



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Decision-making and Memory:  
An Investigation on the Recollection of a Moral Dilemma**

Carolane Loren Helin



**THE UNIVERSITY**  
*of* **EDINBURGH**

Doctor of Philosophy (PhD)

The University of Edinburgh

2022

## **Abstract**

Individuals inevitably face situations where they have to choose between several options with uncertain future outcomes. This decision-making process can occur in various contexts such as getting the COVID vaccine, voting in a referendum, changing electricity provider, buying or renting a house, etc. It is, therefore, not surprising that decision-making has been the focus of a wide range of fields, including economics, philosophy, marketing, neurosciences and psychology (Johnson & Busemeyer, 2010).

In cognitive psychology, there is a consensus around the idea that decision-making is highly related to memory (Fellows, 2018). The relationship between decision-making and memory is thought to be bi-directional, where both cognitive functions have a mutual impact (Tremel et al., 2018). However, empirical research has mainly investigated the effect of memory on decision-making. The impact of decision-making on memory has yet to be fully characterised (Murty et al., 2015).

Previous research has suggested that decision-making may influence specific aspects of memory, such as learning and forgetting (Murty et al., 2015; Murty et al., 2019). This thesis investigates the effect of decision-making on memory through the lens of learning and forgetting. Chapter 1 presents the development of a longitudinal experimental paradigm eliciting decision-making and memory in a set of three experiments - Experiments 1a, 1b, 2 and 3. Experiment 1 describes the retention of a standardised piece of prose (Logical memory Story C subtest from Wechsler Memory Scale -- fourth edition; Wechsler, 2009) under normal forgetting and retrieval practice (a learning strategy that enhances long-term retention) with a delay of up to one month. Experiment 1a illustrates the forgetting curve of Story C over time. Experiment 1b shows how the retention of Story C is enhanced under retrieval practice.

Experiment 2 analyses the benefit of retrieval practice on the retention of Story C and a dilemma story. These results suggest that the dilemma story is more difficult to encode, as it has lower immediate and delayed retention than Story C. However, the difference in retention between the two stories remains constant over time - the stories are forgotten to the same extent. Experiment 3 compares the percentage of moral decision types (i.e., deontological or utilitarian) provided for a set of popular dilemma stories. The results reveal that the *Nobel Prize* dilemma has the most even split in chosen option for decision. From these three studies, I developed an experimental paradigm that concomitantly manipulates decision-making and measures memory retention.

Chapter 3 presents two experiments focusing on the effect of decision-making and moral decision types on forgetting. Experiment 4 focuses on the retention of the dilemma story following a deontological decision, a utilitarian decision, or an unrelated choice (control). The results suggest that decision-making and moral decision types do not influence memory retention after a day. Experiment 5 further investigates the impact of moral decision types on memory retention over a month. The results replicate the absence of an effect found in Experiment 4. Chapter 3 concludes that decision-making may not systematically influence memory, regardless of moral decision types.

Chapter 4 explores the role of learning in the relationship between decision-making and memory. Experiment 6 focuses on the effect of decision-making and moral decision types on memory retention as a function of retrieval practice and forgetting. This experiment aims to test whether learning can change the impact of decision-making on memory. The results revealed that decision-making does not impact forgetting regardless of moral decision types and retrieval practice.

Chapter 5 concludes and discusses the research findings. Overall, my research does not provide evidence of an impact of decision-making on episodic memory. The discussion acknowledges the limitations of the current thesis. Then, explanations for the consistent lack of effects observed in the thesis are considered. Finally, I discuss theoretical and methodological implications and future directions.

## **Lay summary**

Decision-making is the process involved when facing situations involving a choice between several options with uncertain future outcomes. This process can occur in various contexts: Should I get the COVID vaccine? What should I vote for in the next referendum? Should I change my electricity provider? Should I buy or rent a house? Because decision-making impacts health, politics, economics, or even social issues, it has raised the interest of many academic fields.

In cognitive psychology, it is believed that decision-making both rely upon and influence memory. Research has provided supporting evidence that memory of past events influences decision-making. However, there is a dearth of proof that decision-making also impacts memory. This thesis aims to fill the gap between theory and evidence on the effect of decision-making and memory.

In line with previous studies, this research focuses on specific aspects of memory, such as learning and forgetting. Decision-making is investigated using moral dilemmas based on the famous Trolley dilemma. Chapter 2 presents a set of three experiments - Experiments 1, 2, and 3 on the development of a new experimental design. This design measures the forgetting of the dilemma story as a function of the moral decision made. Chapter 3 presents two experiments studying the impact of decision-making and moral decision types on forgetting after one day (Experiment 4) and after one month (Experiment 5). The results suggest that decision-making does not impact long-term forgetting, regardless of the type of moral decision. Chapter 4 describes Experiment 6, studying whether promoting learning can change the impact of decision-making on forgetting. The findings corroborate the previous chapter - decision-making does not influence forgetting regardless of moral decision-type and learning. Overall, the thesis did not bring any proof that decision-making affects

memory. The conclusion implies that theories on decision-making should be considered. Future research should focus on further characterising the relationship between the two processes.

### **Declaration**

I declare that this thesis presented for the degree of Doctor of Philosophy (PhD), has

- i) been composed entirely by myself,
- ii) been solely the result of my own work,
- iii) not been submitted for any other degree or professional qualification.

Carolane Loren Helin

March 2022

*I dedicate this thesis to my step-father, Christian, who got me to where I am today.*

## Acknowledgements

I would like to thank my supervisors, Dr Sarah MacPherson, and Pr Sergio Della Sala. Thank you for your guidance, your support, your advice, and your commitment during the past four years. Thank you for sharing your time with me whenever I needed it. I feel lucky to have the opportunity to work by your side.

I would like to thank my parents, Michèle and Christian, as well as my little sisters Aline and Candice. Thank you for always believing in me. Thank you for your unconditional love and support.

I would like to thank my father, Gilles, and my grandmother, Jeannine. Thank you for being always there for me, any time day or night. Thank you for being so loving and supportive.

I would also like to thank my partner, Julien. Thank you for standing by my side through good and bad. Thank you for reminding me what I am capable of when I lack self-confidence.

Thank you to my entire family, my grandmother Françoise, my great grandparents, Gabrielle and Jacques, my aunts, uncles, and cousins. I am lucky to have such a big and loving family.

