the presence of evaluative others, making them more susceptible to interference when being watched by an experimenter during tasks that require executive attention. Another moderate evidence comes from the study of Maresh et al. (2017), where they used the Attentional Control Theory (Berggren & Derakshan, 2013; Eysenck & Derakshan, 2011), to explain the individuals' ability to select what they attend to or ignore. The Attentional Control Theory suggests the idea that anxiety tolls performance by impairing executive functions. Following this, they predicted that the participants' reaction times during the n-back test would increase but not their accuracy rates would not. 61 participants performed a visual n-back test, with either easy or hard task difficulties, under one of the three social conditions: alone, in the presence of an experimenter ("presence"), or in the presence of an experimenter tasked with evaluating the performance of the participants ("threat") as the study went on. The results indicate that the effect of evaluation by the other is reflected in reaction times but not accuracy; specifically, those who generally have a high Fear of Negative Evaluation showed significant increases in reaction times only under the threat condition and hard difficulty level (Maresh et al., 2017). Therefore, this experiment demonstrated the effect of being watched on working memory load when being observed by an experimenter.

Although the previous studies mentioned before took place in real laboratory settings, being observed can have effects on working memory in virtual environments as well. It is found that even direct eye gaze is not necessary for fear of evaluation (Colombatto et al., 2019). Colombatto et al. (2019) investigated whether the effect of being watched is exclusive to direct eye gaze or is because it indicates a "mind behind the eyes". To test this, they used mouth shapes which can also be detected as an indicator of a watcher, based on whether they are facing the participant or not. The results showed that working memory was disrupted when the mouth shapes were facing the participants. This indicates that the disruption of working memory is not specific to direct eye gaze, but to mouth shapes as well. According to the researchers, these effects can be explained by a phenomenon called 'mind contact', which suggests that the direction of other agents' attention and intentions can be inferred from a "face" and eye contact is not necessary for the effect of being watched (Colombatto et al., 2019, p. 956). Thus, there seems to be evidence suggesting that humans might be interpreting the information related to a mind as a representation of an evaluative mind. This study suggests that there is no need for face-to-face contact (or eye contact) for an observer effect to arise, the idea of being watched is sufficient to generate a load in cognitive functions. This implies that the lack of eye contact on video call platforms would not diminish the effect of being watched on cognitive load.

Self-awareness

One consequence of video calls is that most of the time when people open their cameras and see themselves while there is other information that needs to be attended to, such as the presented information on the screen and the other people on the call. In one study Hassel and Cotton (2017) aimed to investigate the effects of some characteristics of Video Mediated Communication (VMC) on the performance of participants in a task where they needed to communicate information with one another and determine a solution (an adaptation of the International Institute Task) (Hassel & Cotton, 2017). The perceptual demands of VMC on attention and memory are different from face-to-face communication, so various dynamics should be considered while designing meetings and lectures on VMC. Especially, being able to see one's image on the screen and the image being disrupted by the angles of the camera is one of the important aspects that distinguish VMC from face-to-face communication. Hassel & Cotton (2017) state that when more information is transferred through VMC such as visual or auditory cues, the performance of the individual and the team in general declines compared to the cases where such cues are absent. Furthermore, they discuss that access to self-relevant audio-visual information induces additional cognitive load that does not benefit task performance. VMC might have these hindering effects on performance based on Objective Self-Awareness and Communication Overload Theories (Hassel & Cotton, 2017). These findings suggest that additional sensory cues that are present in VMC can lead to detrimental effects on communication due to increased attention and memory demands.

Because most of the VMC applications enable users to see their image besides other people's videos, seeing oneself may lead to increased objective self-awareness (Geller & Shaver, 1976). Objective Self-Awareness Theory (Duval & Wicklund, 1972) states that conscious attention can shift its focus between the environment (subjective self-awareness) and the self (objective self-awareness). When in a state of objective self-awareness, one is more likely to engage in impression management and focus on self-presentation due to actively detecting the discrepancies between their self-expectations and their actual appearance (Duval & Wicklund, 1972). This can lead to focusing their attention on how they present themselves instead of the task that needs to be completed (Hassel & Cotton, 2017; Xu & Behring, 2014). Additionally, this shift in attention may have adverse effects on the user's working memory (Geller & Shaver, 1976). Another explanation for the disruptive effects of seeing one's image is that while more social cues are better for a smooth communication experience, the image of oneself might be distracting and disrupt the task performance (Liebling & Shaver, 1973). This increase in cognitive load might be diverting the attention to the self rather than the task, hindering concentration and performance required for the task at hand (Geller & Shaver, 1976; Liebling & Shaver, 1973).

To explain the adverse effects of cognitive load on task performance, Costley et al. (2020) distinguished between germane load and extraneous load. Germane load refers to resources that are devoted to the acquirement and automation of schemata that facilitate the transformation of information from WM to long-term memory. This enables information to be recallable for longer periods and thus, aids learning. On the other hand, extraneous load emerges from the information irrelevant to the task one is preoccupied with, thus it hinders learning and does not facilitate cognitive functions (Costley et al., 2020). Costley et al. (2020) found that in online classes, germane load correlated with strategies that created more efficient learning opportunities. Such strategies include order of content selection, navigating through the video lecture (pausing, skipping forward/backward, re-watching, adjusting video speed), splitting attention between visuals and text, looking away from the screen to listen more carefully, pausing to focus on something on the screen, and scanning the text to find specific information (Costley et al., 2020). As these findings suggest, self-view on a video conferencing call may also create extraneous load because it does not aid the task at hand, and it can create worries about self- appearance which would again create cognitive load.

The idea that viewing oneself would increase cognitive load was investigated by other researchers by creating conditions that increase self-awareness. According to Ferrari (2001), some possible factors can influence the relationship between performance and procrastination and one of them is objective self-awareness. He suggested that since objective self-awareness draws attention from tasks to the person themselves, it might remind procrastinators of their ineffective performance and deficiency in maintaining a good speed and accuracy balance. Therefore, procrastinators' performance may be affected more negatively than non-procrastinators when a state of objective self-awareness is induced. Ferrari (2001) aimed to measure the performance of procrastinators at a speed-accuracy task by either asking them to perform cognitively-taxing tasks (number recall), subjecting them to tasks aimed at increasing their self-awareness or putting them under time limitations. Participants were asked to complete a shape-matching task, also memorize an 8-digit number in between the tasks to increase their cognitive load. They found an interaction effect between speed and self-awareness, where both procrastinators and non-procrastinators performed slower in the objective self-awareness condition. Additionally, they found that chronic procrastinators in the objective self-awareness condition had lower accuracy compared to other groups. Ferrari (2001) suggested that objective self-awareness may be more hindering for procrastinators due to their higher level of anxiety (Ferrari & Dovidio, 2000) and lower self-esteem (Ferrari & Beck, 1998). This result supports the researcher's hypothesis that objective self-awareness causes more stress on procrastinators (Ferrari, 2001). It also implies that anxiety and self-esteem may have a moderating effect on the relationship between objective self-awareness and self-reflective processes. Moreover, Yamada and Akahori (2009) investigated different settings in video conferencing environments to see whether the videos being turned on or off affects learning. They focused on qualitative aspects of the learning process such as communication, and self-correction. They found that open cameras of the instructor and the learner create an environment that replicates a face-to-face learning process which results in an improvement in communication. This is because VMC provides both the instructor and the learner with social cues utilized to improve communication. In addition to replicating face-to-face contexts, self-awareness draws learners' attention to their errors during communication and their self-presentation, so that they can correct themselves (Yamada & Akahori, 2009). However, this redirected attention may negatively influence students in more demanding tasks, such as online university lectures, since students may have a greater cognitive load and have to maintain attention on the content for longer periods.

The Present Study

Previous studies show that being watched by someone lowers working memory performance (Belletier & Camos, 2018), especially in people with higher memory capacity when the experimenter is observing (Belletier et al., 2015). The effect is more likely to be observed in difficult tasks rather than easy ones, due to a fear of negative evaluation (Maresh et al., 2017). While there is no direct eye-gaze present in video calls, the effect of being watched on working memory remains intact even when there is no direct eye-gaze, and the thought of being watched by another "mind" is sufficient (Colombatto et al., 2018). In light of these findings, we hypothesize that working memory performance will be affected negatively while being watched by an experimenter on video calls.

As well as being watched by others, viewing the self can trigger objective self-awareness and disrupt concentration during video calls further (Hassel & Cotton, 2017; Ferrari, 2001) compared to no self-viewing cases (Yamada & Akahori, 2009). This is thought to be due to an increase in cognitive load because increased extraneous load (i.e., information that does not contribute to learning) hinders the working memory process (Costley et al, 2020). Because of this, our second hypothesis is that viewing the self during a video call would reduce the working memory capacity more, compared to no self-view cases. Thus, we expect camera conditions to influence the performance of participants in a working memory task.

Methods

Participants

There were 90 final participants (37 males, 51 females, and 2 others), who were all recruited from social media through convenience sampling. There were 30 participants per condition. All participants were Turkish and university students between the ages of 18-25 (M = 21.71, SD = 1.54). A young sample was chosen to control for differences in technology use that may arise due to age differences (Hecker et al., 2021). The data from one participant was discarded due to not completing the task. While 88.9% of the participants were undergraduate students, 11.1% were graduate students.

Materials

The working memory of the participants was measured using a verbal missing digit span task. The task had a total of 25 items, where each digit group (from 4 to 8 digits) had 5 items. On each item, the participant was given a string of digits verbally by the experimenter by calling each number by their order in the digit, with the standard pace of one digit per second. Then the numbers of the called string were randomly shuffled, and one digit was removed from a random position. The participant was asked to name the missing digit. The numbers were randomly generated in excel, which also generated a shuffled and missing digit version of the original number. A verbal task was used instead of a visual task so that participants only used the computer screen for Zoom and were able to see themselves. The Pearson correlations of the Missing digit span task with the Backwards Span Task and Subtract 2 Span Task were moderate with r = .26 for both with a significance at the p < .05 level (Waters & Caplan, 2003).

Procedure

Participants were assigned into three separate groups: control group (both the participant's and the experimenter's cameras off), hide self-view (both the participant's and the experimenter's cameras on, but the participant cannot see themselves), and pin self-view groups (both the participant's and the experimenter's cameras on, and the participant pins their image). They were given a Qualtrics link to read and approve the consent form. Afterward, they were asked to provide demographic information (age, gender, and education status) and their participant ID number.

All participants were asked to ensure that Zoom is in gallery mode and to use Zoom in full screen throughout the experiment. They were specifically asked not to write down the numbers or take notes in any way. They were given an example to better understand what they needed to do during the task.

The digits of each item were read as 1 digit per second. After reading the first number, there was a 2-second pause before reading the missing digit number. Their answers were written in the excel sheet, which calculated whether the answers were correct as well as the participants' total score. The first group, called the "Control Group", was asked to keep their camera off and their microphone on throughout the experiment. In the control condition, the experimenter's camera was also turned off. The second group was called the "Hide Self-view Group", where participants were asked to keep their cameras and microphone on throughout the experiment. They were also asked to choose the "Hide self-view" option from settings so that they would not see their video during the experiment. The aim was to manipulate the effect of being observed while the participants saw only the experimenters watching them and did not see themselves. The last group was called the "Pin Self-view Group", in which participants were asked to keep their cameras and microphone on throughout the experiment. They were also asked to choose the "Pin" option on their video from settings so that they would see their video larger than the experimenter's video during the experiment. The aim was to manipulate self-awareness. In both the "Hide self-view" and "Pin self-view" conditions, the experimenters' cameras were turned on.

Results

We conducted two separate independent samples t-tests to analyze the differences between the scores of the control group (condition 1) and the observed group (condition 2), and between the observed group and the self-awareness group (condition 3).

The reason why ANOVA was not used was that there were two different manipulations between condition 1 and condition 2, and condition 2 and condition 3. Between 1 and condition 2, the belief of being watched was manipulated while the perceptual input stayed the same. Between condition 2 and condition 3, the perceptual input was altered by showing their image, while the belief of being watched stayed the same. Therefore, these two variables were analyzed separately.

The results of independent samples t-test revealed that the difference between the scores of the control group (M = 15.8, SD = 4.06) and the observed (M = 17.0, SD = 3.42) group was not significant $t(56) = -1.18$, $p = .878$. Levene's test indicated that the two groups had equal variances in their mean scores (p = .400). Therefore, we failed to reject the null hypothesis, which stated that there would be no difference in working memory performance for those who are watched and those who are not. However, there was a significant difference between the scores of the observed group (M = 17.0, SD = 3.42) and the self-awareness group (M = 14.9, SD = 3.42), $t(54) = 2.16$, $p = .035$ (see Table 2). Levene's test indicated that the two groups had equal variances in their mean scores ($p = .672$). Cohen's d was 0.577, showing a medium effect size. These findings confirm our second hypothesis predicting that the working memory performance of the self-awareness group would be lower than the observed group.

Discussion

Our study aimed to examine the effect of being watched and observing the self on verbal working memory performance among university students. Previous studies showed that working memory is negatively affected by being watched (Bellettier & Camos, 2018), and viewing one's image in video-mediated communication increases one's cognitive load (Hassel & Cotton, 2017).

Our findings showed no significant difference in verbal working memory between the participants who were being observed and those who were not. While this contradicts the previous studies that showed an effect of viewing the self and being watched on working memory, the previous studies were not conducted on video calls. They also did not use a verbal working memory task, thus the variations in the task requirements may have led people to overcome the influence of an observer being present. It follows that these two factors may have caused the contrast in our findings. Another factor might be that as Belletier and Camos (2018) stated, the negative effect of being watched on performance is present in difficult tasks, but for easy or well-known tasks, being watched improves performance. So, it could be that the verbal working memory task was not difficult enough for being watched to have a hindering effect on working memory. One of the potential implications of this is that verbal working memory bypasses the effect of being observed. Previously it has been documented that a decline in the performance of Working Memory occurs through attentional resources being used both by the task at hand and being observed by the experimenter (Belletier & Camos, 2018). Thus, verbal working memory could efficiently function if the task is not difficult enough to create an attentional burden. Another implication could be that with tasks that are easy or mild in difficulty, individuals may be motivated instead of distressed. The presence of an evaluator can promote individuals' performances depending on task difficulty and expected evaluation (Maresh et al., 2017). Because being observed doesn't always yield poor performance, the different aspects of the condition such as task difficulty, individual differences, and level of evaluation must be considered in future research.

Our second hypothesis which predicted the self-awareness group to have significantly lower performance than the being-observed group was supported by the results. One of the factors that give rise to these results could be the difference between the size participants saw the experimenter on their screen. Participants who were observed had the experimenter on the larger screen compared to the self-awareness group. The experimenter's gaze may be a cue to focus on the task for the participants who are observed. On the other hand, in the self-awareness condition, participants see themselves on the big screen and this is irrelevant information about the task. While this kind of information creates a burden on the working- memory (Costley et al., 2020), the appearance of the self may result in further interference with the task because the participants' focus can be directed to themselves. Seeing the self may be a cue to focus on how they are seen from the outside leading to impression management (Hassel & Cotton, 2017). As a result, the cognitive load that is increased by impression management, and evaluating the perceptual information about the self interferes with the task performance and yields a lower performance score in the self-awareness group than in the observed group.

The results could imply that the working memory of the participants who turn their cameras on during video calls is not negatively affected by the belief of being watched when facing content that is not too challenging, but when they are observing themselves on the video feed simultaneously, the performance suffers. However, since we used a verbal working memory task, it would be wrong to generalize these findings to circumstances where video call participants are not expected to verbally respond (e.g., answering questions by speaking, making a comment, etc.).

There are several potential limitations of our study. First of all, the missing digit span task is usually used for children and is not commonly used or validated for adults (Waters & Caplan, 2003). Although the online nature of the experiment limited potential tasks that could have been used, and our choice of task was optimal considering delivery methods, it may have been too easy for the participants to observe the negative effects of being watched on working memory performance. Secondly, it is hard to ensure that everyone who participated in the study followed the zoom instructions correctly. Even though directions were confirmed in every step, since we could not see the participants' screen, we could never be completely sure. Another limitation is that individual differences (such as anxiety, self-image, etc.) may have affected how three conditions affect working memory performance, which we did not investigate in our study. Also, since the participants joined the experiment in their homes, their environment could not be controlled and some external factors may have served as confounding variables (e.g., noise, family members interrupting or distracting, someone else watching them in the room, etc.). Another limitation was that we did not control for perceptual differences between the conditions. Since the visual stimuli are different in all three conditions, this may have also influenced the participants' cognitive load.

For future research, we propose several ways to further investigate the topic. One is using non-verbal working memory tasks since the verbal expression is not very common in students who participate in online classes. This could also resolve the issue of verbal working memory tasks being easy for adults. With a more difficult and ecologically valid task, the effect of the observer must threaten the performance of participants. Additionally, some other cognitive abilities such as sustained attention, attention span, and long-term memory are also required for effective learning to take place throughout the lectures. Attention is a possible mediating factor between the effects of being observed and seeing oneself because it has been documented that attentional resources suffer in both of these conditions (Belletier et al., 2015; Yamada & Akahori, 2009). Thus, it is important to investigate how other cognitive capacities are influenced by online communication platforms. Another suggestion for future research is to consider individual differences and see if factors such as self-image, anxiety, and social anxiety influence performance. As reported by Maresh et al. (2017), social anxiety tied to the fear of negative evaluation influences performance in harder-working memory tasks. Likewise, Belletier et al. (2015) discuss the lack of insight into individual differences in this area. According to their research, the difference in individuals' working memory capacity also affects how being observed alters task performance. Therefore, investigating individual differences can reveal a more complete picture of the complicated relationship between performance and conditions of heightened self-awareness or being evaluated.

In conclusion, our study examined the effects of being watched and self-awareness through video-mediated communication systems. The study included three different groups and the results showed that there was no significant effect between being observed and not being observed. However, our findings showed that there is a significant decrease in performance from when participants observed themselves, thereby increasing self-awareness. Nonetheless, any effect of video-mediated communications remains to be explored. To advance the investigation, other factors that affect performance on video-mediated applications might be considered and different working memory tasks can be applied.

References

Belletier, C., & Camos, V. (2018). Does the experimenter's presence affect working memory? *Annals of the New York Academy of Sciences*, *1424*(1), 212-220. doi:10.1111/nyas.13627

Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: Being watched by the experimenter reduces executive attention. *Psychonomic Bulletin & Review, 22*(5), 1410-1416. doi:10.3758/s13423-015-0804-9

Colombatto, C., van Buren, B., & Scholl, B. J. (2018). Intentionally distracting: Working memory is disrupted by the perception of other agents attending to you — even without eye-gaze cues. *Psychonomic Bulletin & Review*, *26*(3), 951–957. https://doi.org/10.3758/s13423-018-1530-x

Costley, J., Fanguy, M., Lange, C., & Baldwin, M. (2020). The effects of video lecture viewing strategies on cognitive load. *Journal of Computing in Higher Education*, *33*(1), 19–38. https://doi.org/10.1007/s12528-020-09254-y

Day, J., & Verbiest, C. (2021). Lights, camera, action? A reflection of utilizing web cameras during synchronous learning in teacher education. *Teacher Educators' Journal*, *14*, 3–21.

Duval, S., & Wicklund, R. A. (1972). *A theory of objective self-awareness* (Vol. x). Oxford, England: Academic Press.

Ferrari, J. R. (2001). Procrastination as self-regulation failure of performance: Effects of cognitive load, self-awareness, and time limits on 'working best under pressure'. *European Journal of Personality, 15*(5), 391-406. doi:10.1002/per.413.abs

Ferrari, J., & Beck, B. (1998). Affective responses before and after fraudulent excuses by academic procrastinators. *EDUCATION*. *118*. 529-537.

Ferrari, J. R., & Dovidio, J. F. (2000). Examining behavioral processes in indecision: Decisional procrastination and decision-making style. *Journal of Research in Personality, 34*(1), 127–137. https://doi.org/10.1006/jrpe.1999.2247

Geller, V., & Shaver, P. (1976). Cognitive consequences of self-awareness. *Journal of*

Experimental Social Psychology, *12*(1), 99e108. https://doi.org/10.1016/0022-1031(76)90089-5.

Hassell, M. D., & Cotton, J. L. (2017). Some things are better left unseen: Toward more effective communication and team performance in video-mediated interactions. *Computers in Human Behavior*, 73, 200-208. doi:10.1016/j.chb.2017.03.039

Hecker, I., Spaulding, S., & Kuehn, D. (2021). Digital skills and older workers. *Urban Institute*.

Liebling, B. A., & Shaver, P. (1973). Evaluation, self-awareness, and task performance. *Journal of Experimental Social Psychology*, *9*(4), 297e306. http://dx.doi.org/10.1016/0022-1031(73)90067-X.

Maresh, E. L., Teachman, B. A., & Coan, J. A. (2017). Are you watching me? Interacting effects of fear of negative evaluation and social context on cognitive performance. *Journal of Experimental Psychopathology*, 8(3), 303-319. doi:10.5127/jep.059516

Sweller J., van Merrienboer J. J. G., & Paas F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*. *10*, 251–296 doi:10.1023/A:1022193728205

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers,35*(4), 550-564. doi:10.3758/bf03195534

Xu, Q., & Behring, D. (2014). The richer, the Better? Effects of modality on inter-cultural virtual collaboration. *International Journal of Communication*, *8*(0), 22.

Yamada, M., & Akahori, K. (2009). Awareness and performance through self- and partner’s image in videoconferencing. *CALICO Journal*, 27(1), 1-25. doi:10.11139/cj.27.1.1-25

Berggren, N., & Derakshan, N. (2013). Attentional control deficits in trait anxiety: Why you see them and why you don’t. *Biological Psychology*. https://doi.org/10.1016/j.biopsycho.2012.03.007

Mood and False Memory: The Effect of Task Type

Arman Karamanlıoğlu, B. Simay Üner, İrem Türkmen, and Meldanur Çetin
Koç University

How accurately do people remember under the influence of different moods? False memory is a phenomenon where an individual produces inaccurate or missing information. In this study, we explored how mood (i.e., positive or negative) and memory task type (i.e., recognition and recall) affect the memory processes of people in an online experiment (*N* = 148 adults). Participants' mood was manipulated by music pieces before asking them to retrieve lists of words by either a free recall or a recognition task. Results revealed that the number of critical lures in the recognition task was significantly higher than in the free recall task. However, positive and negative mood groups did not significantly differ from each other in terms of the number of critical lures. Our findings did not support the affect-as-information hypothesis possibly due to uncontrolled experimental conditions. To conclude, people tend to retrieve a false memory when there is a cue related to that information in the environment (i.e., recognition), suggesting an effect of task type on false memory creation.

Keywords: mood, affect-as-information, free recall, recognition, DRM task

Are we able to remember everything correctly? People's cognition is formed both by senses and interpretations of the situations, meaning that perception and memory processes are in constant interaction (Neisser, 1976). Because many internal and external factors affect how people perceive the world, memories of any moment can be recreated differently even though they the same situation is experienced (Ofen & Shing, 2013). People are generally certain about what they remember, but the contradictions in memories of the same moments led researchers to investigate memory in a more careful manner in terms of accuracy. In this study, we investigated the relationship between mood and false memory, and the differences of false memory creation between the task types (i.e., free recall and recognition) which are used to measure false memory.

False Memory

False memories occur when people remember events that did not happen or recall the events differently than how actually happened (Knott & Thorley, 2014). In other words, it is the distorted or fabricated perception of an event. For example, sometimes people think they locked the door before going out although they did not. False memory phenomenon has been grabbing the attention of autobiographical memory researchers for a long time. The first study investigating false memory phenomenon was conducted by Loftus and Palmer (1974), successfully showing the integration of one's own perception and external stimuli in the creation of false memories. Following the first experimental work on false memories, researchers focused on false memory in a more detailed way, for instance, the processes such as encoding and retrieval, influences of external and internal elements on memory, developing possible methods to detect false memories (Jacoby & Brooks, 1984; Wilson et al., 1975).

Before investigating memory, it was required to develop some measurement techniques. Researchers worked on several different paradigms to measure different extents of memory, examples of which are Sternberg's short-term memory scanning paradigm for serial processing (Sternberg, 1966), Rivermead Behavioral Memory Test for the assessment of everyday memory (Wilson et al., 1985), or DRM task for the immediate recall (Deese, 1959). Deese/Roediger – McDermott (DRM) paradigm is the most used task in the false memory literature. In the DRM paradigm, participants are shown a list of words where all the words are associated with a single not presented word called *critical lure*. For instance, participants are shown the list of words including cigarettes, cigars, and chimneys which should activate the non-presented main theme critical lure which is *smoke* (Knott & Thorley, 2014). After the participants see or hear those words, they tend to recall or recognize the critical lure as if it was on the list.

Also, if participants are asked to make specific decisions between remember-know specifications to the critical lures (a “remember” response shows participants can re-experience the presentation process of the word list, and a “know” response shows participants consider an item as familiar but cannot recall its presentation), they usually remember responses. Therefore, DRM paradigm creates vivid and bright false memories (Knott & Thorley, 2014). The possible explanation of the vivid false memories, according to Storbeck and Clore (2011), is that semantic activation during the encoding of the words presented in the DRM test. Then, the automatic process of semantic activation brings to mind the words that are related to ones in the list, such as the critical lure which are not presented to the participants. When the critical lure is brought to mind, participants may misattribute this experience as if they heard or saw the word during the encoding process. However, activation of critical lures does not guarantee that the critical lure will be retrieved in the recall process. Effective monitoring might intervene and prevent the critical lure from being remembered (Storbeck & Clore, 2011).

In addition to factors related to retrieval, false memory creation could depend on how information is encoded. The encoding process can be carried out in two different ways: item-specific processing and relational processing (Hunt & Einstein, 1981). Item-specific processing is encoding the items by their distinctive features. On the other hand, relational processing involves encoding the items in the list in relation to the other concepts that are already existed in one’s memory. It was argued that relational processing yields more critical lures than item specific processing (Storbeck & Clore, 2011). Also, it was remarked that when the participants in the DRM test engage in item-specific processing, it harms the relational processing. Therefore, under such circumstances, critical lures are less likely to be recalled (Arndt & Reder, 2003).

Relatedly, in DRM tasks, researchers usually use either a free recall task or recognition task to analyze critical lures. In both tasks, lists of words are presented to the participants. After the presentation, they are expected to remember the words that they have seen from the list. The difference between free recall and recognition task is that participants are given a blank sheet in the free recall task and asked to write down the words they remember from the list. On the other hand, in the recognition task, participants are given a list of words including the presented words, non-presented words, and critical lures, then participants are asked to mark the words that they think they see on the list (Deese, 1959). People’s memory performance might be enhanced or impaired because of the nature of the memory task. For example, an increase of effective monitoring in the free recall task would lead individuals to discriminate between presented and non-presented words better, hence, may reduce the number of the recalled critical lures. On the other hand, being exposed to the critical word on the recognition task may impair the correct decision (Smith et al., 2008).

Affect-as-information Hypothesis

People’s affective states influence their cognitive abilities. For instance, experiencing different mood states may boost or impair decision making, problem solving, memory, and attention (Knott et al., 2014). Many studies examined the effects of mood on cognitive abilities. For example, positive mood enhances creativity whereas negative mood enhances the performance on spatial tasks but impairs the performance on verbal tasks (Gray, 2001; Isen, 1999). More specifically, McDermott et al. (2001) showed that how a moderate level of stress helps improve spatial learning, suggesting that the effect of mood differs according to the tasks.

The effects of mood on false memories have become a major focus recently (Knott et al., 2014). How affect influences cognitive functioning is explained by the ‘affect-as-information’ hypothesis. This hypothesis emphasizes that feelings provide an information source about the tasks, objects, and decisions and influence how input is encoded. In particular, affect plays a task-relevant feedback role and helps people to determine how the information will be processed which is either item-specific or relational processing (Storbeck & Clore, 2005). Fiedler (2001) asserts that positive affective cues strengthen relational processing and lead individuals to process information in relation with the existing input. On the other hand, negative affective cues lead to item-specific processing (Storbeck & Clore, 2005).

The process in which "affect" becomes important to influence false memory is an important question. Storbeck and Clore (2011) investigated whether inducing mood right before the encoding process or retrieval process influences false memory in a recognition task, using the DRM paradigm, results of which supported the affect-as-information hypothesis. Because the memory process is explained by semantic activation (Roediger et al., 2001), negative mood induced right before the encoding processes promotes item-specific processing and decreases the activation of critical lures (Storbeck & Clore, 2011).

Herein, previous studies used various ways to manipulate participants' affective states to examine the role of affect in memory processes. For example, Ruci et al. (2009) separated the participants into three groups as positive mood, negative mood, and control groups and narrated different stories to different groups (except the control group). Then, the participants were asked to imagine themselves as the main character. On the other hand, in the study conducted by Knott and Thorley (2014), there were neutral and negative mood groups. Mood was induced only to the participants in the negative mood group, by showing them a video clip of the final scene of the movie "Dancers in the Dark." The participants in the neutral mood group did not watch a movie scene during the experiment. Storbeck & Clore (2005) preferred to use *Eine Kleine Nacht Musik* by Mozart for positive mood induction and *Adagietto* by Mahler for the negative group and reported those pieces were effective to manipulate mood.

Mood & False Memory

Some previous findings demonstrated a remarkable influence of affective cues on how accurately people remember events. According to Wright et al. (2005), experiencing a change in one's level of emotion triggers some changes in particular brain parts which influence, for instance, attentional resources. Since attention is strongly related to memory processes (Chun & Turk-Browne, 2007), this relation creates a need to investigate the effects of mood on memory. For example, people with depression recognize more negative and depression related words than positive and neutral words in memory tasks (Howe & Malone, 2011). On the other hand, people with positive mood recalled relatively more critical lures as they focus on general content rather than specific information. Therefore, they cannot remember the information fully accurately. However, people with negative mood retrieved less critical lures than people with positive mood because of item specific processing when the retrieval process is made by a free recall task (Storbeck & Clore, 2005).

The Present Study

In the present study, we aimed to investigate whether mood and task type influenced false memory creation of individuals. We used music pieces to induce either positive or negative mood and the DRM test to assess false memory. Depending on the affect-as-information hypothesis, we anticipated that there would be differences between the number of critical lures of people who were in positive and negative mood. Also, we expected a difference between the number of critical lures of the free recall and recognition task because recognition and free recall tasks change individuals' level of effective monitoring (Smith et al., 2008). Therefore, we expect that different levels of effective monitoring during retrieval phase would bring different levels of false memory.

Our main purpose in present study is to replicate the previous findings related to mood and false memory relationship (Storbeck & Clore, 2005) to demonstrate the vulnerability of memory. Also, there was not adequate number of research which used music pieces as mood inducers and compared the false memory rate during recognition and free recall tasks. Therefore, we decided to use both types of tasks to make a comparison between recall and recognition performance of people on false memory.

Method

Participants

We shared the link of the online experiment via our social media accounts and 235 Turkish-speaking participants gave informed consent and took part in the experiment voluntarily without any compensation. The age information of the participants was lost due to an experimental error. Among the 235 participants, 87 individuals did not complete the experiment. To increase the validity of the results, we examined the distribution of experiment completion duration (M = 1728 sec, SD = 906 sec). As a result, nine participants were excluded in the analysis part as their total durations of the experiment were found as outliers. Five participants also were excluded since they have skipped some steps of the experiment. Finally, there were 124 participants (67 females). This study was approved by the Koc University Committee on Human Research.

Materials and Measures

Mood induction

Music pieces were used to manipulate the mood of participants. “Eine Kleine Nachtmusik” by Mozart was used to induce a positive mood, and “Adagietto” by Mahler was given to the negative mood group. Both of the groups were exposed to music pieces for 5 minutes. The effectiveness of these music pieces on mood was shown by previous research (Niedenthal & Setterlund, 1994).

Measures of False Memory

The Deese, Roediger and McDermott (DRM) paradigm was performed to measure the false memory creation of participants in different mood groups. The lists consisting of semantically related words were presented to subjects during the encoding phase. The critical word that was not presented but closely related to presented words was expected to be retrieved by a recall or a recognition task (Knott & Thorley, 2014). In the current study, 6 wordlists were presented to all groups of participants in random order. The wordlists included 12 semantically related words, e.g., “speed”, “red”, “vehicle” (for critical word “car”).

Since the study was conducted in Turkish, the wordlists were taken from Türkçe Kelime Normları [Turkish Word Norms] (Tekcan & Göz, 2005). The neutral wordlists were carefully selected by researchers to eliminate possible mood induction effect as an extraneous variable. Half of the wordlists were randomly chosen and matched to the free recall task, and the remaining half were asked to be retrieved by a recognition task. All participants completed the same type of memory task for a given critical lure (e.g., free recall task after the ‘car’ wordlist).

Positive and Negative Affect Schedule (PANAS)

To assess participants’ current positive and negative affect, we used the Positive and Negative Affect Schedule, which consisted of 20 items (Watson et al., 1988). The subjects rated 10 positive and 10 negative adjectives between 1-5 in a Likert scale according to how they feel at that moment. Higher ratings indicate that the given adjective is a better representative of the participants’ current mood. As participants were native Turkish speakers, we used the Turkish version of the scale which is adapted by Gençöz (2000). Positive and negative affectivity scores of PANAS scores are calculated separately.

Beck Depression Inventory

Beck Depression Inventory is a frequently used self-report questionnaire that consists of 21 multiple choice questions assessing the severity of depression symptoms (Beck et al., 1961). Participants were asked to choose the best fitting answer (or multiple fitting answers if apply) for each question which indicates different levels of one depression-related symptom. Each level is matched with a number between 0-3, 3 representing the most severe condition in a question. Since all participants are Turkish speakers, a translated version of the Beck Depression Inventory is used in the survey (Kapci et al., 2008). One of the questions is eliminated from the calculation of the total depression score since one of its choices was mistranslated but mistakenly presented in the questionnaire.

Procedure

We reached volunteer healthy adults who did not have any hearing problems nor difficulty in reading and understanding Turkish by announcing the experiment on our social media accounts with a standardized pre-message. Participants reached the online experiment via the Qualtrics link in the announcement. We requested them to fill out the study in a quiet room and put on headphones if possible, but no music presentation is mentioned considering the control group. Participants were also asked to complete the experiment without taking a break.

We used a 3x2 mixed factorial design in our study. Participants were randomly assigned to one of the three conditions: positive mood group, negative mood group, and control group. Before any mood induction, participants were informed about the music presentation and asked to focus on the music by closing their eyes and try to avoid any kind of distraction such as checking their phones and talking to somebody. Participants moved on to the next page where they could start the music, once they were ready. In the study, participants in the positive mood group listened to "Eine Kleine Nacht Musik" by Mozart and the negative mood group listened to "Adagietto" by Mahler for five minutes whereas the control group did not listen to any of the music pieces. Following the manipulation step regarding their group, all participants started the experiment by filling out the PANAS. Then, six wordlists were presented to participants in a random order to eliminate the order effect, regardless of their corresponding memory task.

In free recall task steps, since each free recall list had 12 words related to the critical lure and we displayed a word for 1s on the screen, the presentation lasted 12s. Words in each list were shown in the same order to all participants. Then, participants answered math questions in a 6-question-filler task which lasted approximately 30s. Afterwards, we asked participants to write the words they remembered from the list presented. We gave 60s for free recall, and the total time was approximately 2 minutes for each free recall list.

In recognition steps, since each recognition list had 12 words related to the critical lure and we displayed the word for 1s on the screen, the presentation lasted 12s as it happened in each free recall task. Again, each participant solved a filler math task for approximately 30s. Next, we presented 30 words to the participants on one page, the presentation order of given words was randomized for each recognition task and for each participant. Here, 5 words from the list (those were the same words for all participants - i.e., word1, word5, word6, word9, word11), the critical lure for the presented list, 8 related but non-presented words, 8 words that may be associated to the critical lure and 8 unrelated words were given to the participants to select the ones presented before. The duration of the recognition task was 60s, and the total duration of a recognition step was around 2 minutes.

After all of the tasks were completed, the participants again filled out the PANAS along with Beck Depression Inventory. The depression scale was given at the end to avoid any kind of mood change in participants.

Results

We made a comparison between PANAS 1 (before the memory task) and PANAS 2 scores (after the memory task) to detect whether the mood of participants after the mood induction remained the same until the end of the experiment. *T*-test for paired samples revealed that PANAS 1 positive score of subjects (M = 31.1 SD = 7.9) differ significantly from PANAS 2 positive score (M = 27.6, SD = 9.1), $t(123) = 5.01$, $p < .001$. Moreover, there was a significant difference between PANAS 1 negative scores of participants (M = 18.0, SD = 7.5) and PANAS 2 negative score (M = 16.3, SD = 6.5), $t(123) = 3.04$, $p = .003$ (see Table 1). These results showed that the mood of subjects changed statistically during the experiment. Further analyses were made by scores of PANAS 1.

We conducted one-way between-subjects ANOVA to examine whether our mood induction by music pieces had an effect. PANAS 1 positive scores were found to be significantly different among groups $F(2, 121) = 7.77$, $p < .001$. Post-hoc analyses with Tukey correction revealed that positive mood group (M = 34.8, SD = 7.4) scored higher than negative group (M = 31.2, SD = 8.1), p =

.105 and control group (*M* = 28.3, *SD* = 7.1), $p < .001$.

Similarly, one-way between-subjects ANOVA was carried out with negative PANAS 1 scores and we detected that there was a significant difference between groups, $F(2, 121) = 13.2$, $p < .001$. According to the post-hoc tests with Tukey correction, the control group (*M* = 21.6, *SD* = 8.3) scored significantly higher on the negative score than the positive group (*M* =13.9, *SD* = 4.7), $p < .001$. Also, PANAS 1 negative score of the positive mood-induced group was statistically lower than that of negative group (*M* = 17.9 *SD* = 6.6), $p = .01$.

As the scores of the control group showed, they had negative mood more than expected, which demonstrated our manipulation did not have an effect. Therefore, we excluded some participants who were not affected by the manipulation by considering the PANAS 1 scores of the participants. By using the median split approach (Aziz et al., 2010) we detected and excluded participants whose PANAS 1 scores were not adequate for the mood group that they belong to. Participants whose PANAS 1 positive score were higher than the median positive score (32) of the sample (*N* = 124) were included in positive group. The same procedure was applied to the negative group. The sample median of PANAS 1 negative score was 12. The control group remained the same. Finally, there were 92 participants in the sample, 25 of those in the positive group, 21 of those in the negative group and 49 of those in the control group.

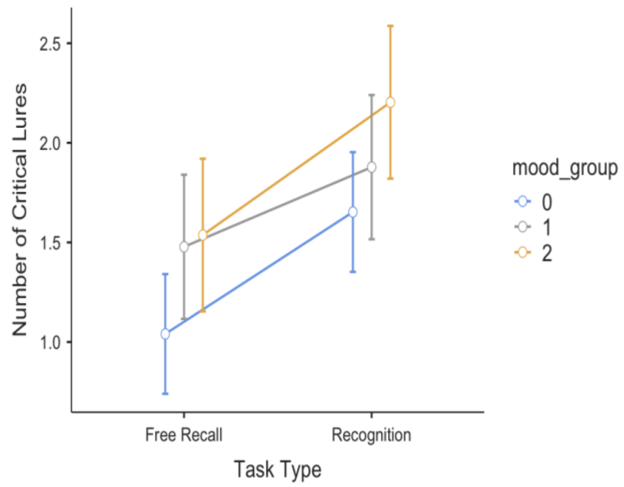
After we finalized the groups in terms of mood scores, we conducted a one-way between-subjects ANOVA, which revealed that there was a statistical difference between PANAS 1 positive scores of the experimental groups, $F(2, 46.9) = 44.1$, $p < .001$. Post-hoc comparisons using the Games-Howell test indicated that the positive group (*M* = 39.2, *SD* = 3.6) had statistically higher PANAS 1 positive score than that of the control group (*M* = 28.3, *SD* = 7.1) and that of the negative group (*M* = 29.0, *SD*= 7.7), $p < .001$. There was no statistical difference between the positive PANAS 1 scores of the control and negative mood groups, $p > .05$.

We also conducted a one-way between-subjects ANOVA, which revealed that there was a statistical difference between PANAS 1 negative scores of the experimental groups, $F(2, 51.8) = 40.9$, $p < .001$. Post-hoc comparisons using the Games-Howell test indicated that the positive group (*M* = 12.4, *SD* = 3.4) had statistically lower PANAS 1 negative score than the negative score of the control group (*M* = 21.6, *SD* = 8.3) and that of the negative group (*M* = 22.6, *SD* = 5.22), $p < .001$. There was no statistical difference in terms of negative PANAS 1 scores between the control and negative mood groups, $p > .05$.

A 3x2 mixed ANOVA was conducted to examine the effect of task type (i.e. free recall and recognition) and mood group (i.e. control, positive, and negative) on the total number of critical lures. Task type (i.e., recognition and recall) had a main effect on the total number of critical lures, $F(1, 92) = 18.423$, $p < .001$, $\eta^2_p = .167$. Post-hoc comparisons revealed that the total number of critical lures of free recall task (*M* = 1.4, *SD* = 0.9) was significantly lower than that of the recognition task (*M* = 1.9, *SD* = 0.9), $p < .001$ (see Figure 1).

Figure 1.

The Interaction of Mood and Task Type on Number of Critical Lures

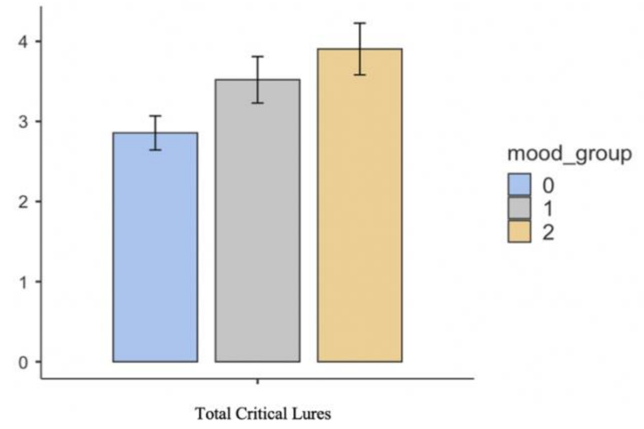


Note. 0: Control Group 1: Positive Group 2: Negative Group

There was also the main effect of mood group (i.e., negative, control, and positive) on the total number of critical lures, $F(2, 92) = 4.22$, $p = .018$, $\eta^2_p = .084$. The total number of critical lures of the negative mood group (*M* = 3.9, *SD* = 1.5) was significantly higher than that of the control group (*M* = 2.8, *SD* = 1.5). The positive mood group (*M* = 3.5, *SD* = 1.5) did not differ significantly either from the negative mood group and the control group, $p > .05$ (see Figure). There was no interaction between the effects of task type and mood group on the total number of critical lures, $F(2, 92) = .355$, $p > .05$, $\eta^2_p = .008$).

Figure 2.