I am very grateful to all my friends. To name but a few, thank you to Nicolas and Elisa. Thank you for your commitment to our friendship despite the 1673 kilometres between us. Thank you for being here when I needed you most. Thank you to Esperanza and Bérengère. Thank you for cheering me up when I felt down. Thank you for always making me laugh. Thank you for all the time spent proofreading this thesis with me. I am very lucky to have you as my friend. Thank you to all my other friends from the psychology department, Riccardo, Shereen, Tanvi, Loris, Jan-Phillip, Andreea and everyone else.

I would also like to thank my colleagues and my friends from Queen Margaret University and Edinburgh Napier University. To name but a few, thank you to Steve, Clare, Toni, Kristen, Olivia, Rory, Duncan, Elli, Lee and everyone else. Thank you for the amazing opportunities you have offered me. Thank you for your guidance and support.

Finally, I would like to thank the three volunteer research assistants, Caroline, Milla and Grace, who assisted me in participant recruitment and data collection. Thank you to all the participants who took part in my experiments.

A warm thought goes to my late grandfather, Gérard, who would have been extremely proud of my achievement.

## Table of Contents

<b>Abstract</b> .....	<b>1</b>
<b>Lay summary</b> .....	<b>4</b>
<b>Declaration</b> .....	<b>6</b>
<b>Acknowledgements</b> .....	<b>8</b>
<b>Table of Contents</b> .....	<b>9</b>
<b>Chapter 1: Introduction</b> .....	<b>12</b>
1.1    Decision-making.....	12
1.1.1    Definition.....	12
1.1.2    Investigation streams and seminal theories.....	18
1.1.3    Categorising decision-making.....	24
1.1.4    Decision-making in moral dilemma tasks.....	26
1.2    Decision-making and episodic memory .....	32
1.2.1    Correlational evidence of a relationship between decision-making and memory.....	36
1.2.2    Causal evidence of the effect of episodic memory on decision-making .....	38
1.2.3    Causal evidence of the effect of on decision-making on episodic memory .....	41
1.2.4    The study of episodic memory with forgetting and learning paradigms. ....	45
1.3    Overview of the present research.....	47
<b>Chapter 2: Development of a new experimental paradigm for investigating decision-making and memory</b> .....	<b>49</b>
2.1    Introduction .....	49
2.2    Experiment 1: Investigating the Retention of a Piece of Prose Under Long-Term Forgetting and Retrieval Practice.....	53
2.2.1    Experiment 1a: The Forgetting of a Piece of Prose. ....	59

2.2.2	Experiment 1b: The retention of a Piece of Prose. Under Retrieval Practice with Incidental Free Recall .....	74
2.2.3	Experiment 1 Discussion .....	82
2.3	Experiment 2: The Long-Term Retention of a Dilemma Story Under Retrieval Practice .....	83
2.3.1	Method.....	84
2.3.2	Statistical analysis.....	90
2.3.3	Results .....	90
2.3.4	Discussion.....	92
2.4	Experiment 3: A Comparison of Moral Decision Types in a Moral Dilemmas Set.....	93
2.4.1	Method.....	95
2.4.2	Results .....	96
2.4.3	Discussion.....	97
2.5	Chapter 2 Discussion.....	98

..... **Chapter 3: The impact of decision-making on long-term episodic memory**

..... **99**

3.1	Introduction .....	99
3.2	Experiment 4: The Impact of Decision-Making on Memory Retention after A Day.....	103
3.2.1	Method.....	104
3.2.2	Statistical Analyses .....	110
3.2.3	Results .....	113
3.2.4	Discussion.....	121
3.3	Experiment 5: The Impact of Moral Decision Types on Long-Term Forgetting Up to One Month Delay	126
3.3.1	Method.....	128
3.3.2	Statistical Analyses .....	132
3.3.3	Results .....	135

3.3.4	Discussion.....	144
3.4	Chapter 3 Discussion.....	148
<b>..... Chapter 4: The role of learning in the relationship between decision-making and memory.</b>		
<b>.....</b>		<b>153</b>
4.1	Experiment 6: The Effect of Decision-Making on Long-Term Forgetting as a Function of Learning 158	
4.1.1	Method.....	162
4.1.2	Statistical analyses .....	169
4.1.3	Results .....	173
4.1.4	Discussion.....	183
4.1.5	Chapter 4 Discussion .....	185
<b>..... Chapter 5: General Discussion</b>		
<b>.....</b>		<b>188</b>
5.1	Summary of the findings.....	188
5.2	Limitations.....	189
5.3	Implications.....	192
5.3.1	Theoretical implications.....	192
5.3.2	Methodological implications.....	193
5.4	Future directions.....	194
5.5	Conclusion.....	195
<b>References.....</b>		<b>196</b>
<b>Appendixes.....</b>		<b>217</b>

## **Chapter 1: Introduction**

From getting vaccinated to voting at a referendum or buying a property, decision-making shapes individuals' lives at many levels. Therefore, most academic disciplines interested in human beings study decision-making (e.g., economics, philosophy, marketing, neurosciences, or psychology). Studies in cognitive psychology are generally concerned with the underpinning functions influencing decision-making. Contemporary research has specifically focused on how the encoding and retrieval of information from memory guide or bias decision-making. Building upon theories on the relationship between decision-making and memory, the current research takes an antithetical approach by focusing on the impact of decision-making on memory. Can the decision one make influence memories of the decision context? Cognitive theories, and potentially intuitive thinking, suggest that it does. However, empirical evidence supporting this assumption is scarce. This research aims to fill the gap between theories and evidence on the impact of decision-making on memory. This first chapter presents the theoretical grounds and rationales for the approach of decision-making adopted in the thesis. Building upon cognitive theories of decision-making, I discuss the consensus around its relationship with memory and the related limitations that motivate my research.

### **1.1 Decision-making**

#### 1.1.1 Definition

Whilst the cross-disciplinary study of decision-making allows for a broader and more complex understanding of the topic, it has also led to divergences and

inconsistencies across decision sciences (Johnson & Busemeyer, 2010). Such discrepancies are prominent when it comes to the definition of decision-making. Even within different branches of psychology such as social psychology, neuropsychology, cognitive psychology and psycho-economics, different definitions of decision-making are adopted. Moreover, it must be noted that studies on decision-making do not systematically state their definition of the process and tend to focus on explaining the potential determinants of the process (e.g., Gutbrod et al., 2006; Gupta et al., 2009; Wimmer & Shohamy, 2012; Cui et al., 2015; Alarcon et al., 2015). When available, definitions can go from the very broad: "taking an action" (Redish, 2013), to the very specific: "selecting a response option, that is, choosing a nonverbal action, a word, a sentence, or some combination thereof, among the many possible at the moment, in connection with a given situation" (p.166), which consequences are uncertain and emotionally meaningful (Damasio, 1994).