The Mean and SD of Mood Groups in terms of Number of Critical Lures



Note. 0: Control Group 1: Positive Group 2: Negative Group.

A 3x2 mixed ANOVA was conducted to examine the effect of task type and mood group on the total number of errors. Task type had a main effect on the total number of errors, $F(1, 92) = 152.777$, $p < .001$, $\eta^2_p = .624$, such that the total number of errors of free recall task ($M = 3.3$, $SD = 2.5$) was significantly lower than that of recognition task, regardless of their experimental group ($M = 6.7$, $SD = 4.4$), $p < .001$ (see Figure 3). Neither the main effect of mood group nor its interaction with task type on the total number of errors is significant, $F(2, 92) = .989$, $p > .05$, $\eta^2_p = .021$, $F(2, 32) = .373$, $p > .05$, $\eta^2_p = .008$, respectively.

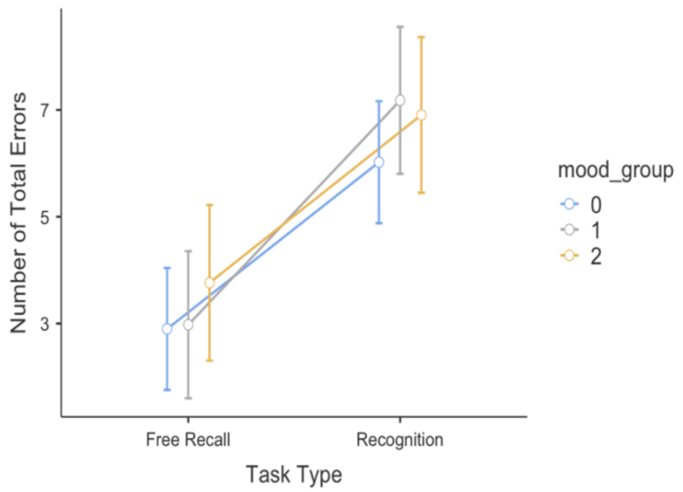
As Beck's depression score was taken as a covariate variable in 3x2 mixed ANCOVA, there was no still significant difference between the effect of mood group and task type on the total number of critical lures, $F(1,91) = .807$, $p > .05$, $\eta^2_p = .009$.

Discussion

The purpose of this study was to investigate the effect of mood (i.e., positive and negative) and the task type (i.e., free recall and recognition) on the creation of false memories. By considering the previous false memory research on the affect-as-information hypothesis (Storbeck & Clore, 2005), we hypothesized that participants in the negative mood group will differ than the positive group, in terms of the number of critical lures that are remembered. The hypothesis was based on the finding that people in negative mood encode the presented words by item-specific processing, whereas people in a positive mood process information by relational processing (Storbeck & Clore, 2011), in either a free recall or recognition task. However, our results did not support our hypothesis because we found no significant difference between positive and negative mood groups in terms of the number of critical words retrieved in both tasks. On the other hand, our hypothesis that states the false memory creation of participants will differ in two task types was supported, since the results showed that people remembered more critical lures in recognition tasks compared to free recall tasks. Similarly, to see whether this situation is different when we consider a number of errors done in either type of tasks, we examined the effect of task type on the number of errors and reported that people do not make more errors in recognition task compared to free recall task, while mood groups were not significantly different from each other.

Figure 3.

The Interaction of Mood and Task Type on Number of Errors



Note. 0: Control Group, 1: Positive Group, 2: Negative Group

While designing our study, even though we did not encounter another study in which participants filled out a *second* mood questionnaire at the end, we wanted to see whether participants maintained their mood throughout the experiment. According to our analysis, participants' moods changed during the experimental part, since PANAS 1 and PANAS 2 scores were found to be significantly different from one another. PANAS 1 scores of participants reflected the mood prior to encoding processes and our main aim was to test the affect-as-information hypothesis which states people in different moods differ in type of processing in the encoding processes using the DRM paradigm (Storbeck & Clore, 2005). Therefore, considering their difference, we decided to use only PANAS 1 scores in the further analysis since PANAS 1 scores reflected the mood right after the mood manipulation and came before the encoding processes. On the other hand, it would be better to see that participants maintained their moods throughout the experiment, since there were other encoding processes in between the times participants filled out PANAS 1 and PANAS 2.

PANAS 1 scores of participants revealed that the mood induction by music pieces were not effective enough since the positive mood scores of the control group is lower than the group who listened to the negative music piece. It may infer either that our negative mood induction was not successful, or the control group consisted of people whose mood were already negative. In addition, even though only PANAS 1 scores were included in the analyses, it is better to point out that our results revealed an expected change in the PANAS 2 scores of participants. When we examined the PANAS 1 and PANAS 2 scores, we realized that both positive and negative scores decreased in all mood groups. The reason why this change was expected is that participants spent a considerable amount of time (not more than 40 minutes) during the experiment, and the stable nature of the procedure probably made participants closer to the neutral scores. Moreover, beyond the potential effect of wordlists or tasks in the procedure, all participants were also subjected to any kind of stimulation in their environment which we were not able to control during the experiment, therefore we did not expect to see that participants successfully maintained their initial moods.

Another important point about PANAS scores is that the negative scores of the control group in both PANAS 1 and PANAS 2 are significantly higher than both positive and negative groups. Since the control group did not receive any kind of manipulation prior to the experiment, we assert that this result could be explained by the impact of type of the music on the mood of the participants. The findings of Rea et al. (2012) suggested that people who are exposed to classical music, as in our study, feel more relaxed and calmer than their previous state. Therefore, the reason why the negative mood group was overall more positive than the control group could be the effect of classical music, in contrast with the finding of Niedenthal and Setterlund (1994) that shows Adagietto by Mahler induces negative mood on people (as cited in Storbeck & Clore, 2005).

From another point of view, because data were collected via Qualtrics from the participants who were in the quarantine at their homes due to COVID-19 pandemic, our participants were likely to be in a negative mood in general. Supporting this view, Kesner and Horáček (2020) suggested that uncertainty and global threats are currently affecting people in a stressful and negative way. When the experimental groups listened to music, we think there was a decrease in their negative mood levels since the music, regardless of its type, may have become a good stimulation for participants who are in the quarantine. Hence, even the presented music was supposed to put negative group in a more negative mood, the participants were more positive than the ones who did not listen any kind of music, depending on the fact that their life in quarantine was lacking stimulation. This idea was also supported by the fact that the positive score of the control group was less than other two groups in PANAS 1. Overall, again regardless of what type of music is introduced to the participant, this manipulation affected people in a more positive way compared to no manipulation case.

Since the PANAS scores were unexpected in terms of the difference among groups, and we chose to analyze our results with PANAS 1 scores because PANAS 1 was prior to all encoding processes and followed the mood manipulation, we needed to exclude some participants from the analyses to be able to test our affect-as-information hypothesis correctly. Thus, we chose to use the median split approach rather than assigning our participants into new groups according to their PANAS scores, because we wanted to see the effect of our manipulation in line with our purpose of conducting this study. If we had grouped our participants according to their PANAS scores, the analysis would have shown only the effect of mood on memory performance, rather than showing the effect of inducing mood with music on the false memory creation. For this reason, we checked whether we were able to induce mood through music and excluded participants that their scores did not match with the criteria of the median split approach. Because the control group did not receive any type of music, no mood manipulation check was required, therefore all data of the control group was included in the further analyses.

Despite the failure in the mood induction procedure, we succeeded to support one of our hypotheses stating that different retrieval techniques (i.e., free recall, recognition) led the number of retrieved critical lures to be statistically different. As our results revealed, people performed worse in terms of false memory in recognition tasks than free recall tasks. One possible reason why people remembered different numbers of critical lures in free recall and recognition processes could be explained by the activation-monitoring framework. According to this hypothesis, whether an individual remembers an item that was actually not on the list is led by semantic activation of the critical lure during the encoding phase and lack of effective monitoring in retrieval (Oliver et al., 2016). Since the task type is related to the retrieval process, we inferred that people use effective monitoring strategies at a higher level during free recall than recognition. In other words, people control the retrieval process better in free recall tasks and do not report the critical lure intentionally even if it was activated in the encoding phase as a result of effective monitoring (Storbeck & Clore, 2005).

Furthermore, one of the reasons why the results of the main effect of mood on false memory did not meet our expectations is that we did not have control over the participants as the experiment was run through Qualtrics. Namely, as we run the experiment online, we did not know whether the participants paid enough attention to the study nor their surroundings let them focus on the lists. Another possible explanation why the results were not satisfying in terms of our hypothesis might be that the induced mood could not be maintained during the study, since the significant difference of PANAS 2 scores from PANAS 1 scores supports this explanation. The fact that participants' mood approached a normal state might have led them to perform differently in the memory tasks as we do not know at what point the participants lost the induced mood. This ambiguity about how effective the induced mood and how long it affected the participants might cause our results to be in conflict with our expectations.

Limitations

There were some limitations to the design of our experiment. During the experiment, we could not prevent the external factors that might have distracted the participants. Experiments should have been carried out in a laboratory environment rather than online. The participants saw each word for 1 second, but in the article that we referred to in our study, the participants saw the words for 250 milliseconds (Storbeck & Clore, 2005). The duration of the word presentation may have an effect on the encoding process, which may potentially result in more successful retrieval processes, decreasing the number of falsely remembered words. However, less than 1 second could not be presented to participants because the experiment was prepared on Qualtrics which did not allow us to present a word for 250ms. Also, we were planning to continue playing the music pieces during memory parts to maintain the mood induction, however, the online system again prevented us from doing that. In future studies, another program may be used instead of Qualtrics if the experiment cannot be conducted in a laboratory setting. Furthermore, the number of wordlists should have been more in further research because the reliability of score on memory tasks would increase in the parallel.

Additionally, the measure of false memory (i.e., DRM task) may have not given the best results in our study. Namely, DRM paradigm associates only one critical lure with the given list and ignores any other semantically related word when they are falsely remembered by participants. For instance, some people may falsely retrieve the word 'time' for the list which can supposed to be associated with the critical lure 'date'. In such situation, DRM task ignores this falsely remembered word by its nature. The evaluation and selection of the critical lures for a particular word list is a subjective task, and the decision of the critical lure may not be convenient for semantic network of each individual. One possible solution for this problem could be to measure false memory by considering each error (i.e., any word falsely remembered) as Murphy et al. did in their study. Relying on this, we also adopted this kind of an approach to evaluate our findings, which did not result in a change in our results. However, we posit that it is important to draw a line in between the falsely remembered words that can be associated with the presented list and the words that are completely unrelated, since we think our results may have been influenced by the measurement technique.

Conclusion and Future Implications

The present study investigated the effect of mood and task type in false memory. We could not find a difference between retrieval of critical lures of people with negative and positive mood, yet we found that task type is important in false memory creation. Although free recall and recognition tasks have been used by previous researchers, we contributed to the false memory literature by combining them in one study and supporting the task type effect. We suggest that in certain conditions such as interrogations and testimonies, retrieval techniques are needed to be chosen considering the results of our study.

References

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language, 48*(1), 1-15

Aziz, S., Wuensch, K. L., & Brandon, H. R. (2010). A comparison among worker types using a composites approach and median splits. *The Psychological Record*, *60*(4), 627-642.

Beck, A.T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current opinion in neurobiology*, *17*(2), 177-184.

Clore, G. L., & Storbeck, J. (2006). Affect as information about liking, efficacy, and importance. *Psychology Press.*

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, *58*(1), 17.

Dewhurst, S., Bould, E., Knott, L. and Thorley, C. (2009). The roles of encoding and retrieval processes in associative and categorical memory illusions. *Journal of Memory & Language*, *60*(1), 154-164. doi: 10.1016/j.jml.2008.09.002

Fiedler, Klaus. "Affective states trigger processes of assimilation and accommodation." (2001)

Gençöz, T. (2000). Positive and Negative Affect Schedule: A study of validity and reliability. *Türk Psikoloji Dergisi*.

Gray, J. R. (2001). Emotional modulation of cognitive control: Approach–withdrawal states double-dissociate spatial from verbal two-back task performance. J*ournal of Experimental Psychology: General*, *130*(3), 436.

Howe, M. L., & Malone, C. (2011). Mood-congruent true and false memory: Effects of depression. *Memory*, *19*(2), 192-201.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497-514.

Isen, A. M. (1999). Positive affect. *Handbook of cognition and emotion, 20*, 522-539.

Jacoby, L. L., & Brooks, L. R. (1984). Memory, perception and concept learning. *Psychology of learning and motivation*, 1.

Kapci, E. G., Uslu, R., Turkcapar, H., & Karaoglan, A. (2008). Beck Depression Inventory II: evaluation of the psychometric properties and cut-off points in a Turkish adult population. *Depression and anxiety*, *25*(10), E104-E110.

Kesner, L., & Horáček, J. (2020). Three Challenges That the COVID-19 Pandemic Represents for Psychiatry. *The British Journal of Psychiatry*, 1-5. doi: 10.1192/bjp.2020.106

Knott, L. M., & Thorley, C. (2014). Mood-congruent false memories persist over time. *Cognition & Emotion*, *28*(5), 903-912.

Knott, L. M., Threadgold, E., & Howe, M. L. (2014). Negative mood state impairs false memory priming when problem-solving. *Journal of Cognitive Psychology, 26*(5), 580-587.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589.

Murphy, K. J., West, R., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2007). Word-list-learning performance in younger and older adults: intra-individual performance variability and false memory. *Aging, Neuropsychology, and Cognition*, *14*(1), 70-94.

Niedenthal, P. M., & Setterlund, M. B. (1994). Emotion congruence in perception. *Personality and Social Psychology Bulletin*, *20*(4), 401-411.

Niesser, U. (1976). Cognition and reality: Principles and implications of cognitive psychology.

Ofen, N., & Shing, Y. L. (2013). From perception to memory: Changes in memory systems across the lifespan. *Neuroscience & Biobehavioral Reviews*, *37*(9), 2258-2267.

Oliver, M. C., Bays, R. B., & Zabrucky, K. M. (2016). False memories and the DRM paradigm: effects of imagery, list, and test type. *The Journal of General Psychology, 143*(1), 33–48. doi:10.1080/00221309.2015.1110558

Quillan, M. R. (1966). Semantic Memory. doi: 10.21236/ad0641671

Rea, C., MacDonald, P., & Carnes, G. (2012). Listening to classical, pop, and metal music: an investigation of mood.

Roediger III, H. L., Balota, D. A., & Watson, J. M. (2001). *Spreading activation and arousal of false memories.*

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic bulletin & review, 8*(3), 385-407.

Ruci, L., Tomes, J. L., & Zelenski, J. M. (2009). Mood-congruent false memories in the DRM paradigm. *Cognition and Emotion, 23*(6), 1153-1165.

Smith, R. E., Hunt, R. R., & Gallagher, M. P. (2008). The effect of study modality on false recognition. *Memory & Cognition, 36*(8), 1439–1449.doi:10.3758/mc.36.8.1439

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652-654.

Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science*, *16*(10), 785-791.

Storbeck, J., & Clore, G. L. (2011). Affect influences false memories at encoding: Evidence from recognition data. *Emotion*, *11*(4), 981.

Storbeck, J., & Clore, G. L. (2006). *Turning on and off affective and categorical priming with mood.* Manuscript submitted for publication.

Tekcan, A. İ., & Göz, İ. (2005). *Türkçe Kelime Normları (Turkish Word Norms)*. Boğaziçi Üniversitesi Yayınevi.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.

Wilson, B., Cockburn, J., & Baddeley, A. (1985). *The Rivermead behavioural memory test (RBMT)*. Thames Valley Test Company.

Wilson, D. L. (1975). Memory and retrieval. *Medical hypotheses*, *1*(5), 183-185.

Wright, D. B., Startup, H. M., & Mathews, S. A. (2005). Mood, dissociation and false memories using the Deese-Roediger-McDermott procedure. *British Journal of Psychology, 96*(3), 283-293

Learning at Higher Speeds: The Effect of Time Compression in Natural and Synthetic Speech on Learner Performance

Tuba P. Öztürk, Maide Körlü, Melike B. Örs, Merve D. Paker, and Güler Z. Sülün
Boğaziçi University

The shift towards online education increased the popularity of multimedia learning. It is possible to increase the speech rate in nearly all multimedia tools. As a result, students choose to consume educational multimedia at higher speeds. In this study, we investigated how listening to compressed audio alters learner performance and interacts with natural and synthetic speech. Participants were presented with educational material produced with natural or synthetic speech compressed to different speed rates. The learning performance was measured with multiple-choice questions. Results indicated no difference in the learning performance of speech type and speech speed. Nevertheless, higher speeds were more efficient, which was more prominent for natural speech. Participants also rated fast and synthetic speech as more difficult and less satisfactory. Implications of our results with the rising need for online education during the COVID-19 pandemic are discussed.

Keywords: time-compression, accelerated speech, synthetic speech, multimedia learning

The beginning of the COVID-19 pandemic marked a new era in education. The shift towards online education emphasized multimedia learning. Having recorded lectures at their disposal gave students the freedom to consume educational material when and however they prefer. Nowadays, multimedia players can speed up or slow down the media presented. Many students use this feature to speed up their lectures (e.g., Lovell & Plantegenest). This practice seemingly allows for more time-efficient learning as the accelerated speech was shown to be more efficient in non-educational material (Adelson, 1975). Our study aims to investigate the efficiency of speed-listening for educational purposes. Is it possible that students learn equally well while spending less time?

The regular speed of speech is approximately 150 words per minute (wpm), but this speech production speed does not reflect the human brain's capacity to process information (Nichols & Stevens, 1957). This is demonstrated by the fact that reading takes place at a speed of 275 wpm (Junor, 1992). Speeding up audio material, referred to as audio compression, makes it possible to increase wpm above what is possible by human speech production.

Despite its recent popularity, the audio compression is not a new technology. For example, Fairbanks et al. (1957) recruited trainees at an air force base and presented them with educational speeches on meteorological instruments compressed to different rates. They tested the trainees' comprehension through multiple-choice questions and found that learning efficiency increased until 287 wpm. Other studies showed that compressed speech is intelligible up to approximately 250 wpm (Foulke, 1968). Furthermore, the retention of compressed speech does not significantly differ from normal rate speech until 225 wpm (Barabasz, 1968). Beyond the efficiency of accelerated speech (Adelson, 1975), people adapt to compressed speech rapidly (Voor & Miller, 1965; Dupoux & Green, 1997) and prefer that (LaBarbera & MacLachlan, 1979). Specifically, university students preferred radio programs presented at %130 speed and found them more interesting, and recalled more information about faster advertisements (LaBarbera & MacLachlan, 1979). Furthermore, the audio compression is advantageous over naturally produced fast speech because there is no overlapping in sounds, and people prefer compressed speech when both are presented (Janse, 2004). These studies show that audio compression can create time-efficient, comprehensible, and preferable speech material.

The means of producing speech make up the second topic of our study. Current technology makes it possible for computer-based systems to generate speech. Speech produced this way is called synthetic speech and is heavily used in telecommunication, information services (Stevens et al., 2005), and assistive technologies for disabled people (Lebeter & Saunders, 2010). One of the most frequently used modern techniques to produce synthetic speech is text-to-speech (TTS) synthesizer (Wester et al., 2016), which is a system converting typed text input into an artificial speech immediately without necessitating any human intervention as long as the text is convertible into ASCII code (Greene et al., 1986). As opposed to synthetic speech, natural speech is the type of speech that include people's words in an everyday setting with their natural intonation and articulation.

There are still crucial differences between synthetic and natural speech despite the recent developments in synthetic speech systems (see Klatt, 1987). For example, Wester et al. (2016) found that modified synthetic speech decreases comprehension, while there was no significant difference between regular synthetic speech and natural speech. On the other hand, Lebeter and Saunders (2010) concluded that synthetic speech results in decreased levels of intelligibility when compared to natural speech in a sentence-verification test. According to Higginbotham et al. (1994), comprehension of synthetic speech is affected by speech rate, the complexity of the text, and the quality of the synthesized voice. Although time-compression is easy to access (Evans, 2008) and is commonly used today (Cheng et al., 2022), there is not enough recent research focusing on the relationship between comprehension and compression. Besides, early research adopted descriptive statistical methods (e.g., Foulke, 1968) rather than the inferential ones the present study uses. Previously, several measures of individual differences were assessed as possible contributors to the comprehension of compressed speech, such as age, working memory, vocabulary, and selective attention (e.g., Rotman et al., 2020). This measure was added because previous studies have argued that attentional resources are needed to adapt to time-compressed sound stimuli (Golomb et al., 2007).

To our knowledge, this is the first study that concerns how speech compression and speech type affect comprehension. The current research also adds to the literature the role of individual differences in comprehension of compressed and synthetic speech. As one of the possible individual differences in compressed and synthetic speech comprehension, we introduced attention deficiency by including an attention-related mistakes measure. Our study used the terms "comprehension" and "learning performance" interchangeably.

We have six hypotheses. We hypothesize that (1) when the audio material is compressed to higher speeds, comprehension will increase, and (2) efficiency will also increase. (3) We predict that natural speech will lead to a higher comprehension level than synthetic speech. (4) Since efficiency is based on the knowledge acquired in a time unit, we predict that speech type will not influence learning efficiency. (5) The hypothesized increased level of comprehension and (6) efficiency by higher speeds will disappear in synthetic speech, meaning that we predict an interaction between speech speed and speech type for both comprehension and efficiency, although we do not predict an effect of speech type on efficiency.