Since the literature in psychology does not provide a unique and consensual definition of decision-making, comparisons between findings can be somewhat hazardous. Therefore, the current section presents and justifies the definition of decision-making used in this thesis. This definition addresses a specific need: distinguishing decision-making from other processes. To fulfil this requirement, I start by presenting the broadest definition available. Then, I specify the description of the process by adding terms in light of theoretical and empirical evidence until I reach an unequivocal definition.

Redish (2013) provides an inclusive definition of decision-making. The author introduces the concept of decision-making as "taking an action". Here, the concept of

*action* refers to any interactions between an individual and their physical world (e.g., muscle movement, salivation, social speeches). Redish (2013) suggests that his definition implies the coexistence between multiple decision-making processes that can be physiological, emotional, or cognitive. Thus, Redish's definition (2013) intends to cover a wide range of processes rather than targeting a specific function.

In my opinion, using a single terminology to encompass such a broad scope of processes is problematic. Redish's definition (2013) places processes that are not comparable under the same label. According to this definition, both visual discrimination tasks and gambling tasks both involve decision-making. However, the two types of tasks differ on several aspects. On the one hand, visual discrimination tasks involve detecting differences between visual stimuli and rely on specific areas of the visual cortex within the occipital lobe (Taylor & Rodriguez, 2021). On the other hand, risk-taking tasks elicit situations where an action must be taken by considering known probabilities. Risk-taking tasks involve a network of brain areas such as the anterior cingulate cortex, insula, dorsolateral prefrontal cortex, orbitofrontal cortex, posterior parietal cortex, ventrolateral prefrontal cortex, ventromedial prefrontal cortex (VMPFC) and ventral striatum (Levy, 2017). The differences between the nature of each task and their neurological underpinning suggest that they do not involve similar processes.

A more parsimonious definition can be found in Newell and Shanks (2014). The authors defined decision-making as the cognitive process of choosing one amongst several actions. The authors indicate that their definition intends to distinguish decision-making from the visual system's computation of low-level properties (e.g.,

visual discrimination). Here, the concept of *choice* refers to the deliberation between multiple courses of actions. Considering decision-making as a cognitive process allows a distinction from processes that are purely physiological or emotional. Thus, Newell and Shanks' (2014) approach to decision-making as choice-making is more specific than Redish's (2013) action-taking approach.

Newell and Shank's (2014) perspective of decision-making is similar to Takemura (2021). The author describes decision-making as the conscious act of selecting an alternative from a group of alternatives (i.e., the choice of action). Takemura's (2021) definition differs from Newell and Shank's (2014) in one aspect: consciousness. Redish (2013) argues against the approach of decision-making as a conscious deliberation by indicating that not all decisions are rational. This argument may be based on the inaccurate premise that consciousness implies rationality. Moreover, assuming that decision-making involves conscious deliberations does not mean that such deliberations cannot be influenced by unconscious factors. In a short paper, Bechara et al. (1998) synthesised their findings from multiple experiments and patient studies on decision-making (Bechara et al., 1994; Damasio, 1994; Bechara et al., 1995; Bechara et al., 1996). Their research is based on an economic decision-making paradigm called the Iowa Gambling Task (Bechara et al., 1994). In the Iowa Gambling Task, participants must bet on four decks of cards. Unbeknownst to the participant, each deck is associated with specific patterns of reward and penalties. Two decks are characterised by small immediate gains and losses and overall long-term gain, while the two other entail large immediate gains and losses but overall long-term losses. Combining neuroimaging, psychophysiological and behavioural

measurements, the authors have shown that conscious reasoning in decision-making is preceded by a nonconscious biasing step that uses neural systems that differ from those that support declarative knowledge. Thus, their research provides strong evidence that decision-making involves, amongst other factors, conscious choices as suggested by Takemura (2021).

The idea that decision-making engages conscious choices seems to be generally admitted in the psychology literature. In the book chapter, Behavioural Decision-Making, Redlaw and Lau (2013) also acknowledge that decision-making implies choosing. However, they argue that thinking of decision-making as nothing more than a choice is a mistake. They suggest differentiating between choosing and deciding. According to the authors, deciding differs from choosing as decisions arise from an evaluation of potential outcomes. Let us consider the following examples. I can choose to stay hydrated by drinking sparkling rather than still water. Such a choice may be spur of the moment. It does not require further consideration of the choice outcomes. Let us now imagine a situation where I must choose a ballot when voting in a referendum. Unlike the previous example, this situation requires careful consideration of short- and long-term outcomes associated with each ballot. The evaluation of options and related outcomes characterise the concept of *judgment*. In sum, Redlaw and Lau (2013) define decision-making as the combination between choice and judgement. This idea that decision-making involves judgments and choice is shared by several authors. For instance, in their introduction to their PLOS ONE collection, Neuroscience of Reward and Decision-Making, Groman et al. (2020) define decision-making as the process of evaluating cost and benefits leading to the choice

of one alternative over others. This perspective is shared by Hastie (2001) in a literature review on contemporary developments in the field of decision-making. According to Hastie (2001), decision-making refers to the whole process of choosing a course of action and judgement is a specific component of this process. Overall, considering the role of judgement in decision-making offer a more specific perspective as it acknowledges the complexity of this process and distinguishes it from basic acts of choosing or taking actions.

Although the definition above differentiates decision-making from other cognitive processes, it may still be an incomplete perspective. Indeed, Ernst and Paulus (2005) consider an additional step in the decision-making process. They suggested that the judgment step is followed by the identification of *preferences*. While the judgement refers to the prediction of options outcome, the preference refers to the subjective weighing of the subjective importance of these outcomes (Fischhoff & Broomell, 2020). According to Brust-Renck (2021) the concept of preference in decision-making is essential as it explains why not all decisions are rational. Thus, considering preferences in decision-making addresses the criticism against the idea of rational deliberation raised by Redish (2013) and mentioned earlier in this section. Decision-making is engaged in situations where risks or uncertainties are associated with options' outcomes (Von Neumann & Morgenstern, 1944/2007). The judgment itself does not suffice in reaching a choice. A choice is made between consequences that are unknown following a preference. This consideration allows distinguishing decision-making from other high-level cognitive processes such as reasoning. To conclude on the approach adopted in this thesis, I will adopt Fischhoff and Broomell's

(2020) elegant definition; decision-making is a process of three steps: “judgment, predicting the outcomes of choosing possible options; preference, weighing the importance of those outcomes; and choice, combining judgments and preferences to make decisions”. Although the literature in psychology provides other definitions, the definition cited here provides a specific framework and is generally acknowledged in the field of judgment and decision-making (for reviews, see Brust-Renck 2021; Fischhoff & Broomell, 2020, Krueger 2011).

### 1.1.2 Investigation streams and seminal theories

Decision-making theories in psychology are generally divided into two categories: normative and descriptive (van der Pligt, 1996; Morelli, 2021). Some reviews also consider a third category, such as prescriptive theories (Brust-Renck, 2021; Takemura, 2021). However, I will not discuss prescriptive theories further since they are rather specific to sub-fields of psychology such as clinical psychology. On the one hand, normative theories aim to model optimal decision-making. They illustrate how decisions should be made using mathematical estimations. On the other hand, descriptive theories are interested in how individuals actually make decisions in context.

Historically, normative theories preceded descriptive ones and were borrowed from economics research (Brust-Renck, 2021, Takemura 2021). Edwards' seminal work, *The Theory of Decision-making* (1954), reviews theories in economics to develop the first psychology normative theory of decision-making. Based on the work of the eighteenth-century mathematician Bernoulli (1735/1954) and the economists von Neumann and Morgenstern (1944/2007), Edwards (1954) introduced the concept

of *subjective expected utility* (SEU) in psychology (van der Pligt, 1996; Weiss & Weiss, 2008). The SEU theory suggests that decisions follow a set of rules that combines beliefs considered in probabilities and preferences considered as values or utilities. The SEU theory can be expressed mathematically as the sum of the products of each outcome's probability and utility:

$$\sum_i p_i \cdot u(x_i),$$

where  $p_i$  is the probability of outcome  $i$  and  $u(x_i)$  the utility of outcome  $i$ .

Following this normative theory, people make decisions by opting for the option with the highest SEU.

Van der Pligt (1996) illustrated how the SEU theory predicts behaviours with the following example. A jobseeker is considering two offers, job A and job B. Three outcomes of equal importance are considered: the salary, career opportunities and the work atmosphere. A utility value is allocated to each outcome, for each job opportunity in terms of which opportunity offers more advantageous outcomes. The utility values are converted into subjective numbers (i.e., utilities), on a score from 1 to 10. Job A offers a higher salary than Job B, so Job A gets a higher utility value ( $u(x_a)=8$ ) than Job B ( $u(x_b)=7$ ) on the salary outcome. The salaries are known for sure as they have been provided by the firm, so their respective probability is 1. The career opportunities seem to be better in Job B ( $u(x_a)=9$ ) than in Job A ( $u(x_b)=8$ ), although the jobseeker does not have an objective measure of career opportunities. They are pretty confident in their judgement of career opportunities in Job B, but rather unsure about their estimation of career opportunities in Job A. Therefore, the utility of career

opportunities in Job A will receive a lower *subjective* probability ( $p_a = 0.7$ ) than Job B ( $p_a = 0.9$ ). Finally, the work atmosphere seems better in Job A ( $u(x_a) = 8$ ) than in Job B ( $u(x_b) = 7$ ), but the job seeker is not quite sure of the atmosphere in Job A ( $p_a = 0.8$ ), while certain in Job B ( $p_b = 0.1$ ). Overall, the SEU value Job B ( $1.0 \cdot 7 + 0.9 \cdot 9 + 1.0 \cdot 7 = 22.1$ ) is higher than Job A ( $1.0 \cdot 8 + 0.7 \cdot 8 + 0.8 \cdot 8 = 20$ ), so Job B should be the chosen option.

The SEU theory has been very influential in social psychology, more specifically in attitude research (van der Pligt, 1996). Ajzen and Fishbein's (1975) theory of reasoned action uses the assumptions above to characterise individuals' attitudes as a function of beliefs about a subject's attributes. However, the authors also highlighted the limitation of this normative theory by estimating between five to nine the capacity of beliefs to be processed. Such findings suggest that human processing of information is limited which in turn limits the descriptive validity of the SEU theory. Van der Pligt (1996) also argues that it is challenging for humans to think probabilistically. Thus, normative theories such as the SEU do not provide an efficient framework to predict or explain human behaviours.

A shift from normative to descriptive models came in 1979 with the work of Kahneman and Tversky. The authors illustrated the limitation of the SEU theory in explaining human behaviour with the *Allais paradox*. Introduced by the economist Maurice Allais, the Allais paradox (Allais, 1953) is presented by Kahneman and Tversky (1979) as follows. Let us consider Problem 1 where a choice is given between:

- Option A, which has a 0.33 probability to win 2500, a 0.66 probability to win 2400 and a 0.01 probability to win 0;

- And Option B, which has a certain probability (1) to win 2400.

When having to choose between the two options, 82% of participants preferred the certain gain (i.e., Option B). Given that SEU theory considers the utility of earning nothing is null  $u(0) = 0$ , then we can infer:

$$1 * u(2400) > 0.33 * u(2500) + 0.66 * u(2400)$$

In other words, the SEU is higher for Option A than for Option B. Participants should generally prefer the sure gain. Let us now consider Problem 2, where a choice is given between:

- Option C, which has a 0.33 probability to win 2500 and a 0.67 probability to win 0;
- Option D, which has a 0.34 probability to win 2400 and a 0.66 probability to win 0.

According to the SEU, the patterns of responses observed in Problem 1 would predict that:

$$0.34 * u(2400) > 0.33 * u(2500)$$

However, the results obtained by the authors violate the SEU theory as 83% of participants chooses Option C. Paradoxically, participants preferred Option B in Problem 1 but preferred Option C in Problem 2. However, Problem 2 is derived from Problem 1 by removing a 0.66 probability of winning 2400. The authors referred to this phenomenon as the *certainty effect*, i.e., the tendency to overweight outcomes that are certain as compared to are merely probable outcomes. The certainty effect, illustrated with the Allais paradox, shows that the SEU theory does not predict human behaviour.