We presented college students with educational material in different compression levels and speech types and asked multiple-choice questions to evaluate their comprehension. We also included perceived difficulty and satisfaction measures. Since this study aims to observe how efficient using the current technology of time compression given the circumstances of online learning, we think that examining the students' perspectives on how difficult and satisfactory the experience is would contribute to the future studies regarding online learning applications. In addition to raw test scores, we calculated the efficiency of learning as several previous studies did (e.g., Fulford, 1993).

Method

Participants

There were 275 participants recruited through the Boğaziçi University's research participation system. They were offered course credit for their participation. All of them were native Turkish speakers. Twenty participants did not choose to begin the experiment. Eight participants withdrew from the experiment after being exposed to the conditions. The dropouts were from the speech conditions of natural-normal-speed (*n* = 5), natural-moderate-speed (*n* = 1), natural-fast-speed (*n* = 1), and synthetic-moderate-speed (*n* = 1). Three of them finished in more than 45 minutes (48, 70, and 104 minutes) (*M* = 15.95, *Mdn* = 14, *SD* = 8.34). Since this was an online experiment, we assumed that these participants might get distracted from the study. One participant completed the experiment in six minutes. That subject was in the natural-fast-speed condition, which lasted four minutes and 6 seconds. We assumed that the likelihood of finishing all items in one minute and 54 seconds accurately was low. Thus, we excluded these four participants from our study. Seven participants were further excluded due to an error in data coding. Consequently, data from remaining 236 participants were used in the analyses.

There were 145 females (M_{age} = 20.55, SD_{age} = 2.37) and 84 males (M_{age} = 21.27, SD_{age} = 2.42). Seven participants marked the "I reject binary gender categorization" option (M_{age} = 20.29, SD_{age} = 1.80). All participants aged between 18 and 40 (*M* = 20.80, *SD* = 2.39). While listening to the audio file, 166 subjects used in-ear headphones, 52 subjects used on-ear headphones, 4 subjects used other types of headphones, and 14 subjects did not use a headphone.

Materials and Procedure

Speech recordings were created via the content obtained from a TED-Ed video about Genetically Modified Organisms (i.e., GMOs) (see Kurzgesagt – In a Nutshell, 2017). The content was already transcribed and translated into several languages by the creators. Since our target subjects were native Turkish speakers, the Turkish translation was used. The transcription text consisted of 968 words. The same content was applied for both natural and synthetic speech. Both were accelerated via a free online speech compressor that changes audio speed without altering its pitch (see audiotrimmer.com). For the natural speech condition, a male vocalist with four years of diction and voice training was asked to read and record the translated transcription via a professional lavalier microphone Boya By-M1. The average wpm rates for the natural normal, moderate, and fast speech were 118, 180, and 236, respectively. A TTS application called VoiceDream (Chen & Rojas, 2012) was used for the synthetic speech condition, and the male voice available in Turkish in IOS was selected. The average wpm rates for the synthetic normal, moderate, and fast speech were 123, 185, and 247, respectively. The data was collected online via a questionnaire developed in Psytoolkit (Stoet, 2010, 2017). Students completed the questionnaire on the research participation system website. The questionnaire consisted of four sections. Students were asked to give demographic information (age, gender, GPA, and class) in the first section.

Participants were randomly assigned to six conditions. Each group was asked to wear a headphone while listening to the recording to minimize the possibility of any distraction caused by external sound sources. Participants were informed that they should listen in a quiet room where they would be free from any distractions. They were further informed that they could not stop or rewind the recording once it started until the ending. When the audio finished, they were required to estimate the speed of the audio file. They had to choose between options ranging from the rates of x.25 and x2 while estimating the speed of the speech material they were exposed to. Then, they answered 20 multiple-choice questions that directly assessed the information from the audio. The questions had one correct answer and three distractors. The questions tested knowledge from all parts of the audio. After completing the test, participants were required to rate an estimation up to 100 about how much they answered solely based on what they learned from the audio. We included the last part to minimize the potential effects of the previous content knowledge (i.e., GMOs) and the possibility of randomly marking an option of the multiple-choice questions without knowing the answer.

We used The Attention-Related Cognitive Errors Scale (i.e., ARCES) (Carriere et al., 2008) to measure participants' proneness to attention-related errors. The authors of the current study did the Turkish translation of the scale. The scale asks the participants to rate the frequency of their absent-minded mistakes on a 5-point Likert scale. Items include sentences like "I have gone to the fridge to get one thing (e.g., milk) and taken something else (e.g., juice).". The Cronbach's alpha revealed that the translated version of the scale was reliable (α = .89).

Finally, we asked five questions about media consumption habits. We presented a 5-point Likert scale with five items to measure the perceived difficulty and satisfaction with the audio material, where higher scores indicated higher perceived difficulty and satisfaction. The reliabilities of difficulty and satisfaction scales were not high, α = .65 and α = .61, respectively.

Results

Jamovi version 1.6 was used for data analysis (2020). We first tested whether groups differed on test performance and efficiency with two-way ANCOVAs. The independent variables were speech speed and speech type, and the dependent variables were learning performance and learning efficiency. Covariates were the estimated amount of learning from the material and the attention-related mistakes. We then compared groups based on how they perceived the difficulty and how satisfied they were with the learning process by using two-way ANOVAs. The independent variables were speech speed and speech type. The dependent variables were the perception of difficulty and satisfaction with the testing material. A Chi-square for independence test revealed that participants' predictions regarding the speed of speech were different among the conditions, $\chi^2(12, N = 236) = 209.73$, $p < .001$, $V = .67$., showing that our speed manipulation was perceptible for participants.

Learning Performance (i.e., Comprehension) and Learning Efficiency

Since the participants were not allowed to progress without marking all questions, it is possible that they randomly selected the correct option without knowing the answer. It is also possible that they marked the correct option with their prior knowledge of the topic. Our experiment was online, and it was not likely to control factors that affected the participants' attention. Subjects' prior familiarity and general attention levels might interfere with their performance on multiple-choice items. Thus, we took estimation rates of learning (M = 62.50, SD = 29.4) and attention-related mistake scores (M = 38.00, SD = 9.39) as covariates in testing the learner performance and test efficiency.

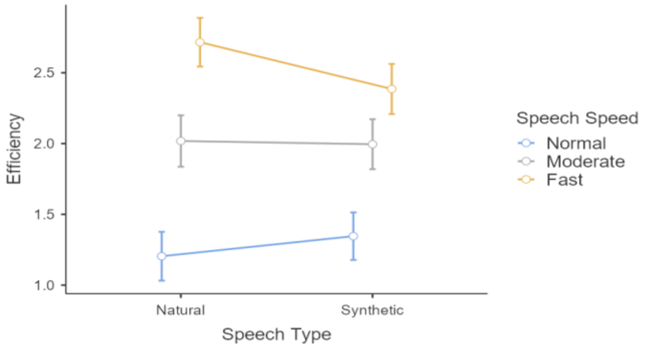
The performance of participants on multiple-choice items was calculated by the total number of correct answers. For covariates; there was an effect of learning perception, $F(1, 228) = 57.24$, $p < .001$, $\eta^2 = .19$, there was no significant effect of attention-related mistake levels, $F(1, 228) = 2.42$, $p > .10$, $\eta^2 = .01$. Contrary to hypotheses (1) and (3), there was no main effect of speech type and speech speed on the learning performance, $F(1, 228) = 3.01$, $p > .05$, $\eta^2 = .01$, $F(2, 228) < 1$, respectively. Contrary to hypothesis (5) there was no interaction, $F(2, 228) = 2.86$, $p = .06$, $\eta^2 = .02$.

The efficiency scores were calculated by dividing the total number of correct answers by the duration of the speech (Foulke & Sticht, 1969). The assumptions for normality and homogeneity of variances were violated according to Shapiro-Wilk and Levene's tests. We assumed the robustness of ANCOVA (Olejnik & Algina, 1984) alongside with visual depictions, descriptive statistics, and the sample size for each condition. There was an effect of learning perception on efficiency, $F(1, 228) = 55.77$, $p < .001$, $\eta^2 = .11$ and no effect of attention levels, $F(1, 228) = .60$, $p > .10$, $\eta^2 < .01$. Natural and synthetic speech were not different in terms of efficiency, $F(1, 228) = .95$, $p > .10$, $\eta^2 < .01$. There was a main effect of speed of speech, $F(2, 228) = 104.88$, $p < .001$, $\eta^2 = .42$. Tukey's HSD tests revealed that fast (x2) speech condition (M = 2.50, M_{adj} = 2.55, SD = .81) was more efficient than moderate (x1.5) speech condition (M = 2.02, M_{adj} =

2.01, SD = .58), which in turn is more efficient than normal (x1) speech condition (M = 1.31, M_{adj} = 1.28, SD = .42). There was also an interaction (see Figure 1), $F(2, 228) = 3.78$, $p < .05$, $\eta^2 = .02$. Simple main effects analyses showed that for natural voice, fast speech (M = 2.69, M_{adj} = 2.72, SD = .76) was more efficient than moderate speech (M = 2.04, M_{adj} = 2.02, SD = .62), and moderate speech was more efficient than normal speech (M = 1.29, M_{adj} = 1.20, SD = .42), all $ps < .001$. For synthetic voice, efficiency scores of fast speech (M = 2.32, M_{adj} = 2.39, SD = .84) were also higher than of moderate speech (M = 1.99, M_{adj} = 2.00, SD = .56) that were larger than of normal speech (M = 1.34, M_{adj} = 1.35, SD = .41), all $ps < .01$. The differences between each speed condition were significantly higher for natural voice than for synthetic voice. That is to say, the increase in efficiency from normal speed to fast speed was greater for natural speech than synthetic speech (see Figure 1).

Figure 1.

Changes in Efficiency as a Function of Speech Type and Speed Speed



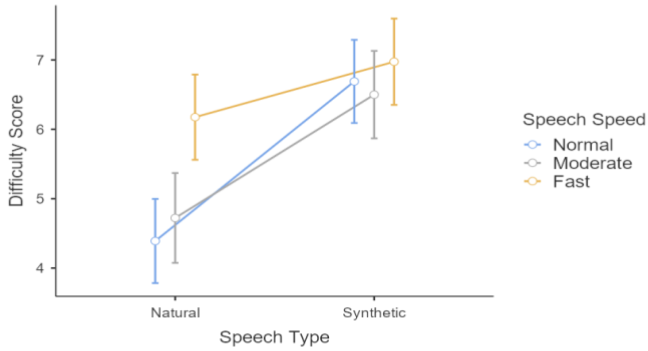
Note. The mean efficiency scores of speech types are shown for normal, moderate, and fast speech conditions (error bars reflect 95% CI for means)

Perception of Difficulty and Satisfaction

The scores of perceived difficulty (M = 5.92, SD = 2.19) were calculated by summing the responses that the participants gave to each question, which could range from 2 to 10. An analysis of two-way ANOVA showed that there was a main effect of speech type on the perception of difficulty, $F(1, 230) = 40.01$, $p < .001$, $\eta^2 = .14$. Participants rated synthetic speech (M = 6.72, SD = 2.14) as significantly more difficult to comprehend than natural speech (M = 5.10, SD = 1.93). There was a main effect of speed on the difficulty perception of participants, $F(2, 230) = 6.80$, $p = .001$, $\eta^2 = .05$. Tukey's HSD tests indicated that fast speed (M = 6.57, SD = 1.93) was rated as significantly more difficult than moderate (M = 5.61, SD = 2.26) and normal speech (M = 5.54, SD = 2.24), all $ps < .01$. There was no significant difference between normal and moderate speech. There tends to be an interaction (*see Figure 2*), $F(2, 230) = 11.67$, p = .052, $\eta^2 = .02$. Although the p value that is slightly larger than .05, due to the low effect size, we interpreted this finding to be non-significant.

Figure 2.

Changes in the Perception of Difficulty as a Function of Speech Type and Speech Speed



Note. The mean perceived difficulty scores of speech types are shown for normal, moderate, and fast speech conditions (error bars reflect 95% CI for means).

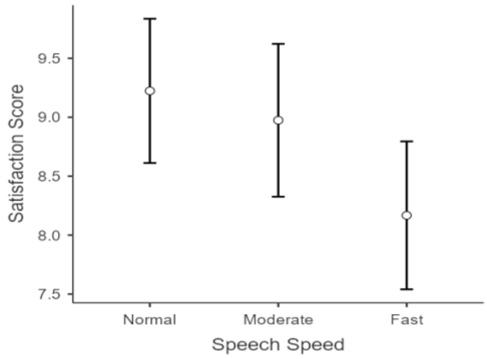
Total satisfaction scores (M = 8.78, SD = 3.08) were calculated by summing the responses given to each question on the scale. The scores could range from 3 to 15. There was a main effect of speech type, $F(1, 230) = 37.57$, $p < .001$, $\eta^2 = .14$. Participants were significantly more satisfied with natural speech (M = 9.92, SD = 2.99) than synthetic speech (M = 7.66, SD = 2.74). A main effect of speech speed tended to occur (*see Figure 3*), $F(2, 230) = 3.04$, $p = .05$, $\eta^2 = .02$. We did not refer this finding as significant because of the low effect size.

Discussion

This research investigated how the audio material's speed rate and speech type affect college students' learning performance. We included covariates in our analyses to eliminate possible confounds of prior familiarity and individual differences in attention-related mistakes. We hypothesized that increased speech speed would increase comprehension and learning efficiency. Although comprehension rate did not differ among different speech groups, efficiency increased as speech got faster. We predicted that natural speech would result in higher comprehension but would not affect efficiency as efficiency was a measure that took time into consideration. We also predicted an interaction between speech type and speech speed on comprehension and efficiency. In other words, we expected that the effect of speech speed would differentially be prevalent. Our results revealed that speech type affected neither comprehension nor efficiency. There was also no interaction on comprehension. However, there was a significant interaction between speech type and speech speed. The effects of speech speed on efficiency were more pronounced in natural speech. The implications of our results were discussed.

Figure 3.

Changes in Satisfaction Scores as a Function of Speech Speed



Note. The mean satisfaction scores for normal, moderate, and fast speech conditions (error bars reflect 95% CI for means).

Firstly, we did not find a main effect of speech rate and speech type on learner performance. Also, there was no interaction between these two variables. Which speed rate or speech type participants listened to did not significantly affect their performance.

Although the effects were not significant, we found the implication of this finding necessary. With the saving of time in compressed speed conditions, participants can perform as well as in normal speed conditions. Goldhaber and Weaver (1968) and Foulke (1968) found a main effect of the rate. However, their speed levels differed from ours in terms of wpm rates. They found that after 250 wpm, the decline in performance began. This previous finding parallels ours because the fast speed condition contained 247 wpm, below the declining level. That may be the reason for the lack of effect in our study. For speech type, Lebeter and Saunders (2010) found the main effect of speech type on comprehension measured by the sentence verification task: participants in the natural speech condition performed better than those in the synthesized speech condition. We interpreted these contrasting findings to be due to the differences in measurement techniques.

We further calculated participants’ efficiency scores by dividing the performance scores by the duration of the audio material. Firstly, we found a main effect of speed rate. As the audio becomes faster, the efficiency increases. Secondly, we did not find a main effect of speech type because the durations of audios did not change across natural and synthetic speech for each speed group. Performance scores were not found to be different from each other. Thirdly, we found an interaction between speed rate and speech type. Although the efficiency scores increase as the speed rate of the audio increases for both groups, the amount of increase was higher for natural speech conditions than synthetic speech conditions. However, the effect size of this interaction was not large. Thus, it should be interpreted with caution. Adelson (1975) applied the same formula for efficiency index and found that efficiency increases until 280 wpm point; because our fast condition contained 247 wpm, these findings were consistent. However, Adelson (1975) criticized the formula for the efficiency index by arguing that this index takes the role of neither the number of ideas nor the difficulty of ideas per time unit into consideration. We had similar concerns and thought it necessary to develop more sophisticated indexes for efficiency.

We measured participants' attention-related mistakes and the extent to which their answers were based only on listening to the audio material to rule out their possible effects. The attention scores were not significantly correlated with participants' performance and performance efficiency. However, the number of correct responses and performance efficiency were significantly correlated with the answers participants rated as giving by listening to the audio material. This could be both because of possible prior knowledge of the participants regarding the audio content and answering by not knowing but rather randomly due to the nature of forced multiple-choice questions.

We gathered students' perceived difficulty data and compared them across speed rate and speech type conditions. Firstly, we found that participants in the synthetic speech condition rated the difficulty higher than those in the natural speech. Secondly, there was a main effect of speed rate on perceived difficulty; participants in the fast condition perceived more difficulty than to those in moderate and normal speed conditions, whereas normal and moderate speed did not significantly differ. This finding is consistent with the previous work on perceived difficulty, which found that after a 1.5-speed rate, the perceived difficulty increases (Zemlin et al., 1968). We interpreted these findings with the familiarity of natural speech and normal speed in real everyday life. Although the learner performance did not differ across speech types and speed rates, perceived difficulty scores did. Even though they perceived the experience of listening as difficult, it did not affect their performance, or their performances in different speed conditions did not affect their difficulty perception. There was no interaction. Since the perceived difficulty items independently were developed for the purpose of this study, their validity and reliability should be further investigated. Additionally, participants might have responded to the questions for difficulty based on either the technical properties of the audio or the content, potentially not measuring the same construct for all participants. Future research may develop better questionnaires to measure perceived difficulty.

Lastly, we looked at the participants' satisfaction ratings for the audio across conditions. There was a main effect; natural speech was more satisfactory than synthetic speech. The main effect tended to occur but with a very small effect size. No interaction was found. For the findings on the difficulty and satisfaction, we thought that if the synthetic speech quality can be improved with technology, it will have many advantages. As the comprehension scores did not differ from natural speech, they can be used widely in the future to enhance its attractiveness.

One limitation of our study was that we only asked to what extent the participants answered the questions with the knowledge they gained by being exposed to the audio. Future research may also ask how much participants' responses were affected by prior knowledge. Future studies investigating the effect of speech compression and speech type could be cautious and develop new methodologies while assessing the performance based on multiple-choice questions.

The main contribution of our study was that we investigated the interaction between speed rate and synthetic speech on learner performance accounting for individual differences. This was particularly important due to the reasons mentioned earlier. Previous research focused on intelligibility and comprehension; however, we measured learning performance because our goal was to optimize the learning experience. To our knowledge, previous research did not include individual differences or any possible covariate while measuring comprehension. Therefore, we added learning perception, which is the students' estimation of how much they learned about the topic from the audio, and attention-related mistakes as covariates to measure learning performance better to understand the effects of speed rate and speech type. Also, previous research was based mainly on English-speaking students; we conducted our study with Turkish students, and our results were consistent with previous research.

We did not investigate how the difficulty of the material content would affect the performance. Would using a more difficult or a more accessible material interact with our main variables? This could be a question for future research. Goldhaber and Weaver (1968) found that the difficulty of the content influences listeners' comprehension, and it follows a curvilinear trend in that medium-difficult content was more comprehensible than easy and difficult ones. Yet, they did not find an interaction between speed rate and difficulty and did not measure the results for efficiency. Thus, to see the interaction between difficulty and speech type and difficulty's effect on efficiency, extended research should be conducted.

To conclude, we found that accelerating the speech and changing the speech type did not affect learning. Nevertheless, increasing the speed of speech lead to more efficient learning when the amount of knowledge is acquired per second. The efficiency of learning did not differ among speech types. Both speech types showed efficiency when the audio was compressed to higher speeds, whereas natural speech was more efficient when the audio was compressed to higher rates. Today it is essential to study speech technologies and their implications since many students study online due to the ongoing COVID-19 pandemic. Therefore, it is necessary and inevitable to analyze the effects of speed rate and the type of speech to test the efficiency of these practices.

References

Adelson, L. (1975). Comprehension by college students of time-compressed lectures. *The Journal of Experimental Education*, *44*(1), 53–60. https://doi.org/10.1080/00220973.1975.11011513

Barabasz, A. F. (1968). A study of recall and retention of accelerated lecture presentation. *Journal of Communication*, *18*(3), 283–287. https://doi.org/10.1111/j.1460-2466.1968.tb00077.x

Carriere, J. S., Cheyne, J. A., & Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and Cognition*, *17*(3), 835-847. https://doi.org/10.1016/j.concog.2007.04.008

Chen, W., & Rojas, S. (2012). *Voice Dream* (Version 4.14.6) [Mobile app]. App Store. https://www.voicedream.com

Cheng, L., Pastore, R., & Ritzhaupt, A. D. (2022). Examining the accelerated playback hypothesis of time-compression in multimedia learning environments: A meta-analysis study. *Journal of Educational Computing Research*, *60*(3), 579-598. https://doi.org/10.1177/07356331211043535

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914-927. https://doi.org/10.1037/0096-1523.23.3.914

Evans, C. (2008). The effectiveness of m-learning in the form of podcast revision lectures in higher education. *Computers & Education*, *50*(2), 491-498. https://doi.org/10.1016/j.compedu.2007.09.016

Fairbanks, G., Guttman, N., & Miron, M. S. (1957). Auditory comprehension of repeated high-speed messages. *Journal of Speech and Hearing Disorders*, *22*(1), 20-22. https://doi.org/10.1044/jshd.2201.20

Foulke, E. (1968). Listening comprehension as a function of word rate. *Journal of Communication*, *18*(3), 198-206. https://doi.org/10.1111/j.1460-2466.1968.tb00070.x

Foulke, E., & Sticht, T. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, *72*(1), 50-62. https://doi.org/10.1037/h0027575

Fulford, C. (1993). Can learning be more efficient? Using compressed speech audio tapes to enhance systematically designed text. *Educational Technology*, *33*(2), 51-59.