Kahneman and Tversky (1979) drew on the limitations of the SEU in explaining decision-making to develop the *prospect theory* (updated in Tversky and Kahneman, 1992). The prospect theory suggests that the probability of each decision outcome is assigned a value and a weight. The prospect theory differs from the SEU theory as the utilities are replaced with the concept of values,  $v(x_i)$ , expressed in terms of gains and losses, and the objective probability  $p$  is given a weight  $\pi(p)$ . The outcome's value is considered in terms of gains or losses relative to a psychologically neutral point. The value function is concave for gains, convex for losses, and steeper for losses than for gains. Moreover, judgments for small probabilities are overweighted and for high and moderate probabilities are underweighted. When objective probabilities cannot be known, the judgment of probability associated with the decision outcome is estimated through intuitive strategies and biases (Kahneman and Tversky, 1979; for a review, see Brust-Renck, 2021). The author identified three basic intuitive strategies, called *heuristics*. The three heuristics are *representativeness* (Kahneman & Tversky, 1972), *availability* (Tversky & Kahneman, 1973), and *anchoring and adjustment* (Tversky & Kahneman, 1974).

The first heuristic presented by Kahneman & Tversky (1972) is representativeness. The representativeness heuristic stipulates that the estimated probability of an event relies on two factors: (1) the degree of similarity to its parent population, and (2) the reflection of randomness. The authors illustrate this probability with the following example. Consider a population of families of six children from a specific city. A city survey suggests that in 72 families, the exact order of birth of boys (B) and girls (G) was GBGBBG. In how many families would the order of birth be

BGBBBG? Although both sequences of birth have the same probability, the mean estimation was 30 families. Such results illustrate the first factor of representativeness: the sequence BGBBBG does not reflect the proportion of boys and girls in the population suggested by the survey (e.i, more boys and fewer girls than the parent population). In the same study, participants also judge the sequence BBBGGG as less likely than GBBGBG. The second sequence appears more random than the first one. Thus, when the judgment of probability for each decision outcome is based on its representativeness, the outcome is considered as a sample from a population to which it is compared.

If the representativeness heuristic estimates probability by assessing its connotative distance, Tversky & Kahneman (1973) suggest that probabilities can also be assessed by associative distance or availability. The availability heuristic suggests that the probability for an event to occur depends on the ease with which instances of related events can be retrieved from memory. For instance, the estimated probability of divorce rate by recalling divorces amongst one's acquaintances. In a set of ten studies, the authors provide evidence that the availability in memory leads to systematic biases in frequency judgement such as illusory correlations. The authors noted that the assessment of available does not require retrieving instances but simply estimating the ease of retrieval (this assumption is further discussed in Section [1.2](#)).

The third heuristic presented by Tversky & Kahneman (1974) is anchoring and adjustment. Following this heuristic, the estimation of values is anchored from the initial value, or starting point, and is adjusted as a function of this anchor value. One example is the estimation of the results of the following products: (A)  $8 \times 7 \times 6 \times 5 \times 4$

x 3 x 2 x 1, and (B) 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8. In both cases, the result is 40320. However, because of anchoring, the median estimates were 512 for B and 2250 for A. The authors specified that such a heuristic is employed when a relevant value is provided. Overall, the heuristics presented above provide a framework for explaining judgments of probabilities under uncertainties. The use of heuristics leads to systematic and predictable errors (Tversky & Kahneman 1974). Such findings have revolutionised research on judgment and decision-making and still influences recent development in decision-making research (Gilovich & Griffin, 2012).

### 1.1.3 Categorising decision-making

Decision-making can take various forms (Takemura, 2021). Nevertheless, decision-making tasks can usually be categorised based on two factors: the context and the outcome. The decision context refers to the nature of the decision to be made. The historical perspective presented in the previous section (1.1.2) greatly influenced the contexts used to study decision-making. Indeed, the most popular contexts for studying decision-making are economic and social situations. Economic decision-making refers to an experimental or clinical task involving a financial bet. One examples of economic decision-making is The Iowa Gambling Task (Bechara et al, 1994) presented in the previous section. Social decision-making tasks involve integrating individual interests and social values and evaluating social information (Ni & Li, 2021). A well-known example of social decision-making tasks is moral decision-making tasks. The most popular moral decision-making task was presented in Greene et al. (2001). In Greene et al. (2001), the participant read a set of 60 dilemma stories. The stories present scenarios involving conflicting moral reasons. The participants

must decide which of the two conflicting options they would endorse. This task is considered a social decision-making task since the suggested course of action often involves harming other people (MacPherson et al., 2015). Social and economic decision-making are not mutually exclusive. Indeed, experimental or clinical tasks may elicit both contexts such as the Ultimatum game. In the Ultimatum game the participant is engaged in a financial deal with a proposer. The proposer offers a certain amount of money that the participant can either accept or reject. If the participant accepts the offer, both parties will receive the agreed amount. If participants reject the offer, the none of the parties earn money. Although the participant should always accept the deal to earn money, research shows that people tend to reject offer that are unfair (less 20% of the total amount) as a way of punishing the proposer (for a review see MacPherson et al., 2015).

Economic and social situations are convenient frameworks for studying decision-making as such situations usually involve different degree of certainty about possible outcomes. In decision-making, the outcomes related to each option can be categorised as risky or uncertain. The first distinction between decision-making under risk vs uncertainty was formulated in 1921 by the economist Frank H. Knight: "there is a fundamental distinction between the reward for taking a known risk and that for assuming a risk whose value itself is not known." In other words, decision-making under risk involves decision outcomes with well-specified probabilities, while decision under uncertainty involves ambiguous outcomes (Johnson & Busemeyer, 2010). Moreover, a study by Wilson and Vassileva (2018) provides neuropsychological evidence of a distinction between the two categories of decision-making. Their