Goldhaber, G., & Weaver, C. (1968). Listener comprehension of compressed speech when the difficulty, rate of presentation, and sex of the listener are varied. *Speech Monographs*, *35*(1), 20-25. https://doi.org/10.1080/03637756809375562

Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *The Journal of the Acoustical Society of America*, *121*(3), 1701-1708. https://doi.org/10.1121/1.2436635

Greene, B., Logan, J., & Pisoni, D. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. *Behavior Research Methods, Instruments, & Computers*, *18*(2), 100-107. https://doi.org/10.3758/BF03201008

Higginbotham, D., Drazek, A., Kowarsky, K., Scally, C., & Segal, E. (1994). Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication*, *10*(3), 191-202. https://doi.org/10.1080/07434619412331276900

Janse, E. (2004). Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, *42*(2), 155-173. https://doi.org/10.1016/j.specom.2003.07.001

Junor, L. (1992). Teaching by tape: Some benefits, problems, and solutions. *Distance Education*, *13*(1), 93-107. https://doi.org/10.1080/0158791920130108

Klatt, D. (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America*, *82*(3), 737-793. https://doi.org/10.1121/1.395275

Kurzgesagt – In a Nutshell. (2017, March 30). *Are GMOs Good or bad? Genetic engineering & our food* [Video]. YouTube. https://youtu.be/7TmcXYp8xu4

LaBarbera, P., & MacLachlan, J. (1979). Time-compressed speech in radio advertising. *Journal of Marketing*, *43*(1), 30–36. https://doi.org/10.1177/002224297904300103

Lebeter, J., & Saunders, S. (2010). The effects of time compression on the comprehension of natural and synthetic speech. *Working Papers of the Linguistics Circle of the University of Victoria*, *20*(1), 63-81.

Lovell, K., & Plantegenest, G. (2009). Student utilization of digital versions of classroom lectures. *Medical Science Educator*, *19*(1), 20-25.

Nichols, R. G., & Stevens, L. A. (1957). *Are you listening?* New York: McGraw-Hill

Olejnik, S. F., & Algina, J. (1984). Parametric ANCOVA and the Rank Transform ANCOVA When the Data are Conditionally Non-Normal and Heteroscedastic. *Journal of Educational Statistics*, *9*(2), 129–149. https://doi.org/10.3102/10769986009002129

Rotman, T., Lavie, L., & Banai, K. (2020). Rapid Perceptual Learning: A Potential Source of Individual Differences in Speech Perception Under Adverse Conditions? *Trends in Hearing*, *24*, 2331-2165. https://doi.org/10.1177/2331216520930541

Stevens, C., Lees, N., Vonwiller, J. & Burnham, D. (2005). Online experimental methods to evaluate text-to-speech (TTS) synthesis: Effects of voice gender and signal quality on intelligibility, naturalness and preference. *Computer Speech & Language*, *19*(2), 129-146. https://doi.org/10.1016/j.csl.2004.03.003

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24-31. https://doi.org/10.1177/0098628316677643

The jamovi project (2020). *jamovi* (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org

Voor, J. B., & Miller, J. M. (1965). The effect of practice upon the comprehension of time-compressed speech. *Communications Monographs*, *32*(4), 452-454. https://doi.org/10.1080/03637756509375469

Wester, M., Watts, O., & Henter, G. E. (2016). Evaluating comprehension of natural and synthetic conversational speech. *Speech Prosody*, *8,* 766-770.

Zemlin, W. R., Daniloff, R. G., & Shriner, T. H. (1968). The difficulty of listening to time-compressed speech. *Journal of Speech and Hearing Research*, *11*(4), 875-881. https://doi.org/10.1044/jshr.1104.875

The Role of Statistical Learning in Language Acquisition

Sıla Sevi Çapar
Koç University

The ability to extract statistical regularities in the sensory input, namely statistical learning, is one of the key learning mechanisms underlying language acquisition. By using a variety of input, infants' and adults' sensitivity to statistical regularities has been repeatedly documented across modalities. Although initial findings implied domain-generality of this learning mechanism, recent studies have documented modality-specific constraints on statistical learning, indicating its domain-specificity. The current review begins by presenting the major findings on the role of statistical learning in language acquisition and outlines a framework of the underlying processes. The second part discusses statistical learning across modalities, mainly how it is constrained differently across modalities due to the differential aspect of input to which sensory modalities are attuned. The last part of the review discusses the constraints driven by general cognitive abilities of the learner, which determine the type of computations executed over a kind of input and limit the amount of statistical information extracted from it. The review concludes with directions for future research, highlights of potentially yielding research concerning the role of statistical learning in different levels of language acquisition and the evolution of human language.

Keywords: Statistical learning, language acquisition, language development

The ease and the rapidity with which infants acquire a system as complex as language has always been a source of fascination (Hoff, 2013). The assumption that understanding how infants accomplish this task would provide clues to how the human mind works has led to the scientific study of language development. The behaviorist account, which sought to explain language acquisition through associative learning mechanisms was soon displaced by the nativist view, which proposed Language Acquisition Device as the innate mental capacity of infants to acquire language without sufficient input and reinforcement (Hoff, 2013). The deep similarities underlying the surface differences between the languages of the world are taken to imply a biological endowment of linguistic knowledge in the form of Universal Grammar (Christiansen et al., 2002). While there is a consensus about the existence of constraints determining the structure from which natural languages emerge, there is a disagreement on the nature of these constraints (Christiansen & Ellefson, 2002). Learning-oriented theories of language acquisition have suggested that similarities among languages are due to constraints on human learning mechanisms, rather than a neurologically hardwired innate capacity to process linguistic information (Saffran, 2002; Frost et al., 2015). In particular, the domain-general cognitive approach suggests that the problem of language acquisition is solved by applying of general-purpose reasoning abilities to linguistic input, such as memory capacity, symbolic representation ability, sensitivity to patterns in environmental input, and the ability to extract statistical regularities from a variety of input, namely statistical learning (Kelly & Martin, 1994; Hoff, 2013).

Statistical learning is a domain-general learning mechanism that enable organisms to detect the (statistical) structure underlying complex environmental input, including language (Saffran, 2002; Krogh et al., 2013; Frost et al., 2015). To date, exceptional statistical learning ability of infants and adults have been documented empirically in linguistic and non-linguistic domains (Saffran et al., 1996; Saffran et al., 1999; Fiser & Aslin, 2002;

Roseberry et al, 2011). Although earlier studies suggested that statistical learning is a domain-general learning mechanism that operates across sensory modalities to process different types of inputs, this view was later challenged by the finding that there are modality-specific constraints on statistical learning (Conway & Christiansen, 2005). Neuroimaging studies showing differential activation of brain regions during the extraction of statistical regularities from different types of input and the lack of correlation between measures of statistical learning across different sensory modalities further supported the claim of domain-specificity (Frost et al., 2015; Milne et al, 2018). Besides modality-specific constraints, comparative studies showing sensitivity to statistical regularities in sensory input in other species such as songbirds, macaque monkeys, cotton-top tamarins, and rats (Santolin & Saffran, 2018) have indicated that constraints on general cognitive abilities also affect statistical learning processes. These constraints on general cognitive abilities may have exerted a selection pressure on languages, favoring the linguistic patterns that are most learnable by infants and thus causing to the similarities among the world's languages (Christiansen et al., 2002).

The ability to extract statistical regularities has been shown to contribute linguistic processing at multiple levels, from low-level processes such as discrimination of speech sounds, learning prosodic and phonotactic regularities, abstraction of phonological rules, to higher level processes like word learning and syntax acquisition (Gómez, 2007; Romberg & Saffran, 2010). The current review will first summarize the major findings on the role of statistical learning in language acquisition with a focus on word segmentation and outline a framework of the processes underlying statistical learning. This will be followed by a discussion on the domain-specific versus domain-general nature of this learning mechanism in the light of research investigating the constraints on statistical learning across modalities. Lastly, constraints driven by learners' general cognitive abilities, which limit the amount of statistical information extracted will be discussed in the light of comparative findings.

Statistical Language Learning

One of the biggest challenges infants face in the process of language acquisition is to segment fluent speech into words in the absence of word boundaries marked by a consistent perceptual feature, such as white spaces between words in print (Saffran et al, 1996). Although it was known that infants could do this easily, the learning mechanisms involved were unknown. Saffran et al. (1996) suggested that a type of conditional statistical information, the transitional probability (TP), which expresses the probability of event Y occurring given the information that event X has occurred, can be used to discover word boundaries in sound sequences. To test this argument, researchers first familiarized twenty-four 8-month-old infants with a 2-minute continuous stream of speech from an artificial language consisting of three-syllable nonsense words ("words") without any cues to word boundaries (e.g., prosodic cues such as pauses, pitch, stress, and duration differences) except for different transitional probabilities between syllables within words (high TPs) and syllables at word boundaries (low TPs). In the testing phase, each infant was presented with two "words" and two "part-words" formed by combining the last syllable of one word with the first two syllables of the next word, all encountered during the familiarization phase. Infants showed a novelty preference for "part-words" as indexed by longer listening times, meaning that they can detect word boundaries solely based on the statistical structure of string. In this study, however, the frequency of "words" was higher than that of "part-words" in the auditory material used in the familiarization phase, therefore the design of the study did not allow the researchers to distinguish the specific statistical information (transitional probability or frequency of occurrence) used to segment words from fluent speech. In another study in which the frequency of occurrence of words and part-words was equalized, infants were able to distinguish words from part-words based only on the transitional probabilities between syllables (Aslin et al., 1998).

However, the artificial languages used in the studies presented so far consist of either entirely of disyllabic or trisyllabic words, in contrast to natural languages, which consist of words of varying lengths. Johnson and Tyler (2010) examined infants' ability to segment an artificial language consisting of words of varying lengths and found that neither 5.5 nor 8-month-old infants were able to track the transitional probabilities between syllables in this input language. While this finding seems to cast doubt on the applicability of the ability to compute conditional statistics in segmenting natural language, Thiessen et al. (2005) suggest that natural speech offers other cues that facilitate infants' access to and use of statistical information. To test this hypothesis, they compared infants' ability to discriminate between words and part-words after exposure to infant-directed (ID) and adults-directed (AD) speech. While a natural ID speech contains some auditory cues that facilitate infants' location of word boundaries (e.g., longer pauses at phrase boundaries), the ID speech used in this study was constructed to include only one feature of ID speech: exaggerated pitch contours, which provided no additional cues to word boundaries, other than the statistical structure of the speech stream identical in the ID and AD conditions. Infants were successful in word segmentation only after ID speech, suggesting that the prosodic properties of ID speech facilitate the extraction and use of statistical information to segment the speech stream.

Just as prosodic cues increase sensitivity to conditional probabilities, the ability to extract statistical regularities can increase sensitivity to other linguistic cues, which is proposed to occur through further analysis of the word inventory provided by initial statistical learning processes (Thiessen & Saffran, 2007). Indeed, the finding that 7.5-month-old infants can use stress cues to detect word boundaries, the youngest age at which infants have been shown to be able to segment words from fluent speech, raises the possibility that stress may be the earliest cue used by infants for word segmentation. For this possibility to be true, when conditional statistical cues contradict stress cues in terms of the word boundaries they indicate in a given linguistic input, infants should be shown to prefer stress cues over conditional statistics to locate word boundaries and segment words. In a series of experiments, Thiessen and Saffran (2003) showed that 9-month-olds rely on stress cues as an indicator of word boundaries when stress cues contradict probabilistic cues.

Nevertheless, the hypothesis that stress is the earliest cue that infants use to segment words leaves an important question unanswered: How do infants discover the dominant stress pattern in a language? Thiessen and Saffran (2003) suggested that infants can first identify a sufficient number of words using conditional statistics and then detect the dominant stress pattern of the language, which can serve as a reliable cue to word boundaries. To test this claim, they investigated whether at an earlier age, 6.5 to 7 months of age, infants give conditional statistical cues more weight than stress cues when segmenting words from fluent speech. Although their findings confirm the claim that 7-month-olds make greater use of conditional statistical cues when segmenting words, this does not negate that they are aware of the dominant stress pattern in the segmentation process. Indeed, they found the input with a stress pattern different from the one used in familiarization phase to be unusual, as indicated by longer looking times to the input with unusual stress pattern. According to the researchers, infants may attend to different cues at different time points, such that they can use conditional statistics as the primary cue for word segmentation, then learn the dominant stress pattern through repeated exposure to isolated words, and gradually integrate multiple cues for subsequent language processing. In this way, different types of statistical information involved in language acquisition, which will be discussed in the following paragraphs, feed each other in the process of language development.

To better understand the processes underlying statistical learning, Thiessen et al. (2012) proposed a framework that combines three types of statistical information to which learners are sensitive to: conditional statistics, distributional statistics, and cue-based statistics. Conditional statistics provide information about the predictive relationship between two events and transitional probability is a type of conditional statistic used by infants in most word segmentation studies, including those reviewed so far. Distributional statistics convey the prototypical characteristics of a set of exemplars (i.e., the central tendency of a category). Sensitivity to the distributional structure of the input, the frequency and variability of exemplars, and the context in which exemplars occur, enables learners to discriminate between exemplars of two categories and make judgements about stimuli close to a category boundary (e.g., the categorical distinction between the phonemes /d/ and /t/).

Cue-based statistics reflect the correlation between perceptible features of the input and features that cannot be directly perceived, such as word stress indicating word boundaries. The Extraction and Integration Framework proposed by Thiessen et al (2012) is based on the combination of two processes: one is *extraction*, the identification and long-term storage of statistically consistent clusters, and the other is *integration*, the comparison of exemplars stored in long-term memory to identify commonalities between them and the central tendency of the exemplars. According to this framework, learners first segment exemplars from a continuous input using conditional statistical cues. The exemplars stored in long-term memory then serve as input for the integration process, where the computation of distributional statistics comes into play to make comparisons between exemplars. Accordingly, the different cue weighting observed between 7- and 9-month-old infants in Thiessen and Saffran (2003) is explained as follows: conditional statistical cues enable infants to identify various word forms, which are further analyzed to discover phonological cues to word boundaries. In this framework, regularities discovered during the integration process (i.e., distributional features) inform subsequent extraction by biasing attention to words that match the acoustic properties (e.g., stress pattern) of the exemplars stored, a process in which cue-based statistical learning is at play.

The primary claim on which the Extraction and Integration Framework is built is that learning results in discrete representations (Thiessen et al., 2012). Successful discrimination between words and part-words alone does not evidence the infant's ability to discover "words" in fluent speech. Rather than representing sound sequences as discrete units available for mapping to meanings, infants may simply be discriminating between parts that occur with different TPs. Therefore, it is important to show that the elements extracted through statistical learning are represented in a unitized way. To address the issue, Romberg and Saffran (2010) tested infants on a label-object association task after the familiarization phase, rather than testing their ability to distinguish words from part-words. Once infants were habituated to label-object pairs, where labels are sequences from the stream of speech used in the familiarization, the "Switch" procedure was used to test whether infants learned labels and objects only separately or also the mapping between them. In the "same" trials, infants were presented with the same label-object pairs to which they were habituated. In the "switch" trials, new pairs of the same labels and objects seen in the habituation were presented. Infants who have not learned the mapping between object and label are expected to be equally attracted to the pairs presented in "same" and "switch" trials, whereas infants who have learned the association between label and object are expected to look longer in "switch" trials where the learned association is violated. The researchers showed that only the group of infants who saw words as labels, compared to those who saw part-words as labels, looked longer on Switch trials than on Same trials. This finding suggests that units segmented from the speech stream are represented as candidate words available for mapping to meaning in the label-object association task, and thus statistical cues facilitate word learning.

Similarly, Saffran and Wilson (2003) expected infants to accomplish two linguistic tasks after the familiarization phase: word segmentation and learning grammatical patterns. After identifing words in input sentences by tracking transitional probabilities at the syllable-level, infants were shown to be able to track the ordering relationship between words and learn simple grammatical patterns, as indexed by longer listening times for grammatical versus ungrammatical test stimuli. This study reveals the interplay between statistical learning mechanisms operating at different levels of linguistic structure and demonstrates how the output of one learning process serves as input for the learning process at the next level.

However, the studies documented so far have not addressed the acquisition of semantic properties of word categories, which is an integral part of language learning. To investigate the acquisition of semantic properties of word categories, Lany and Saffran (2010) exposed two groups of infants to two-word phrases sequences from an artificial language, one, the experimental language, contained word categories distinguished by statistical cues, while the other, the control language, contained the same vocabulary yet lacked cues to these word categories. In the referent training phase, both groups were presented with pictures of animals and vehicles paired with labeling phrases heard during auditory familiarization. In the generalization trial, all infants saw pictures of animals and vehicles which were not present in referent training and heard labeling phrases from auditory familiarization. Only infants in the experimental group were exposed to statistical cues indicating category membership of labeling phrases during the familiarization phase were found to learn associations between words and pictures. Furthermore, only these infants were able to identify which of the unfamiliar pictures are likely to be referred to by phrases. Thus, experience with reliable statistical cues enabled infants to discover the semantic properties of words and generalize these semantic properties to novel category members. These findings suggest that experience with statistical cues paves the way for the identification and acquisition of different linguistic information, such as the semantic properties of word categories.

Statistical Learning Across Modalities

While the studies outlined so far demonstrated the role of statistical learning in language learning, they did not address the domain-generality of statistical learning. Saffran et al. (1999) was first to investigate whether statistical learning is language-specific or operates over non-linguistic input as well. In their study, adults and 8-month-old infants were exposed to non-linguistic continuous auditory sequences consisting of tones combined into tone words that were identical in their statistical structure to the syllable-based words used in Saffran et al. (1996). This allowed researchers to make direct comparisons between linguistic and non-linguistic input types in terms the discoverability of their underlying statistical structures by individuals. Both groups performed equally well on the tone segmentation task and on the speech segmentation task used in Saffran et al. (1996), suggesting that the learning mechanism employed to segment speech elements (linguistic stimuli) also serves to compute transitional probabilities between sequences of adjacent tone units (non-linguistic stimuli). While this study revealed that statistical learning is not a language-specific mechanism, it was not enough to show its domain-generality.

To investigate visual statistical learning in infancy, Kirkham et al. (2002) familiarized 2-, 5-, and 8-month-old infants with a series of discrete visual stimuli ordered in a statistically predictable sequence; only the conditional statistical pattern determined the structure of this sequence. After the habituation phase, infants in all age groups showed a preference for novel sequences consisting of the same stimulus components but violating the statistical structure of the familiar sequence. While this study shows that infants are sensitive to statistical regularities in visual sequences, it is not clear whether they relied on the frequency of shape co-occurrences or on the transitional probabilities between shapes, as in the case of Saffran et al (1996). Fiser and Aslin (2002) clarified this question by showing that 9-month-old infants are sensitive to both the frequency of co-occurrence of elements and the conditional probability relations between elements. To test infants’ ability to use statistical learning to parse continuous dynamic events that resemble those encountered in the natural environment, Roseberry et al. (2011) familiarized infants with a videotaped sequence of three hand movements consisting of “units” arranged in a “frequency-balanced” manner as in Aslin et al. (1998). In the test phase, in which infants were presented with units and part-units formed in the same way as the part-words in the auditory domain studies, infants looked at the units longer than the part-units. In this study, in which infants were shown to be able to compute transitional probabilities during the analysis and parsing of continuous dynamic events, infants showed a familiarity preference rather than the novelty preference they showed in the aforementioned studies, possibly due to the stimulus complexity.

In addition to hearing and vision, Conway and Christiansen (2005) presented adult participants with tactile sequences conforming to an artificial grammar and showed that the experimental group exposed to training sequences was significantly better at discriminating between test sequences using adjacent element statistics, suggesting a tactile statistical learning ability. The studies reviewed so far suggest that infants are sensitive to statistical information in linguistic, auditory, and visual inputs, thus statistical learning is a domain-general process. However, without using comparable materials and identical procedures across modalities, it would not be possible to conclude that SL is a domain-general ability. Accordingly, following the experiment with tactile stimuli, the same researchers compared participants' performance quantitatively and qualitatively across modalities, controlling for low-level perceptual abilities and training performances. They found that auditory statistical learning was better than both visual and tactile learning (Conway & Christiansen, 2005). As for the qualitative difference, the aspect of the sensory input to which the modalities are attuned were different. For example, in the tactile domain, participants were most sensitive to the length of sequence and the information at the beginning of a sequence. In the auditory domain, participants were most sensitive to the information in the last part of a sequence. Although these findings suggest the existence of modality constraints on statistical learning, the source and nature of these constraints have yet to be determined. Conway and Christiansen (2005) suggested that modality-specific brain areas may mediate learning as they become more efficient over time in processing certain types of stimuli that contain some form of statistical structure. For example, while the auditory system may specialize in processing statistical relationships between temporal elements, the visual system may excel in processing spatial relationships.

Indeed, the fact that temporal and spatial information play different roles in different modalities has been demonstrated in other perceptual and cognitive tasks (Emberson et al., 2011). While sound is temporally variable and transient in space, vision is spatially constant, therefore it seems adaptive for these sensory systems to be more sensitive to corresponding aspects of the environmental input. According to the "law of proximity", which accounts for the spatial bias in visual processing that leads to the perception of closer visual stimuli as a single unit, sounds presented in greater temporal proximity are not perceived as parts of the same information stream. To investigate the role of perceptual nature of input in SL processes, Emberson et al. (2011) presented participants with statistically equivalent auditory and visual stimuli at fast and slower rates during the familiarization phase and showed that auditory SL improved while visual SL was impaired when stimuli were presented at a faster rate. On the contrary, visual SL was better than auditory SL when stimuli were presented at a slower rate. These results suggest that temporal and spatial information is processed differently across modalities, thus SL is mediated by the perceptual nature of the input. In addition to the mechanism proposed in Conway and Christiansen (2005), the authors of this study propose an embodied architecture to account for the observed modality-specific patterns. According to this account, learning processes are not separate from perceptual processes, but rather learning mechanisms are grounded in perceptual processing.

Next, to explore commonalities and differences between statistical learning operating in different domains, the researchers investigated the relationship between SL performance across different sensory modalities (Frost et al., 2015; Milne et al, 2018). Although there are mixed findings (Milne et al, 2018), in general, the transfer of learning in one modality to other modalities was limited, a statistical regularity captured in one modality did not interfere with the SL process in other modalities, and SL in different modalities was not correlated (Conway & Christiansen, 2005; Frost et al., 2015). These results point to the existence of independent modality constraints and thus to the domain-specific nature of statistical learning.

In line with the proposal of Conway and Christiansen (2005), Frost et al. (2015) proposed a set of computational principles that are instantiated by discrete neural networks according to unique constraint sets determined by specific input properties. Neuroimaging studies comparing activation patterns for structured and unimodal stimuli have shown stronger activation in higher-order visual networks for structured visual stimuli and analogous auditory networks for structured auditory stimuli, suggesting that SL occurs in brain regions dedicated to processing a specific stimulus type (Frost et al., 2015). Besides local computations, some brain regions, such as the medial temporal lobe, have been shown to be involved in SL processes regardless of stimulus modality (Frost et al., 2015; Milne et al., 2018). This finding may clarify the findings outlined in Milne et al. (2018), which point to the existence of transfer, facilitation, and interference of learning across modalities.

Statistical Learning Across Species

In addition to constraints arising from input characteristics, there are constraints related to the cognitive and perceptual state of the learner. To illustrate such constraints, SL abilities in species with different general perceptual and cognitive abilities can be examined. For example, Zebra finches were found to be able to track transitional probabilities when learning sequences of song syllables (Santolin & Saffran, 2018). However, unlike Budgerigars, which used the underlying statistical structure of the training string to discriminate between two novel sets of sound strings, Zebra finches made this discrimination based on the sequential positions of specific items in the training strings (Spiering & Cate, 2016). Zebra finches failed to generalize learned sequences to strings of novel song syllables, while Budgerigars were able to do so. This finding is consistent with the vocal learning characteristics of each species. The vocal repertoire of zebra finches consists of repetitive, linear patterns of syllable sequences, whereas budgerigars' songs have more vocal variation and show a syntactic-like organization (Santolin & Saffran, 2018). This suggests that species-specific learning abilities may have shaped song structures by limiting vocal learning abilities.

Memory ability is also crucial both for temporarily holding elements of the incoming input for future processing and for retaining the representations formed during the extraction process (Santolin & Saffran, 2018). Wilson et al. (2013) showed that when presented with sequences in which an artificial grammar familiar to common marmosets and Rhesus macaques was violated at different positions in the sequence, common marmosets were only able to detect violations that occurred at the beginning of sequences, in contrast to Rhesus monkeys, which demonstrated deeper levels of learning. This finding suggests that common marmosets, which are evolutionarily more distant to humans, have limited memory abilities compared to Rhesus monkeys. Budgerigars, which were shown to be skilled at extracting regularities underlying auditory strings, were also found to have an exceptional memory for complex acoustic stimuli. Thus, it can be argued that memory capacities that constrain statistical learning ability by limiting the information available to the learner (embedded in the input) may affect the structural organization of vocalizations in vocal learners (e.g., songs and human language). These arguments for the role of general cognitive abilities such as learning and memory are in line with the account of both cross-linguistic similarities in world languages and the ease and rapidity with which infants learn human language (Christiansen et al., 2002). That is, the constraints on general human learning abilities may have shaped the structure of natural languages through natural selection of the linguistic structures most learnable by human infants.

Conclusions

Statistical learning has been shown to play a crucial role in language acquisition. As the Extraction and Integration Framework suggests, language experience provides infants with a repository from which they extract various language-specific cues, and with the help of these cues, infants segment words from the speech stream through statistical learning. Moreover, the interaction between the outputs of statistical learning processes at different levels facilitates the acquisition of semantic properties of word categories and syntax.

Although the findings that statistical learning operates in visual, auditory, and tactile domains were initially interpreted as indicating domain-generality, it has been shown to be differently constrained across modalities according to the aspect of input to which sensory modalities are attuned. Findings that there was limited transfer, interference, and correlation between SL in different modalities and that different brain areas performed computations in different modalities support the domain-specificity claim. In addition to input characteristics, constraints on general cognitive abilities were found to influence statistical learning by limiting the amount of statistical information extracted from the input and or by limiting the type of computations performed to discover statistical patterns in the input. Furthermore, the correlation between constraints on statistical learning and the complexity of vocalization structure in songbirds implies that cross-linguistic similarities may have resulted from constraints on general learning mechanisms. To gain insight into the human-specific adaptations that contributed to the emergence of human language, it is crucial to investigate differences in the type and scope of statistical computations between species. Future studies investigating the neural basis of statistical learning across modalities in non-human species will be particularly informative in this respect. However, since there are inevitable methodological differences between studies on various species, researchers should be cautious when making comparisons across species and drawing evolutionary inferences.

To better understand the role of statistical learning in language acquisition, future research should investigate the interaction between SL and other learning mechanisms hypothesized to be involved in language acquisition. It would be highly informative to understand which learning mechanisms are employed by infants to tackle what kind of regularities. Furthermore, investigating the relative weighting of different types of statistical information at different time points in development and for different inputs will provide important insights into language acquisition.

In addition to experimental studies, the role of statistical learning in language development can be investigated through longitudinal designs that aim to uncover the relationship between children's SL skills in the laboratory and native language outcomes such as language comprehension, lexical knowledge, and reading proficiency. Furthermore, examining quantitative and qualitative individual differences in the statistical learning performance in both typically and atypically developing children will reveal the relationship between SL ability and other cognitive measures. Investigating the relative weighting of different linguistic and statistical cues by children with different language impairments during language acquisition will provide insights into the nature of developmental language disorders.

References

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321-324. https://doi.org/10.1111/1467-9280.00063

Christiansen, M. H., Dale, R. A., Ellefson, M. R., & Conway, C. M. (2002). The role of sequential learning in language evolution: Computational and experimental studies. *Simulating the Evolution of Language* (pp. 165-187). Springer, London.

Christiansen, M.H. & Ellefson, M.R. (2002) Linguistic adaptation without linguistic constraints: The role of sequential learning in language evolution. *Transitions to Language* (pp. 335–358). Oxford University Press.

Conway, C., & Christiansen, M. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *31*(1), 24-39. https://doi.org/10.1037/0278-7393.31.1.24

Emberson, L., Conway, C., & Christiansen, M. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal Of Experimental Psychology*, *64*(5), 1021-1040. https://doi.org/10.1080/17470218.2010.538972

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822-15826.

Frost, R., Armstrong, B., Siegelman, N., & Christiansen, M. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends In Cognitive Sciences*, *19*(3), 117-125. https://doi.org/10.1016/j.tics.2014.12.010

Gómez, Rebecca L. (2007). Statistical learning in infant language development. In Gareth Gaskell (Ed.), *Oxford Handbook of Psycholinguistics* (pp. 601–616). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0036

Hoff, E. (2013). *Language development*. Belmont, CA: Wadsworth, Cengage Learning

Johnson, E., & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339-345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Kelly, M., & Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua*, *92*, 105-140. https://doi.org/10.1016/0024-3841(94)90339-5

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42. https://doi.org/10.1016/s0010-0277(02)00004-5

Krogh, L., Vlach, H., & Johnson, S. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers In Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00598

Lany, J., & Saffran, J. (2010). From statistics to meaning. *Psychological Science*, *21*(2), 284-291. https://doi.org/10.1177/0956797609358570

Milne, A., Wilson, B., & Christiansen, M. (2018). Structured sequence learning across sensory modalities in humans and nonhuman primates. *Current Opinion In Behavioral Sciences*, *21*, 39-48. https://doi.org/10.1016/j.cobeha.2017.11.016

Romberg, A., & Saffran, J. (2010). Statistical learning and language acquisition. *Wires Cognitive Science*, *1*(6), 906-914. https://doi.org/10.1002/wcs.78

Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R., & Shipley, T. (2011). Babies catch a break. *Psychological Science*, *22*(11), 1422-1424. https://doi.org/10.1177/0956797611422074

Saffran, J. (2002). Constraints on statistical language learning. *Journal Of Memory And Language*, *47*(1), 172-196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52. https://doi.org/10.1016/s0010-0277(98)00075-4

Saffran, J., & Wilson, D. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273-284. https://doi.org/10.1207/s15327078in0402_07

Santolin, C., & Saffran, J. (2018). Constraints on Statistical Learning Across Species. *Trends In Cognitive Sciences*, *22*(1), 52-63. https://doi.org/10.1016/j.tics.2017.10.003

Spierings, M., & ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proceedings Of The National Academy Of Sciences*, *113*(27), E3977-E3984. https://doi.org/10.1073/pnas.1600483113

Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71. https://doi.org/10.1207/s15327078in0701_5

Thiessen, E., Kronstein, A., & Hufnagle, D. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792-814. https://doi.org/10.1037/a0030801

Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706-716. https://doi.org/10.1037/0012-1649.39.4.706

Thiessen, E., & Saffran, J. (2007). Learning to learn: infants' acquisition of stress-based strategies for word segmentation. *Language Learning And Development*, *3*(1), 73-100. https://doi.org/10.1080/15475440709337001

Wilson, B., Slater, H., Kikuchi, Y., Milne, A., Marslen-Wilson, W., Smith, K., & Petkov, C. (2013). Auditory artificial grammar learning in macaque and marmoset monkeys. *Journal Of Neuroscience*, *33*(48), 18825-18835. https://doi.org/10.1523/jneurosci.2414-13.2013

Emotion Recognition in Humans: A Literature Review Investigating the Role of Embodiment in the Process of Emotion Recognition

İlayda Arslan
Koç University

Emotion recognition lies at the core of social interaction, so it has a crucial role for social functioning. Theories of embodiment introduce a practical approach to explain how we make use of imitation of others' emotional expressions to understand how they feel. This paper presents a literature review to explain the role of embodiment in the process of emotion recognition in humans. The aim is to examine both the effects of visual cues, such as facial expression and whole-body cues, and the underlying neural mechanisms, as well as the individual and contextual factors.

Keywords: embodiment, emotion, facial expression, body

Emotion recognition, the ability to infer an emotional state of another individual, is a significant domain of social cognition, which is the underlying mechanism of social functioning (Barbato et al., 2015). Since humans are social beings, emotion recognition, mainly acoustic and visual features like vocalization and facial expressions (Mier et al., 2010), is a fundamental capability. Making sense of emotional expressions and knowing how to react to them enriches interpersonal communication (Varghese et al., 2015). Following the early findings that mere perception of another's behavior increases the likelihood of engaging in that behavior, studies show that observers also unintentionally imitate emotional expressions and gestures and then report associated emotional states (Chartrand & Bargh, 1999). Embodied cognition approach provides a distinct account to explain the underlying mechanism. The embodied cognition perspective implies that the cognitive processes are grounded in the body's interactions with the world (Wilson, 2002). Specifically, some social activities involving the perception of own bodily behaviors, such as language understanding or mindreading, can be qualified as embodied. Empathy is also one of them. For example, seeing another person's disgust expression elicits a disgust feeling in oneself (Goldman & de Vignemont, 2009).

In terms of emotional understanding, embodied cognition perspective implies that imitating another's emotional expressions turns into a bodily reenactment of another's emotional state. In other words, embodiment theories claim that a perceptual, somatovisceral, and motoric re-experience of another's emotion in oneself lies at the core of perceiving an emotion (Niedenthal, 2007). On the other hand, since re-experiencing emotions is claimed to be situated, embodied cognition perspective considers individual and contextual differences to account for emotional experience and understanding (Niendenthal & Maringer, 2009). This paper will investigate emotion recognition in people from the theoretical perspective of embodiment; present how embodiment theories can account for emotional inference based on visual cues (facial expressions and whole-body cues), the neural mechanisms underlying these processes, and the role of moderating factors (individual or contextual factors).

Facial Expressions

The term "emotional expressions" is thought to be nearly synonymous with facial expressions of emotion. Therefore, most studies on emotion recognition have focused on faces (Heberlein & Atkinson, 2009). Theories of embodied cognition mainly demonstrate three steps to explain how facial feedback plays a role in emotion perception (Neal & Chartrand, 2011). As a first step, the perceiver subtly and unconsciously mimics the facial expressions of the target. In the second step, subtle face muscle contractions of the perceiver generate a muscular signal as feedback to the brain. In the third step, the perceiver uses this feedback to reproduce perceived expression and thus makes an emotional inference. In a neuroimaging study investigating the recognition of emotional facial expressions, Wicker et al. (2003) gave participants odors to inhale which generated feelings of disgust. Then, the same participants watched the videos of other individuals while expressing disgust. Their results showed that the same areas of the anterior insula and the anterior cingulate cortex were activated both when individuals felt disgust themselves also when they perceived it from the videos. These findings highlight the main idea of the embodied emotion account, which is built on the relationship between experiencing and recognizing emotion.

To assess the causal role of embodiment in emotion recognition, Niedenthal et al. (2009) manipulated embodiment and instructed participants to make an emotion-focused judgment on emotional words. As a manipulation of embodiment, half of the participants were allowed to move their faces freely, and the other half of the participants were prohibited from using some facial responses. More specifically, they made half of the participants hold a pen between their lips and teeth, which prevented certain facial responses. Inhibitory effects of the pen manipulation captured particular emotions, which were disgust and joy. Their results revealed a main effect of word type on accuracy and a main effect of pen manipulation, suggesting that the prohibiting effects were limited to emotions that engage relevant muscles. According to these findings, their proposed claim that the production of facial expressions contributes casually to the processing of emotion concepts was supported.

Building on the existing studies, a study by Neal and Chartrand (2011) pointed out the bidirectional effects of facial feedback, such that emotion recognition is improved in accuracy when facial feedback is increased, and it is dampened when facial feedback is decreased. Similar to Niedenthal et al. (2009), they manipulated the embodiment process in their experiment; to dampen the facial feedback signals of the participants, they used Botox, and to amplify the facial muscle signals, they used a restricting gel. The participants were then required to make judgments about the emotional content of the faces. According to their results, the facial feedback effects were present for both positive and negative emotions. However, the researchers noted that some previous findings showed differential effects for positive or negative emotions or some specific emotions. They interpret these findings in a holistic way, saying that magnitude of the facial feedback effects may vary for specific emotions and are based on the valence of emotions. Therefore, as a result, they found a moderating role for facial feedback in the process of emotion recognition, rather than a mediating role.

According to the evidence from these empirical studies, thinking of facial feedback as a general mediator in the process of emotion recognition can be misleading. The embodiment account to emotion perception of facial expressions requires a more comprehensive approach, mainly assigning facial feedback a moderating role and suggesting possible moderating factors which may integrate with facial feedback in the emotion-perception process. Present literature suggests many possibilities for moderating factors, such as the perceivers' processing goals at the time or various features of interpersonal context (Neal & Chartrand, 2011).

Whole-Body Cues

Although many studies on emotion recognition have focused on faces, other non-verbal cues such as body posture, gestures, and locomotor patterns convey emotional information (Heberlein & Atkinson, 2009). Studies showed that viewing facial expressions and affective body gestures together improves emotion classification accuracy compared to viewing only one of these modalities (Gunes & Piccardi, 2007; Castellano et al., 2008). A long but sparser history of research, which may be traced back to Darwin (1872) has examined how emotional information is conveyed by whole-body cues. Another early line of experiments was conducted by William James (1932), showing that participants who were looking at the photos of posed whole bodies seemed to be imitating those postures. However, he was skeptical about the possibility of getting high inter-subject agreement about the emotion portrayed by any given posture.

More recent studies with both static and dynamic whole-body stimuli have found a high degree of agreement, by using a combination of stimuli that were specifically constructed to convey emotional expressions and more constrained methods rather than free responses of participants (Heberlein & Atkinson, 2009). Many of these studies have focused on identifying visual cues, bodily expressed emotions, and isolated body parts, by using techniques such as the point-light display. For example, some movements (from point-light emotional movement movies) were recognized as fearful by participants, although the figures were unlikely to even be recognizable as human forms without kinematic (regarding the motion of the body) information (Heberlein et al., 2004). Thus, the results show that not just stationary body postures, but motion cues, especially kinematics, convey important information about others' emotional states.

Neural and Cognitive Underpinnings

Classically, the concept of embodiment refers to the body as the vessel for introspection, bringing together insights about the state of the self and the environment. There were some objections to this classical explanation, arguing that bodily feedback is too undifferentiated and sluggish to constitute emotional experience (Cannon, 1927). Modern embodiment theories clarify the subject, claiming that peripheral changes in the body do not constitute an emotion on their own. Rather, peripheral input works with the brain's modality-specific systems to recreate bodily states of emotion very rapidly (Barsalou, 2010; Barsalou et al., 2003).

To explain the underlying neural mechanisms of theories of embodied cognition, different accounts have been proposed. "Mirror neuron system", which supports the "simulation model", is one of the most prominent accounts (Rizzolatti et al., 2001). Mirror neurons map the correspondences between the observed and performed actions. Some of the original work on mirror neurons with monkeys emphasize a distinctive role of neurons located in the inferior parietal and inferior frontal cortex, which discharge both when a monkey performs an action and when it observes another individual's action (Rizzolatti et al., 2001). Extending the implications to humans, some scientists argue that humans have a distinct mirror neuron area (Niedenthal, 2007). Applying this model, we can explain the mechanism behind the application of embodiment process to visual cues in humans. Visual cues, which are important for discriminating facial emotional expressions, are extracted at early stages of visual processing, largely in cortices specialized for face and body processing; and are fed forward to somatosensory cortices. These regions are implicated in simulating certain aspects of the body states associated with the viewed emotional state, including proprioceptive and somatovisceral sensations (Heberlein & Atkinson, 2009).

In line with these explanations, a multi-stage model for processing the emotional expressions of others by Dricu & Frühholz (2016) presents four main stages: Extracting sensory information from visual cues (middle fusiform gyrus & inferior occipital gyrus), integration of sensory information and inference of basic intentionality from the perceived expression (posterior superior temporal sulcus), maintaining choice options in working memory, accessing semantic representations of emotional constructs and inhibiting self-perspective (inferior frontal cortex), and lastly, inferring the mental state of the perceived individual (dorsomedial frontal cortex). According to the findings, the first two stages are sufficient for emotion perception and understanding. Latter stages are necessary to evaluate the perceived emotion further (Dricu & Frühholz, 2016).

Moderating Factors

During emotion perception processes, there is an integration of other moderating factors which are concurrently accompanying the visual body cues, especially facial feedback. These moderating factors may be brought by the perceiver themself, as well as the context. One factor may be the current processing goals of the perceiver. Niedenthal et al. (2009) demonstrated that the manipulation of situational factors can alter the nature of the process of representing a concept. An embodied simulation occurs only under the expected conditions, notably, when generation of the embodied information would provide information that is useful for the task at hand. Another aspect that may be considered as an example of contextual moderating factors of simulation is eye contact. The study of Rychlowska et al. (2012) depicts that eye contact is associated with greater imitation of smiles compared to averted gaze.

Additionally, Jospe et al. (2018) made an important contribution to the literature on embodiment by emphasizing individual differences. They noted that while the most widely held theory states that embodiment is an automatic process, its utilization is subject to individual differences, such that empathetic individuals can be capable of more efficient utilization.

Another factor affecting the process of emotion recognition is culture. For example, the emphasis is on the eyes when perceiving smiles for Asian cultures, like Japan, whereas, for Western cultures, the emphasis is on the mouth (Niedenthal & Maringer, 2009). Furthermore, another study (Mondillon et al., 2007) on understanding facial expressions focuses on imitating in-group versus out-group members' expressions. The researchers showed that Chinese participants living in France both imitated the angry facial expressions of Chinese and French faces, whereas French participants living in France only imitated French faces.

Conclusions and Future Research

For human beings, inter-personal interactions and communication are fundamental parts of life. Thus, emotion recognition, since it is a crucial step for understanding others, is one of the key processes. Embodiment theories stand out as providing valuable and dynamic explanations for emotional understanding. The embodiment account allows us to analyze the brain's rapid recreation process of emotional states in order to interpret the emotional state of the individual that we perceive. However, it also has limitations. First, the process of emotional understanding via embodiment is not restricted only to facial cues. Whole-body cues are also crucial in understanding emotions, which should be researched more, especially in the context of embodiment.

Embodied cognition theories highlight the role of individual and contextual factors, yet they have not been investigated thoroughly. For example, most studies use only face stimuli, keeping it apart from the natural settings that we understand emotions. In naturalistic settings, we do not focus on the facial expressions of speakers while listening to the emotional narrative; we also look at, for example, the hand gestures of the speakers. The reenactment of or the motor feedback from their gestures might also enhance the understanding of speakers' emotions, which should be addressed in future research.

Furthermore, though we know that embodiment effect depends on the valence of the emotion (Niedenthal et al., 2009), the underlying reason why it is not strong for some emotions needs further attention. Additionally, the possibility that embodiment effect can be present only for in-group members (Mondillon et al., 2007) might make the embodiment account less important in emotion understanding from out-group members. How and why we understand the emotions of out-group members differently should be researched further considering the interrelations among stimuli characteristics (e.g., facial and whole body cues) and individual differences (e.g., prior experiences with other cultures, empathy levels, and participants' use of facial and whole body cues in emotion communication).