experiment examines the relationship between abstinent substance users' performance on the Iowa Gambling Task (decision-making under uncertainty, Bechara, et al., 1994), and the Cambridge Gambling Task (decision-making under risk, Rodger et al., 1999) and pathological gambling. In the Cambridge Gambling Task, the participant is presented with a set of six red and blue boxes. The ratio of each boxes presented varies from one trial to another. The participant must bet on one box to find a hidden token to win points. When a red box is selected, the participant wins 30 point if there is a token presented but losses 30 point otherwise. If the participant successfully bet on a blue box, they win 70 points, but they would lose 70 points if the token was not hidden there. Participant performance is reflected through the final point balance. Pathological gambling was assessed using the gambling subscale of the Addiction Severity Index-Lite (McLellan et al., 1992) and the DSM-IV criteria (American Psychiatric Association 1994). Their results suggest that pathological gambling in abstinent substance users is predicted by decision-making under risk but not under uncertainty. Additionally, Adjeroud et al. (2017) investigated decision-making skills in participants with premanifest and manifest Huntington's disease. Their results show that the progression of Huntington's disease is associated with impairments in decision-making under uncertainty while decision-making under risk is preserved. To conclude, a distinction can be made across social and economic contexts of decision-making and risky and uncertain decision outcomes.

#### 1.1.4 Decision-making in moral dilemma tasks.

Amongst psychology research, moral dilemmas are an extremely popular method of studying decision-making (Christensen & Gomila, 2012). However, moral

dilemmas were originally developed as philosophical thought experiments. Moral dilemmas are elicited through short stories involving a moral conflict calling upon opposing ethical reasons. The most famous example of a moral dilemma is the Trolley Dilemma (Foot, 1967/2002). In this dilemma, a trolley is running toward five railway workers. However, the trolley driver can divert it to another track where it will hit only one worker. The moral conflict resides in doing nothing and allowing the trolley to continue travelling along its original path or diverting it and sacrificing one worker to save the life of five others. The first option refers to the philosophical principle of deontology, considering that the morality of an action depends on its intrinsic nature independently of its consequences (Kant, 1785/2001) i.e., to kill is immoral. The second option follows the philosophical principle of utilitarianism, according to which the morality of an action is dependent on its consequences (Mill, 1861/1998) i.e., killing one person can be moral if it minimises the overall harm. Thus, philosophers consider moral decisions endorsing either the deontological or the utilitarian principle.

Moral dilemmas' philosophical tradition has led to confusions in decision-making research. Indeed, ambiguities between the philosophical and the psychological level of analysis arise when considering philosophical commitments as reasons for dilemma responses. Conway (2022) illustrates such a confusion by highlighting the role of antisocial personality traits in dilemma judgment. The author points out a frequently reported correlation between antisocial traits and endorsing sacrificial harm to maximise outcomes. Conway (2022) argues that such findings should neither be considered as evidence of utilitarian concerns in antisocial personality nor as evidence that dilemma responses do not entail any form of moral

concerns. Research suggests that preferences for accepting harm to maximise outcomes in antisocial personality are actually driven by low concerns for outcomes. Conway (2022) concludes that judgments are based on combination of psychological mechanisms which do not necessarily reflect the endorsed philosophical principle. Therefore, the author highlights the relevance of distinguishing philosophical theories, actual judgments and psychological mechanisms when analysing dilemma responses.

The use of moral dilemmas for investigating social decision-making in psychology has been popularised by Greene et al. (2001, 2004, 2008). The authors used moral dilemmas to investigate the relationship between decision-making, emotion (Greene et. 2001, 2004) and cognitive load (Greene et al., 2008). Greene et al. (2001) formulated the assumption emotional processing in judgement by comparing the trolley dilemma to the footbridge dilemma (Thomson, 1985). In the footbridge dilemma, the trolley headed toward the five works can be stopped by pushing a stranger off the footbridge that spans the track between the trolley and the workers. While people generally consider diverting the trolley (an killing the person on the other track) as acceptable, they usually consider pushing the stranger as immoral. On the one hand, Greene et al. (2001) consider the trolley dilemma as impersonal as the death of the person is a side effect of hitting the switch to divert the trolley. On the other hand, the footbridge dilemma is considered as personal since the death of the stranger is not a side effect but a direct 'close-up' act. They suggest that the footbridge dilemma is, therefore, more emotionally engaging than the trolley dilemma. To test this hypothesis, the author used functional magnetic resonance imagery (fMRI) to measure the activity of brain areas involved in emotional processing such as the

medial frontal gyrus, the posterior cingulate and the angular gyrus, when answer moral dilemmas similar to the trolley vs. the footbridge dilemma<sup>1</sup>. Their results suggest that the activation of brain areas involved in emotional processing increased when answering personal than impersonal dilemmas. Moreover, Greene et al. (2004) complemented the previous findings by showing greater activation of brain areas associated with abstract reasoning and cognitive control (i.e., the dorsolateral prefrontal cortex and the anterior cingulate cortex) and longer response times when resolving personal dilemmas in a fashion consistent with utilitarian response. Finally, Greene et al. (2008) investigated the role of controlled cognitive process in utilitarian responding. In their experiment, participants responded to parts of moral dilemmas under cognitive load by completing a concurrent digit search task. Their results showed increased response times when making utilitarian judgements under cognitive load. This increase was not observed in non-utilitarian judgements. Thus, the authors suggested that answering dilemmas under cognitive load interfered with utilitarian judgement.

Through the research mentioned above, Greene et al. (2001, 2004, 2008) have developed the dual-process model of decision-making. The model suggests that utilitarian judgments result from a deeper level of cognitive processing while deontological judgments involve more automatic emotional reactions. Moore et al. (2008) tested the dual-process theory by investigating the impact of individual

---

<sup>1</sup> Greene et al. (2001) also used behavioural measures such as response time to inform on the automatic process involve in answering the dilemma. However, the results from the response time analyses were retracted in Greene et al. (2008) as they were due to a methodological glitch in the formulation of some moral dilemmas.

differences in control on moral judgement. Moreover, the author provided a modified version of Greene et al.'s (2001) dilemma set to match stories for their length and avoided repeated exposure to the same scenario. They controlled for individual differences in executive control using Working-Memory-Capacity (WMC) tasks. The WMC tasks involved three measures of complex-span: operation span, reading span, and symmetry span. In line with Greene et al. (2001, 2004), the authors included measures of response times. Their results showed that WMC scores predicted moral judgements in personal moral dilemmas especially when the victim's death was inevitable. They also found that higher WMC predicted longer response times when endorsing utilitarian judgments and higher consistency when answering personal dilemmas. Interestingly, their improved version of the dilemma set led to shorter response times when answering personal dilemmas compared to impersonal dilemmas. Overall, their findings highlight the importance of distinguishing between personal and impersonal moral dilemmas (Moore et al., 2008).