References

Barbato, M., Liu, L., Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A., McGlashan, T. H., ... & Addington, J. (2015). Theory of mind, emotion recognition and social perception in individuals at clinical high risk for psychosis: Findings from the NAPLS-2 cohort. *Schizophrenia Research: Cognition*, *2*(3), 133-139.

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716-724.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Cciences*, *7*(2), 84-91.

Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion recognition through multiple modalities: face, body gesture, speech. In *Affect and emotion in human-computer interaction* (pp. 92-103). Springer, Berlin, Heidelberg.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893.

de Bruin, L., Strijbos, D., & Slors, M. (2014). Situating emotions: From embodied cognition to mindreading. *Topoi*, *33*(1), 173-184.

Dricu, M., & Frühholz, S. (2016). Perceiving emotional expressions in others: activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience & Biobehavioral Reviews*, *71*, 810-828.

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied?. *Trends in Cognitive Sciences*, *13*(4), 154-159.

Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, *30*(4), 1334-1345.

Halberstadt, J., Winkielman, P., Niedenthal, P. M., & Dalle, N. (2009). Emotional conception: How embodied emotion concepts guide perception and facial action. *Psychological Science*, *20*(10), 1254-1261.

Heberlein, A. S., Adolphs, R., Tranel, D., & Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience*, *16*(7), 1143-1158.

Heberlein, A. S., & Atkinson, A. P. (2009). Neuroscientific evidence for simulation and shared substrates in emotion recognition: beyond faces. *Emotion Review*, *1*(2), 162-177.

Jospe, K., Flöel, A., & Lavidor, M. (2018). The interaction between embodiment and empathy in facial expression recognition. *Social Cognitive and Affective Neuroscience*, *13*(2), 203-215.

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, *47*(6), 1028-1039.

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.

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673-678.

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002-1005.

Niedenthal, P. M., Augustinova, M., & Rychlowska, M. (2010). Body and mind: Zajonc's (re) introduction of the motor system to emotion and cognition.

Niedenthal, P. M., & Maringer, M. (2009). Embodied emotion considered. *Emotion Review*, *1*(2), 122-128.

Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, *96*(6), 1120.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661-670.

Rychlowska, M., Zinner, L., Musca, S. C., & Niedenthal, P. M. (2012, October). From the eye to the heart: Eye contact triggers emotion simulation. In *Proceedings of the 4th workshop* *on eye gaze in intelligent human machine interaction* (pp. 1-7).

Schubert, T. W., & Semin, G. R. (2009). Embodiment as a unifying perspective for psychology. *European Journal of Social Psychology*, *39*(7), 1135-1141.

Varghese, A. A., Cherian, J. P., & Kizhakkethottam, J. J. (2015, February). Overview on emotion recognition system. In *2015 International Conference on Soft-Computing and Networks Security (ICSNS)* (pp. 1-5). IEEE.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655-664.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625-636.

Music as an Emotion Regulation Tool Among Adolescents: Underlying Mechanisms and Clinical Implications for Depression

İlayda Akınkoç
Istanbul Bilgi University

The development of emotion regulation skills during adolescence is crucial for a healthy emotional life. Researchers investigated the effects of musical activities on adolescents' developing emotion regulation skills with the increased use of music through listening, playing, or singing. There are multiple theories regarding how music is used to regulate emotions, emphasizing that music can be used by adolescents to manage their emotions. In addition, different emotion regulation strategies used during musical activities influence how one experiences emotions. Subjective, physiological, contextual, and behavioral factors play a big role in how underlying mechanisms of emotion regulation operate. Impractical and ineffective use of emotion regulation strategies are also observed in several mental disorders (e.g., depression). Moreover, music use as an emotion regulation tool can have both positive and negative impacts on adolescents' depression symptoms. What determines this difference is (1) the reason why adolescents choose musical activities to regulate their emotions, and (2) which emotion regulation strategies are used during these musical activities. Several intervention programs have been developed to make adolescents more aware of the effects of music use on their emotions. This article further discusses what future studies might consider improving the understanding of adolescents' experiences.

Keywords: emotion regulation, emotion dysregulation, adolescence, music use, depression

Adolescence is a period of internal and external transitions in biological, cognitive, emotional, and social domains (Baltazar, 2019). These potentially dramatic changes require adolescents' emotion regulation skills to adjust and make sense of the world. However, similar to the changes in other domains, such as increased emotional capacity and changes in attention adolescents' regulatory skills also change and develop during this period when most needed (Yurgelun-Todd, 2007; Dingle et al., 2016). With technological advancements, this developmental period is also accompanied by increased use of music, changes in musical preferences, and varying reasons for listening to music (Gurgen, 2016; Londsale & North, 2011; Miranda, 2013; Schwartz & Fouts, 2003). Researchers increasingly wonder whether music can be a tool to enhance emotion regulation skills and develop intervention programs to help adolescents experience this transition period more smoothly (Dingle et al., 2016; Uhlig et al., 2018).

Some researchers suggest that although music is popular among adolescents, little attention is given to the scientific study of music and how it relates to their daily lives and specifically, emotion regulation skills (Moore, 2013; Rentfrow, 2012). Studying music use in the context of emotion regulation is important given the prevalence of emotion dysregulation is in mental disorders. The evidence supporting the idea that mental disorders' onset is generally early adolescence further suggests that fostering adaptive emotion regulation skills may have long-lasting and beneficial effects on adolescents' well-being (Hou et al., 2017; Moore, 2013; Sakka & Juslin, 2018). Therefore, understanding how music operates in emotion regulation can help develop therapeutic interventions, which may aid adolescents in improving their regulation skills (Moore, 2013; Papinczak, 2015; Stewart et al., 2019). In addition, listening to certain types of music can act as a maladaptive emotion regulation attempt, where music may adversely affect individuals' emotional experiences (Sakka & Juslin, 2018; Stewart et al., 2019). Understanding these specific music types can also help adolescents understand the potential antecedents (i.e., music type) of their increased negative affect and make their music choices accordingly (Moore, 2013).

Hence, findings indicate a relationship among music, emotion regulation, and mental health. This paper will review (1) the literature on how adolescents use music to regulate their emotions by focusing mainly on the underlying mechanisms and strategies, (2) the relation between music use and depression, which is characterized by emotional dysregulation, and (3) key points to consider in future research.

Underlying Mechanisms of Emotion Regulation Through Music

Londsdale and North (2011) explain that one of the main reasons why people listen to music is to regulate their emotions, both positive and negative. This common function of music uses further sparked researchers' interests to identify the underlying mechanisms of how music regulates emotions. However, there is not a single mechanism explaining adolescents' all emotion regulation experiences (Juslin & Västfjäll, 2008). Therefore, multiple theoretical models will be examined in relation to each other.

The choice of musical activity (e.g., listening to music, writing songs, singing, and dancing) starts with adolescents' need to control their emotions and to manage their affective states (Saarikallio and Erkkilä, 2007). On the other hand, some researchers suggest that the choice of musical activity can be based on three goals: intensification of positive emotions, improvement in negative emotions, and intensification of negative emotions, which also influence the choice of emotion regulation strategies and the underlying mechanisms (e.g., Papinczak et al., 2015; Baltazar, 2019).

Emotion Regulation Strategies

Saarikallio and Erkkilä (2007) suggest that musical activities help regulate emotions through several strategies: *entertainment*, *revival*, *strong sensation*, *diversion*, *discharge*, *mental work*, and *solace*. These strategies mediate the interaction between use of musical activity and the changes in emotions. That is, they can be considered emotion regulation strategies within the context of music.

According to Schäfer et al. (2012), one function of music is entertainment. Thus, it is reasonable to consider entertainment as an emotion regulation strategy. More specifically, the strategy of *entertainment* is most likely to be used by adolescents when they feel alone or bored, which may result in feelings of energy and contentedness, and lead to enhancement or maintenance of positive mood by creating a positive atmosphere (Saarikallio & Erkkilä, 2007; Silverman, 2021; Thomson et al., 2014).

The strategy of *revival* includes recreation, personal renewal, and relaxation (Saarikallio, 2012; Saarikallio & Erkkilä, 2007; Silverman, 2021). Similar to the *entertainment* strategy, it also results in being energetic and cheerful. However, it is used especially when adolescents need to relax due to stress or tiredness (Saarikallio & Erkkilä, 2007; Silverman, 2021). This strategy includes singing and playing music to deal with the difficulties emerge in one's life due to, for example, schoolwork (Saarikallio & Erkkilä, 2007). There also seems to be a sex difference in the use of this strategy, with men reporting use of music for revival more than women (Saarikallio et al., 2012).

The strategy of *strong sensation* influences adolescents' emotions by creating a mechanism through which adolescents feel immersed in music. In other words, it entails "putting ones' soul into music" (Saarikallio, 2012). For example, adolescents are less likely to hear people calling them when they are strongly concentrating on the music because they are merged with the music (Saarikallio & Erkkilä, 2007). This strategy is more likely to be used when adolescents want to feel or maintain intense emotions regardless of emotional valence (Saarikallio & Erkkilä, 2007; Silverman, 2021). However, some critiques of this strategy suggest that it reflects more of an outcome of an emotion regulation process rather than an emotion regulation strategy (e.g., Sakka, & Juslin, 2018).

Furthermore, the strategy of *diversion* is an early modification of attention towards more positive affective states (Saarikallio, 2012). Alongside with *mental work*, *revival*, and *solace*; *diversion* is also a strategy used before an affective response is given to a situation (Saarikallio, 2011). Adolescents are more likely to benefit from this strategy when they experience negative emotions (e.g., sadness, anger, and anxiety) because singing and listening to music help adolescents focus on the music and lyrics, hence divert their attention away from negative emotions (Saarikallio & Erkkilä, 2007).

The strategy of *discharge* gives an opportunity to express current negative emotions through using music which expresses those emotions (Saarikallio, 2012). For example, Saarikallio and Erkkilä (2007) posit how anger can be healthily expressed through listening to aggressive music (e.g., metal music genres). Additionally, with this strategy, music provides a space for adolescents to express and discharge their emotions in adaptive ways (Saarikallio & Erkkilä, 2007). Unlike the strategies of *mental work*, *revival*, *solace*, and *diversion*, this strategy is response-focused and aims to change the behavioral manifestation of an already aroused emotional state (Saarikallio, 2011; Saarikallio, 2012).

In contrast to emotion-focused strategies, the strategy of *mental work* involves cognition and modification or reappraisal of the thinking patterns to regulate emotions (Saarikallio, 2012). Musical lyrics and writing songs may help adolescents regulate their emotions successfully by providing them with new perspectives to reflect on reality and their experiences (Saarikallio & Erkkilä, 2007).

Last, with the strategy of *solace*, music provides comfort and consolidation when negative emotions are experienced. Similarly, through the lyrics of the songs, feelings of being understood may be observed with the implementation of the *solace* strategy (Saarikallio & Erkkilä, 2007; Schwartz & Fouts, 2003). However, van Goethem and Sloboda (2011) argue that the strategies of *strong sensation* and *solace* should not be considered strategies but rather mechanisms through which musical activity influences emotions, which will be explained in the next section.

From another point of view, Baltazar (2019) mentions six emotion regulation strategies used by adolescents when engaging in musical activities, some of which overlaps with Saarikallio and Erkkilä's (2007) model: *cognitive work*, *entertainment*, *affective work*, *distraction*, *focus on situation*, *revival*. She further categorizes them into two main categories: *change-oriented and pleasure-oriented strategies*. Baltazar (2019) suggests that change-oriented strategies are mainly the attempts to regulate and change emotions by actively dealing with the emotional experiences. These change-oriented strategies include cognitive tools (e.g., reappraisal of a situation and consideration of multiple perspectives), affective tools (e.g., intensification of emotions and not hiding or suppressing emotions), and focus on situation (e.g., directing the attention towards the emotional experience itself to reflect on it). Pleasure-oriented strategies are the attempts to deal with personal experiences and regulate emotions by inducing further physical and emotional experiences. These include *entertainment*, *distraction*, and *revival* (e.g., using music to induce relaxation and energy) (Baltazar, 2019).

Underlying Mechanisms

Saarikallio and Erkkilä (2007) suggest that strategies mentioned above lead to emotional changes through regulation of specific elements of subjective, physiological, cognitive, and behavioral experiences of adolescents (Juslin, & Västfjäll, 2008; Saarikallio, & Erkkilä, 2007; Scherer, 2004). Specifically, music and these emotion regulation strategies affect subjective experiences through (1) changes in emotional valence by emphasizing the experiences of positive emotions, (2) changes in intensity of the emotions, be it in a negative or positive direction, and (3) changes in clarity of adolescents' thoughts and reflections (Saarikallio & Erkkilä, 2007). Additionally, emotion regulation strategies during music use affect physiological experiences through alteration of the experienced energy, excitement, and movement (Saarikallio & Erkkilä, 2007; Scherer, 2004). Last, adolescents may experience emotional changes and regulations because music and these strategies also influence how emotions are expressed and behaviorally manifested (Saarikallio & Erkkilä, 2007).

On the other hand, Baltazar (2019) mentions two mechanisms through which these emotion regulation strategies work: individual-dependent and feature-dependent mechanisms. Individual-dependent mechanisms are based on adolescents' personal experiences, such as identification with the singer/songwriter and the lyrics of the songs, recall of music-related memories and experiencing a sense of familiarity with the music and the emotions it induces (Baltazar, 2019). In contrast, the feature-dependent mechanisms are based on the characteristics of the musical activity, such as genre, rhythm, valence, and other sound-related properties. Last, adolescents' past experiences and the contextual influences are important, which might also influence the choice of the musical activity, strategies, and the strength of these mechanisms' impact on emotion regulation (Baltazar, 2019).

Emotion Dysregulation in Depression and Music Use

As mentioned above, emotion regulation is critical for physical and mental health, which is why implications of the disruptions in emotion regulation are studied in the context of mental disorders along with the understanding of how music can be used to support emotion regulation in adolescents who are at high risk for experiencing mental health issues. Emotion regulation is especially impaired in mood disorders (Millgram et al., 2020). Besides, most research have been conducted with populations who experience clinical depression. Therefore, this paper will also focus on studies investigating emotion dysregulation and the use of music among adolescents with depression.

Due to the disruption of emotion regulation mechanisms in depression, some researchers suggest that depressed individuals may not use emotion regulation strategies adaptively (Carlson et al., 2015; Stewart et al., 2019). For example, Millgram et al. (2015) suggest that depressed individuals are more likely to use emotion regulation strategies of *rumination* and *avoidance* while listening to music concerning negative emotions. This impairment in the use of emotion regulation strategies can also be observed in music choice (Sakka & Juslin, 2018).

One study reveals that when depressed individuals are presented with options to listen to either neutral, sad, or happy music; they are more likely to choose the sad music (Millgram et al., 2015). One explanation for this finding may be depressed individuals' propensity to direct their attention towards negative stimuli. (Garrido & Schubert, 2015). Additionally, they suggest that individuals show this preference for negative stimuli although it is maladaptive because it attenuates negative experiences. This attenuation of negativity then will lead to an experience of catharsis and *discharge* from negative emotions, which has beneficial effects of emotional release.

Another reason is depressed individuals' lack of awareness about the damaging effects of their listening habits on their mental health (Garrido & Schubert, 2015; Stewart et al., 2019). Therefore, Stewart et al. (2019) suggests raising awareness through intervention programs regarding the unhealthy effects of some emotion regulation skills (e.g., rumination), so that it would enhance adolescents' use of emotion regulation strategies through music. Although depressed individuals are likely to use maladaptive regulation strategies, Yoon et al. (2020) state that when depressed individuals choose to listen to sad music, they experience lower levels of sadness in the long run.

In contrast, Sakka and Juslin (2018) elucidate that depressed individuals do not differ from the individuals in the control group in their use and choice of maladaptive strategies. Some researchers emphasize that what is important to consider for the emotional outcome is the intention and the strategies with which adolescents choose what type of music they want to listen to (Sakka, & Juslin, 2018; Yoon et al., 2020). These are important because different intentions and the use of different strategies during music listening may lead to different outcomes regarding whether individuals with depression feel better or worse after listening to music. Additionally, Garrido and Schubert (2015) suggest a distinction between *rumination* and *reflection* strategies to explain these two contradictory findings for the adaptiveness of music choice by depressed adolescents. They suggest that adolescents who use a more ruminative style may get *absorbed* into sad music, which will perpetuate existing depressive feelings and tendencies towards these feelings. Contrary, individuals who use the *reflective* style with sad music may also think about negative emotions and experiences. However, they tend to find solutions to the problems that can lead to depressive feelings and acknowledge the emotional states within that thinking process of finding solutions (Garrido & Schubert, 2015). In short, according to Garrido and Schubert (2015), the effects of listening to sad music may depend on the extent to which individuals use ruminative or reflective emotion regulation strategies.

Similarly, Miranda and Claes (2009) studied the mediator effects of three coping styles: problem-oriented, emotion-oriented, and avoidance-oriented coping styles, on different intensities of depression in adolescents. The findings of this study suggests that problem-oriented emotion regulation strategy while listening to music may help adolescents reflect on their experiences and possibly find solutions to the problems which led them to experience depressive feelings as a result of this reflection process, which has been found to be related with lower levels of depressive symptoms especially among girls. In addition, Miranda and Claes (2009) accentuate the lyrics' importance and how they might provide adolescents with solutions and different ways of looking at their problems. For example, an adolescent who experience difficulties might benefit positively from a song containing lyrics about how the singer/songwriter dealt with his/her breakup. These adaptive ways of looking at the problems might decrease the severity of depressive symptoms when problem-focused coping strategy is used (Schwartz & Fouts, 2003; Saarikallio & Erkkila, 2007).

Furthermore, the higher use of emotion-focused regulation strategies is associated with more severe depressive symptoms, especially among young males (Miranda & Claes, 2009). Researchers explain this finding by suggesting that men are more likely to engage in emotional venting of negative emotions (e.g., anger), which in turn is associated with more depressive feelings among boys; however, this interpretation warrants further investigation (Miranda, & Claes, 2009). In other words, the use of maladaptive emotion-focused regulation strategies (e.g., using music to vent without coming to a conclusion or solution) might lead to more depressive symptoms, especially among boys. Future studies might investigate whether these gender differences can be explained by differences in the strategies of *rumination* or *reflection*, and the differences in the intention behind using music for emotion regulation. Last, Miranda and Claes (2009) study reveal that avoidance-oriented regulation and coping style are related with higher levels of depression among adolescent girls. They further explain this difference by stating that girls might be more likely to use music as a way to avoid dealing with problems, which might perpetuate the depressive symptoms. However, more research is needed to understand why these differences exist, whether they are due to chance, and whether they are replicable in other studies.

Since using music has both healthy and unhealthy effects on emotion regulation, integrating awareness about the effects of music to intervention programs is necessary for future research. There have been some attempts to integrate these empirical findings to daily life practices. For example, Uhlig et al. (2018) have worked on an intervention program for adolescents to help them develop emotion regulation tools with music use. They have found that adolescents experienced a reduction in several mental disorder measures through a space given to them to express their feelings with music. In addition, Dingle et al. (2016) also have developed a therapeutic intervention program in which they have observed developments in adolescents' abilities to become more aware of their emotions, label and regulate them through music, which is associated with improved mental health. These attempts are promising as some researchers suggest that cognitive and behavioral interventions have significant positive effects on depression and individuals' emotional and cognitive reactions towards emotional triggers (e.g., Hou et al., 2017).

Limitations in the Literature and Further Research

This research on emotion regulations has its own limitations. The most prominent one is the interchangeable use between the terms *emotion* and *affect* (Sakka & Juslin, 2018). It is important to distinguish between these terminologies because use of different concepts require different operationalizations and measures. Another limitation is the limited choice and use of music genres in empirical studies (Thoma et al., 2012). For example, it has been observed that 48% of the musical stimuli used in studies are from classical music (Cook et al., 2019; Eerola & Vuoskoski, 2013). The lack of variation in music choices is an important limitation because different genres of music create different contextual effects (Thoma et al., 2012). Therefore, relying solely on specific genres of music will not help researchers uncover the differential effects of various music genres on adolescents' emotional experiences. Some researchers also have criticized that the most research in this area are conducted with samples from West and non-minority groups, which poses a great threat to external validity of the studies (Miranda, 2013; Moore, 2013).

In addition, to get a complete understanding of adolescents' experiences, Stewart et al. (2019) and Moore (2013) suggest conducting studies with samples from different cultures so that researchers can observe the variations within the reasons for use of music and the musical strategies used to regulate emotions. By taking the abovementioned limitations into account, Cook et al. (2019) suggest future studies to investigate the reasons behind adolescents' choices of musical strategies and music genres for emotion regulation. Last, since emotion regulation is of high importance to mental health, researchers also emphasize the need for studies with clinical samples to understand whether musical interventions actually improve adolescents' psychological health in the long term (Dingle et al., 2016).

Conclusion

Music plays a crucial role in adolescents' lives. Some researchers emphasize the importance of studying the music's effect on adolescents since it is used widely. One of the ways in which music is used by adolescents is through emotion regulation. Researchers have observed different emotion regulation strategies, goals, and underlying mechanisms that are responsible for emotional changes through music. These effects differ among individuals, musical and contextual factors.

Due to emotion dysregulations' implications on adolescents' experiences, many researchers study how emotion regulation through music can both hinder and improve mental health. Different intentions for listening to the music and use of emotion regulation strategies are considered to be responsible different effects of music on several depressive symptoms. The impact of healthy and unhealthy emotion regulation strategies through music and the importance of raising awareness about these impacts eventually have led to the development of intervention programs and therapeutic interventions, given how damaging emotion dysregulation can be to adolescents.

Nevertheless, even though both contradicting and promising findings exist, this is an emerging field of research which requires further attention and investigation on (1) understanding more complex functioning of emotion regulation while listening to music and (2) whether culture, gender and music genres make a difference in this functioning.

References

Baltazar, M. (2019). Musical affect regulation in adolescents: A conceptual model. In K. McFerran, P. Derrington, & S. Saarikallio (Eds.), *Handbook of Music, Adolescents, and Wellbeing* (pp. 65–74). Oxford University Press.

Carlson, E., Saarikallio, S., Toiviainen, P., Bogert, B., Kliuchko, M., & Brattico, E. (2015). Maladaptive and adaptive emotion regulation through music: A behavioral and neuroimaging study of males and females. *Frontiers in Human Neuroscience*, *9*(466). https://doi.org/10.3389/fnhum.2015.00466

Cook, T., Roy, A. R. K., & Welker, K. M. (2019). Music as an emotion regulation strategy: An examination of genres of music and their roles in emotion regulation. *Psychology of Music*, *47*(1), 144–154. https://doi.org/10.1177/0305735617734627

Dingle, G. A., Hodges, J., & Kunde, A. (2016). Tuned in emotion regulation program using music listening: Effectiveness for adolescents in educational settings. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00859

Garrido, S., & Schubert, E. (2015). Music and people with tendencies to depression. *Music Perception*, *32*(4), 313–321. https://doi.org/10.1525/mp.2015.32.4.313

Hou, J., Song, B., Chen, A. C. N., Sun, C., Zhou, J., Zhu, H., & Beauchaine, T. P. (2017). Review on neural correlates of emotion regulation and music: Implications for emotion dysregulation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00501

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*,*31*(5),559–621. https://doi.org/10.1017/s0140525x08005293

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, *102*, 108-134. https://doi.org/10.1348/000712610X506831

Millgram, Y., Huppert, J. D., & Tamir, M. (2020). Emotion goals in psychopathology: A new perspective on dysfunctional emotion regulation. *Current Directions in Psychological Science*, *29*(3), 242–247. https://doi.org/10.1177/0963721420917713

Millgram, Y., Joormann, J., Huppert, J. D., & Tamir, M. (2015). Sad as a matter of choice? Emotion-regulation goals in depression. *Psychological Science*, *26*(8), 1216–1228. https://doi.org/10.1177/0956797615583295

Miranda, D. (2013). The role of music in adolescent development: Much more than the same old song. *International Journal of Adolescence and Youth*, *18*(1), 5–22. https://doi.org/10.1080/02673843.2011.650182

Miranda, D., & Claes, M. (2009). Music listening, coping, peer affiliation and depression in adolescence. *Psychology of Music*, *37*(2), 215–233. https://doi.org/10.1177/0305735608097245

Moore, K. S. (2013). A systematic review on the neural effects of music on emotion regulation: Implications for music therapy practice. *Journal of Music Therapy*, *50*(3), 198–242. https://doi.org/10.1093/jmt/50.3.198

Papinczak, Z. E., Dingle, G. A., Stoyanov, S. R., Hides, L., & Zelenko, O. (2015). Young people's uses of music for well-being. *Journal of Youth Studies*, *18*(9), 1119–1134. https://doi.org/10.1080/13676261.2015.1020935

Rentfrow, P. J. (2012). The role of music in everyday life: Current directions in the social psychology of music. *Social and Personality Psychology Compass*, *6*(5), 402–416. https://doi.org/10.1111/j.1751-9004.2012.00434.x

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of Music*, *39*(3), 307–327. https://doi.org/10.1177/0305735610374894

Saarikallio, S. (2012). Development and validation of the brief music in mood regulation scale (B-MMR). *Music Perception*, *30*(1), 97–105. https://doi.org/10.1525/mp.2012.30.1.97

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, *35*(1), 88–109. https://doi.org/10.1177/0305735607068889

Saarikallio, S., Nieminen, S., & Brattico, E. (2012). Affective reactions to musical stimuli reflect emotional use of music in everyday life. *Musicae Scientiae*, *17*(1), 27–39. https://doi.org/10.1177/1029864912462381

Sakka, L. S., & Juslin, P. N. (2018). Emotion regulation with music in depressed and non-depressed individuals: Goals, strategies, and mechanisms. *Music & Science*, *1*, 1–12. https://doi.org/10.1177/2059204318755023

Schäfer, T., Tipandjan, A., & Sedlmeier, P. (2012). The functions of music and their relationship to music preference in India and Germany. *International Journal of Psychology*, *47*(5), 370–380. https://doi.org/10.1080/00207594.2012.688133

Understanding Childhood Amnesia: Why and How Do Infants Have It?

Deniz Gülyurt
Koç University

Childhood amnesia is the inability to remember events from one's infancy and early childhood and is a widely discussed phenomenon in memory development (Hayne & Jack, 2010). Predominant explanations highlight a barrier that is needed to be overcome for autobiographical memories to be remembered (Fivush & Nelson, 2004); such as the development of the sense of self. Another view on the opposing side argue that childhood amnesia is a result of the inaccessibility to early memories, rather than the absence of them (Howe, 2018). More recent understanding of development of autobiographical memory, however, points that childhood amnesia is not an all-or-none phenomenon (Fivush & Nelson, 2004). It neither solely depends on the development of certain cognitive skills to enable encoding autobiographical memories, nor does it on the increasing similarities between the cognitive skills of a person between the time at encoding and the time at retrieval. The interplay between these two mechanisms is the key to explain overcoming childhood amnesia. Bauer's (2015) Complementary Processes Account proposes a comprehensive theory on this interplay between the traditionally argued mechanisms. In this review, both the traditional accounts and Bauer's (2015) more recent account is discussed to demonstrate that the Complementary Processes Account best explains childhood amnesia.

Keywords: Childhood amnesia, autobiographical memory, Complementary Processes Account.

Scarcity of memories from the first years of one's life has drawn attention in memory research for many decades. The topic still holds its mystery, and there is yet to be a consensus over the underlying mechanisms of what researchers refer to as *childhood amnesia* (Newcombe et al., 2007). The term is used to define lack of episodic memories for events and experiences that are about one's self (Bauer, 2015), namely, autobiographical memories. Childhood amnesia is restricted to autobiographical memories as the phenomenon is not observed for children's semantic or non-declarative memories. Hayne & Jack (2010) highlighted the importance of a clear conceptualization of autobiographical memories to understand the mechanisms involved in childhood amnesia; however, there is not a consensus on the definition of autobiographical memory (Bauer, 2015). This review focuses on different perspectives regarding autobiographical memory and their implications for possible mechanisms of childhood amnesia, which will be discussed as support for Bauer's (2015) Complementary Processes Account.

Autobiographical Memory and Early Memories

A memory is named autobiographical only if it is about one's personal and specific past events and experiences (Bauer, 2015). Akhtar et al. (2019) claimed that the formation of autobiographical memories requires the development of the sense of self, which is necessary for the organization and regulation of an experience, and consequently, its memory. The sense of self develops after the first 2 years of life; hence, Akhtar et al. (2019) propounded that autobiographical memories cannot exist before that age. Akhtar et al. (2019) found the "specific event" criteria mentioned in Bauer's (2015) definition worth discussing: Conway (2005, 2009) showed that memories of specific events (i.e., episodic memories) are fragmentary and time-compressed, requiring conceptual-knowledge networks to coherently encode these memories. Therefore, Akhtar et al. (2019) claimed that the development of such knowledge networks is also necessary for the development of autobiographical memories. Nelson (1993, as cited in Bauer, 2015) added another criteria for autobiographical memories as being of discrete and unique events (in contrast to recurring events), being expressed verbally, and being long-lasting. Some researchers tend to include narrative thinking (Akhtar et al., 2018) and expression in life stories (Fivush et al., 2011). Narrative thinking refers to thinking through internal mental representations (Goldie, 2012) that are formed through conceptual-knowledge networks with the sense of self.

This disagreement on the criteria for autobiographical memory brings up the question of whether childhood amnesia is suddenly overcome at distinct stages or continuously lost with the development of autobiographical memory throughout childhood. According to Akhtar et al. (2019), the sense of self is required for encoding autobiographical memories. For instance, there would be a sudden improvement in autobiographical memory at the age of 2 with the development of the sense of self. On the other hand, if the self can be integrated into memories of events that occurred before the sense of self developed, there will be a gradual improvement in autobiographical memories. In fact, Hayne & Jack (2010) reported that when the number of memories people had from their childhood were averaged between participants, the mean number of memories from age 2 to 8 yielded a smoothly increasing slope. However, the distribution of one individual's memories showed two distinct stages, the timings of which varied across participants (Hayne & Jack, 2010).

Newcombe et al. (2007) proposed a two-stage model of childhood amnesia that explained the abrupt changes in the number of memories. Newcombe et al. (2007) identified the first two years of life with dense amnesia as people had no memories of this period. The offset of dense amnesia was around age 2, overlapping with the emergence of the sense of self. The second stage Newcombe et al. (2007) identified was relative amnesia from the age of 2 until the ages of 5 to 7, when adults had few memories. The number of memories they had from this period was significantly smaller than the number of memories the forgetting curve would predict (Bauer, 2015). This is why the period was identified with amnesia although adults had memories from that time. The two-stage model is one of the traditional views that offers a one-way mechanism of how childhood amnesia is overcome: development of autobiographical memories depended on the development of certain cognitive skills that enabled encoding, consolidation, and consequently retrieval of autobiographical memories. It is also a popular view as it proposes a simple mechanism depending on the development of the sense of self, without the complications of the continuous integration of the other cognitive skills (semantic knowledge networks, narrative thinking, verbal expression) argued by other researchers.

A problem with the two-stage model of Newcombe et al. (2007) is crucial for overall childhood amnesia research: the stages are identified based on participants' dating of reported memories. Dense amnesia is identified not only with the absence of sense of self but also with the absence of memories (that is proposed as a consequence of the former), the duration of which is terminated with people's reports of their earliest memory.

The earliest memories of people, however, are problematic to set a boundary for the emergence of autobiographical memory. Peterson (2021) argues that such accounts underestimate the fluidity of early memories regarding content and dating. What people report as their earliest memory is open to change due to the time passed between the event and its report, one's lifestyle, experiences at the time of report, and the methodology used in memory research. For example, using written interviews for earliest memory recall, priming participants before memory recall or participants overhearing others recalling an early memory, and additional memory tasks before earliest memory recall significantly lower participants' age of first memories (Peterson, 2021). Such factors of memory fluidity imply that the emergence of autobiographical memory does not necessarily have to overlap with the emergence of specific cognitive abilities such as sense of self or narrative thinking that enable memory encoding.

The fluidity of earliest memories raises the following question: If the earliest memories are prone to be influenced by various factors, are they real or fictional memories? Akhtar et al. (2019) argued that early memories are not evidence of autobiographical memories because many details of memory reports could be based on one's conceptual knowledge about the world or of one's knowledge about their childhood that is acquired throughout life (e.g. from family stories). The memory fragments can also be altered to fit into one's life story (Akhtar et al., 2019). Additionally, Akhtar et al. (2018) hypothesized that early memories (especially before age 2) were only fragments of memories lacking rich descriptions. They conducted a large-scale web-based survey to collect participants' estimates of their age of first memories (i.e., age at encoding). Surprisingly, 39% reported memories before age 2, and 14% before age 1. Because children at those ages lack critical cognitive abilities, Akhtar et al. (2018) argued that the results can be explained by misdating and integration to life story, approaching the dating of autobiographical memories as predominantly inferential. However, they could not verify the exact dates of those memories from external sources (e.g., dated photographs) and validate their argument. In many studies with date verification, dates of reported memories are indeed dominantly accurate (Bauer, 2019). If anything, people tend to forward telescope, date a memory chronologically later than when it actually happened (Peterson, 2021), as they believe that an event must have happened more recently if they can recall it despite the change in their cognitive abilities and in the contextual cues between the age at event and age at recall (Howe, 2018; Bauer, 2019).

Another argument for early memories being fictional is memory integration to life stories. Episodic memory-like fragments are enriched in detail and personalized by knowledge derived from external sources like photographs or family stories without conscious efforts (Akhtar et al., 2018). Akhtar et al. (2018) explained that all improbably early memories (before the age of 2) are about age-appropriate events, coherent in a life story with successive periods defined by distinctive content and events. Together with the findings indicating that earliest memories' can be experimentally manipulated (Peterson, 2021), Akhtar et al. (2018) elucidated that all memories contained some degree of fiction, suggesting that the memory is constructive. However, their argument is refuted by several research findings such as Simcock and Hayne's (2002), showing that children could talk about their preverbal experiences as demonstrated by usage of newly acquired words to describe events before word acquisition. Similar findings suggest that children can still encode autobiographical memories in the absence of specific cognitive concepts such as language and the sense of self (Simcock & Hayne, 2002).

So far, the literature shows that the development of autobiographical memories does not depend only on memory encoding with specific cognitive skills; yet, developing cognitive skills are important for memory storage and retrieval too. In other words, an autobiographical memory can be encoded in the absence of the sense of self, and the sense of self can later be integrated into the encoded memory, which will strengthen its consolidation and facilitate its retrieval. Successful storage of memories is important as memories encoded early in life are fragile due to relative paucity of conceptual-knowledge networks and cognitive skills to bind elements, which makes them prone to be rapidly forgotten (Howe, 2018). Furthermore, successful retrieval is critical as early memories are constrained by the context in which they are encoded (Howe, 2018).

On the other hand, even if these memories are held in storage, they might never be retrieved if necessary conditions (similar contexts and cues) are not provided. The argument is known as Tulving's Encoding Specificity Hypothesis and highlights the discrepancy between the way early memories are encoded and the cues used to retrieve them (Hayne & Jack, 2010). This hypothesis also claims that encoding specificity is at its extreme in early development: young children can recall their memories only when the cues almost exactly match the stimuli present at the time of encoding (Hayne & Jack, 2010). Hayne and Jack (2010) reported that bilingual children could remember events that occurred in one of their languages better when the cues were in the same language. Thus, the discrepancy between cognitive and environmental conditions at the time of encoding and retrieval is important for explaining childhood amnesia. This is another traditional view that explains childhood amnesia as a result of inaccessibility to encoded and stored memories due to the mentioned discrepancy. This view suggests that the effects of childhood amnesia might be reduced by improvements in retrieval as a result of reduced discrepancy in cognitive skills between the time of encoding and retrieval through developments such as conceptual-knowledge networks and relational thinking.

Complementary Processes Account

Bauer (2015) developed a novel account of childhood amnesia that differs from the traditional views. According to Bauer (2015) the traditional views focus on two opposite mechanisms: late development of skills and processes that enable encoding, consolidation, and retrieval of autobiographical memories; or decrease in the functional disappearance and inaccessibility to encoded and stored memories. Considering research findings supporting both accounts, Bauer (2015)'s Complementary Processes Account is a plausible explanation for childhood amnesia, which is based on a definition of autobiographical memory as a system involved in the formation, retention, and retrieval of episodic memories that are spatially, temporally localized and are self-relevant. Bauer's account (2015) suggested a gradual increase in the quality of autobiographical memories and a gradual decrease in the vulnerability of those memories to forgetting in parallel. In contrast to traditional accounts that are built on discontinuous processes due to sudden improvements in encoding or retrieval, Complementary Processes Account accentuates the continuities in memory over development. Autobiographical memories are formed even in infancy as a function of immature neural substrates. Newly formed memory representations' quality increases with development and forgetting past events slows down as a result of improved cognitive operations that support memory recollection.

According to Bauer (2015) the core of the first process is that with increasing autobiographical features included in the memory trace due to increasing cognitive and linguistic development, the quality of autobiographical memories improves, resulting in more integrative encoding, more elaborative storage, and easier retrieval. This process rejected the traditional view that children lack abilities to detect and encode autobiographical features of events (e.g., the relation of self to event); because evidence showed cases of poor autobiographical memory from periods of life where the cognitive and linguistic skills were present. Bauer (2015) discussed that knowledge of "who, what, where, when, why, how" was present before children were 2 years old. This information was a critical feature of autobiographical memories, yet it was missing in children's reports although their understanding of that information existed. Therefore, Bauer (2015) rejected the conclusion that autobiographical memory did not fully emerge. Instead, she concluded that memories gradually become more autobiographical. An appropriate example to compare these conflicting processes stems from children's verbal reports of their early memories. When these verbal reports lacked some of the autobiographical features, researchers from traditional view concluded that children did not yet have the faculties to encode and store those features, and therefore, their memories were not autobiographical (Bauer, 2015). In Bauer's (2015) account, however, absence of some of the features in verbal reports do not imply the absence of those features from autobiographical memory, and memories can still be autobiographical if they miss some of the features but not all.

The second process that Bauer (2015) explained as complementary to the first one is the decline of the vulnerability of memories to forgetting over development. Bauer et al. (2007) found that children's forgetting function is different from adults' as the former is an exponential and the latter is a power function. The different forgetting functions show that younger children forget their memories more rapidly than older children and adults. Bauer (2015) concluded that childhood amnesia was partly a result of exponential forgetting in childhood. In contrast to the traditional view that explained forgetting as early memories becoming inaccessible for later recollection, Bauer (2015) claimed that forgetting is exponential because it is a process of inefficient cognitive functioning for recollection of memory traces that were already poorly encoded due to the paucity of the same cognitive skills. As the cognitive skills to recollect memory traces develop while the traces also start to be encoded with enriched details, the forgetting slows down exponentially. As an example, the traditional view would argue that memories before age 2 cannot be remembered as the sense of self does not exist before age 2. Because of the present sense of self at age of retrieval, the memories before age 2 become inaccessible. According to the Complementary Processes Account, however, the self can be integrated into existing memories (Bauer, 2015). Forgetting, in this sense, is a result of the combination of absence of self in the present memory representations and poor cognitive operations to integrate self into existing memories. With both processes developing, it becomes much easier for older children to detect their self-relevance in their memories they encoded with self-relevance in the first place.

Conclusions

The Complementary Processes Account is plausible in the sense that it can explain most findings, even the seemingly contradicting ones. Children’s preverbal memories might be more difficult to verbally report, but they can still be remembered. Neither the existence of specific cognitive skills nor the encoding specificity can fully determine whether an early memory is encoded or can be retrieved. The Complementary Processes Account, at this point, proposes a complex, yet comprehensive account of how the present cognitive skills at the time of encoding, developing cognitive skills during their storage, and the newly acquired cognitive skills at the time of retrieval are all important factors in the development of autobiographical memory. It also successfully explains how autobiographical memory contributes to the continuity of one’s narrative and the sense of self over time, which the traditional views could not achieve due to sudden changes in autobiographical memory they suggested. Even though a full consensus over what autobiographical memory and childhood amnesia is hardly achievable, this framework can at least provide a common ground for all lines of research to operate and prevent overlooking possible mechanisms of childhood amnesia.

References

Akhtar, S., Justice, L., Morrison, C., & Conway, M. (2018). Fictional First Memories. *Psychological Science*, *29*(10), 1612-1619. doi: 10.1177/0956797618778831

Akhtar, S., Justice, L., Morrison, C., Conway, M., & Howe, M. (2019). What Are Autobiographical Memories? A Reply to Bauer, Baker-Ward, Krøjgaard, Peterson, and Wang (2019). *Psychological Science*, *30*(9), 1400-1402. doi: 10.1177/0956797619868994

Bauer, P. (2015). A complementary processes account of the development of childhood amnesia and a personal past. *Psychological Review*, *122*(2), 204-231. doi: 10.1037/a0038939

Bauer, P., Baker-Ward, L., Krøjgaard, P., Peterson, C., & Wang, Q. (2019). Evidence Against Depiction as Fiction: A Comment on "Fictional First Memories" (Akhtar, Justice, Morrison, & Conway, 2018). *Psychological Science*, *30*(9), 1397-1399. doi: 10.1177/0956797619834510

Bauer, P. J., Burch, M. M., Scholin, S. E., & Güler, O. E. (2007). Using cue words to investigate the distribution of autobiographical memories in childhood. *Psychological Science*, *18*, 910–916. http://dx.doi.org/10.1111/j.1467-9280.2007.01999.x

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*, 594–628.

Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, *47*, 2305–2313.

Fivush, R., & Nelson, K. (2004). Culture and Language in the Emergence of Autobiographical Memory. *Psychological Science*, *15*(9), 573-577. https://doi.org/10.1111/j.0956-7976.2004.00722.x

Hayne, H., & Jack, F. (2010). Childhood amnesia. *Wires Cognitive Science*, *2*(2), 136-145. doi: 10.1002/wcs.107

Howe, M. (2018). Unravelling the nature of early (autobiographical) memory. *Memory*, *27*(1), 115-121. doi: 10.1080/09658211.2019.1537140

Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review*, *104*(3), 499–523. https://doi.org/10.1037/0033-295X.104.3.499

Nelson, K. (1993). Events, narratives, memory: What develops? In C. A. Nelson (Ed.), *The Minnesota symposium on child psychology: Volume 26*. *Memory and affect in development* (pp. 1–24). Hillsdale, NJ: Erlbaum.

Newcombe, N. S., Lloyd, M. E., & Ratliff, K. R. (2007). Development of episodic and autobiographical memory: a cognitive neuroscience perspective. *Advances in child development and behavior*, *35*, 37–85. https://doi.org/10.1016/b978-0-12-009735-7.50007-4

Peterson, C. (2021). What is your earliest memory? It depends. *Memory*, *29*(6), 811-822. doi: 10.1080/09658211.2021.1918174

Simcock, G., & Hayne, H. (2002). Breaking the barrier? Children fail to translate their preverbal memories into language. *Psychological Science*, *13*, 225–231. doi:10.1111/1467- 9280.00442