The dual-process theory (Greene et al., 2001, 2004) has been the most popular model of moral decision-making for over a decade. However, part of this model has been updated since then. The original version of the dual-process theory (Greene et al., 2001, 2004) considers the emotional processing as immediate, automatic and leading to a deontological decision. The model also suggests that the immediate emotional response can be overcome with a slower cognitive deliberation allowing the endorsement of the utilitarian decision. This temporal dynamic between emotional and cognitive processes is also known as "default-interventionist" (Evans, 2008). However, more recent evidence supports a concurrent rather than sequential activation of

deliberative and emotional systems in moral decision-making. Koop (2013) tested the default-interventionist hypothesis using mouse tracking, also called response dynamics of moral dilemma responses. Following the default interventionist principle, the author hypothesised that utilitarian responses would show initiative mouse movement toward the deontological response option before reversing trajectories to select the utilitarian option. A direct mouse trajectory was assumed for deontological responses. The results failed to support the mouse trajectory reversal expected in utilitarian responses. Interestingly, more reversals were observed before providing deontological responses as compared to utilitarian responses. Further evidence against the default-interventionist hypothesis was provided in Baron et al., (2012). Using moral dilemma response time, the authors showed that longer response times were not associated with the dilemma response itself but with the difficulty of the dilemma. Overall, contemporary evidence supports concomitant emotional and deliberative processing in moral decision-making.

Considering the lack of support for the temporal aspect of the dual-process model, Conway (2022) proposed a “revised soft dual-process”. According to the soft version of the dual-process model, moral decision-making is driven by parallel activations of the emotional aversion to the thought of killing someone and the logical deliberation around the overall outcomes. Following this model, both the emotional and deliberative systems vary independently, and may concomitantly increase. Nonetheless, Conway (2022) indicated that the soft dual process model is incomplete. Emotional and deliberative processing are only two of the many aspects influencing

responses in dilemma tasks. To conclude, the author invited researchers to differentiate dilemma responses from the processes involved.

## 1.2 Decision-making and episodic memory

*“The relationship between memory and decision-making is circular and iterative: we learn about what we choose to experience, whether directly or vicariously, and our memory representations are adjusted through repeated experience.”*

Fellows, 2018

Tversky and Kahneman’s (1973) availability heuristic has popularised the idea that decision-making and memory are intrinsically related. The availability heuristic states that the accessibility of instances retrieved from memory influences judgments and decisions. If the availability heuristic has generated a plethora of research, Schwarz and Vaughn (2002) raised the issue of ambiguity around the underlying processes and proposed a revisited<sup>2</sup> version of this heuristic. The authors highlighted that most contemporary evidence of the availability heuristic was based on recall tasks. However, recall tasks do not only render the experienced ease of recall as a source of information but they also yield recalled content. The authors showed that confounding the two sources of information can lead to misinterpretation of the underlying processes. Their findings suggest that the ease of recall is used as a source of information for decisions only when the value of such an experience is not questioned. When the value of recall experiences is contested, the recall content is

---

<sup>2</sup> It is the authors’ intention to refer to their version of the availability heuristic as ‘revisited’ rather than ‘revised’.

used as a substitute. Furthermore, the ease of recall experiences is preferred in judgement tasks with low personal relevance. When the personal relevance of the task is high, individuals will use recall content as a source of information. Overall, Schwarz and Vaughn's (2002) revisited availability heuristic considers two sources of information for judgment and decision: the ease of recall and the content of recall.

The issue raised by Schwarz and Vaughn (2002) about the nature of the processes underpinning the availability heuristic was further considered by Stewart et al. (2006). Focusing on the content of recall as source of information, Stewart et al. (2006) started from the assumption that memory content reflects the structure of the individual's environment, or the real world. When making a judgment, a sample or related items is drawn from memory and the targeted item is compared to the memory sample. Accordingly, the authors labelled their theory as *decision by sampling*.

In contrast to the work conducted by Kahneman and Tversky in the 70s, Stewart et al.'s (2006) model does not rely on psychoeconomic scales. The authors highlighted that unlike classic theories on decision-making, they do not assume that value, probabilities and temporal duration are represented by stable, long-term internal scales. They consider that only a sample of items can be retrieved from memory and the targeted items will be judged as larger or smaller than the sample. In other words, a value judgment is based on binary or ordinal comparisons to a sample of reference. By focusing on the content of recall from sampling, the authors aimed to explain the main findings of the prospect theory (Kahneman and Tversky, 1979): the shape of the utility function, the steeper slope of losses compared to gains, the hyperbolic temporal discounting, and the respective over- and underweighting of small and large

probabilities. For instance, the author examined the distribution of credit to and from the UK's private bank accounts. They showed that the distribution of credits follows a power law with many small gains and a few large gains. They also noted that several natural phenomena follow similar distributions such as earthquakes or the number of words within a language. Thus, the authors concluded that the value function of gains is concave because it is based on a memory sample that reflects the real-world distribution of gains. Similar, the asymmetry of the distribution of losses and gains mirrors the natural distribution of gains and losses. In summary, the theory of decision by sampling assumes that judgment is made by binary or ordinal comparisons between a targeted value and a sample of values retrieved from memory that depicts a real-world distribution. In other words, because of limited working memory capacities, a small sample of similar items cued from memory is selected and compared to the targeted items. In line with what is experienced in the daily life, the samples constructed for losses are high fidelity but the samples for gains are low fidelity.

Initiated with the availability heuristic and further developed with the decision by sampling theory, the relationship between decision-making and memory has become a prosperous field of research. In 2010, Johnson and Busemeyer reviewed the literature on models of decision-making under risk and uncertainty. They highlighted the influence of "memory-based approaches" on decision-making and the empirical support that these theories have received. More recently, Redish and Mizumori (2015) have edited a special issue of the journal *Neurobiology of Learning*

*and Memory* on decision-making and memory highlighting the growing interest in the topic.

In the following sections, I review correlational and causal evidence for the relationship between decision-making and episodic memory. The focus on episodic memory was motivated by two main reasons. The first reason relates to a matter of consistency with contemporary descriptions of the relationship between decision-making and memory. Let us consider Fellow's (2018) citation presented above. This citation highlights the mutual influence between decision-making and memory of previous experiences. A similar description can be found in Tremel et al. (2018) who suggest that learning about outcomes and retrieving memories of prior decisions inform decisions in the present. This type of memory system, the memory for past experiences, refers to episodic memory as defined by Tulving (1972). Tulving (1972) defined the episodic memory system as the underpinning capacity to remember specific events. Moreover, extrapolation of classical episodic memory effects would suggest enhancement of memory retention following decision-making. In their 1972 paper, Craik and Lockhart presented their theory on levels of processing as a framework for memory research. In their paper, the authors reviewed contemporary literature on perception and the concept of "depth of processing". Following this concept, greater "depth" implies a greater degree of semantic or cognitive analysis. By reviewing existing data, the authors showed that deeper levels of processing lead to better retention. Since then, the impact of the level of processing on memory has been widely studied, and this framework is considered as a classic paradigm for memory research (for a review see Craik, 2002). Thus, by extension, if decision-

making requires in-depth processing of the decision context at (and maybe beyond) semantic level, then decision-making should enhance the memory retention of the decision situation. The second motivation for focusing on episodic memory is to fill a gap in the literature. Indeed, Murty et al. (2016) state that the relationship between decision-making and episodic memory is highly assumed but rarely explored. Schacter et al. (2015) agree with this observation and insist on the relevance of focusing on episodic memory, which could, in the authors' words, "potentially illuminate the relationship between memory and decision-making".

#### 1.2.1 Correlational evidence of a relationship between decision-making and memory

A relationship between decision-making and episodic memory has been observed at a neuroanatomical level. Studies in functional neuroimaging show that a specific set of brain regions is involved in decision-making: the thalamus, the striatum, the orbitofrontal cortex (OFC), and the ventromedial prefrontal cortex (VMPFC; for a review see Fellows, 2018). Interactions between the hippocampus and the VMPFC support memory integration and considerations of options for decision (for a review, see Schlichting & Preston, 2017). Additionally, studies in psychophysiology have shown correlations in brain activation between decision-making and episodic memory (Jacobs et al., 2006). Overall, correlations between brain areas involved in decision-making and memory support a physiological relationship between the two functions.

Research in cognitive psychology also shows a relationship between decision-making and episodic memory. In 2013, Del Missier et al. investigated the relationship between decision-making and working memory, episodic memory and semantic memory. Their experiment measured decision-making using the Adult Decision-

Making Competence battery (A-DMC; Bruine de Bruin et al., 2007). The A-DMC assess decision-making as a function of six components: resistance to framing, applying decision rules, under/overconfidence, recognising social norms, resistance to "sunk" costs (i.e., the ability to discontinue failing commitments following unrecoverable costs), and consistency in risk perceptions. The authors evaluated working memory, episodic memory and semantic memory using a battery specifically developed for a wider research project which included their experiment. Their results revealed that each A-DMC component was correlated with at least one of the three memory systems. Furthermore, a study from Schaper et al. (2019) provided further evidence of a relationship between *source memory* and social decision-making. The authors measured decision-making and source memory with the Prisoner Dilemma task. Here, the authors referred to source memory in the sense of remembering that a face was encountered in the Prisoner Dilemma task. The Prisoner Dilemma task is similar to the Ultimatum game presented in Section 1.1.3. In the Prisoner Dilemma task, participants are presented with fictitious partners. The participant can cooperate with each partner by investing money or refusing to cooperate and not investing anything. The game has two types of partners "cooperator" or "cheater." The participant earns or loses money depending on whether they cooperated with a cheater or a cooperator. In Schaper et al.'s (2019) experiment, a picture of the partner is presented during the bet. Their experiment ends with a memory task where participants are given pictures of partners to identify whether the partner is a cheater, cooperator, or new partner. Their results show a correlation between participants' performance on the decision-making task (i.e., total gain) and their performance on

the memory task (i.e., partner type correctly recognised). Better source memory was associated with more cooperation with cooperators and less cooperation with cheaters. Thus, the literature provides neuroimaging and cognitive evidence of a relationship between decision-making and episodic memory.

### 1.2.2 Causal evidence of the effect of episodic memory on decision-making

In this section, I present experimental research on the impact of decision-making on episodic memory. As a matter of consistency, I review only the studies in which the decision-making tasks follow the definition of decision-making provided in Section 1.1.1. A decision-making task meets the definition provided if it involves choosing between several options based on judgements and preferences for the option's outcomes. Therefore, studies involving visual discrimination or associative learning tasks (e.g., Giguere & Love 2013; Tremel et al., 2018) are not reported.

As mentioned earlier in this chapter, Kahneman and Tversky's (1979) seminal work on heuristics and biases has a long-lasting impact on the study of decision-making and memory. In line with this perspective, contemporary research has mainly investigated memory as a bias for decision-making. In 2014, Madan et al. studied the impact of memory bias for extreme outcomes on decision-making under risk. In their experiment, participants are presented with pictures of doors. The task involves four doors: one associated with a fixed gain (+20), one associated with a fixed loss (-20), one associated with a random gain (50% of 0 and +40) and one associated with a random loss (50% of 0 and -40). Doors are presented two-by-two, and participants must successively choose one of the two doors in a sequence of multiple trials. A surprise memory task followed, during which participants were asked to recall the

outcome associated with each door and the frequency at which they appeared. Their results suggested that participants were more risk-seeking toward random gains than random losses. Indeed, participants bet on the door associated with a random gain more frequently than the doors associated with a random loss. They also estimated the extreme outcomes as more frequently occurring. The authors concluded that memory for extreme outcomes biases decision-making under risk. The authors replicated and expanded these results in a later experiment. Madan et al. (2017) investigated whether the memory bias observed in the Madan et al. (2014) study varies according to the type of decision-making task adopted. The authors differentiated two types of tasks: decision-making from experience and decision-making from description. In decision-making from experience, participants completed the door gamble task used in the earlier experiment (Madan et al., 2014). In the description task, pictures of the doors were replaced by descriptions of outcome probabilities (e.g., "Win 20" or "Gamble, win 0, win 20"). Their results replicated the memory bias observed in decision-making by experience but not in decision-making by description. In other words, a memory bias was observed where participants had to bet on the doors but not when they bet on description of probabilities. The authors concluded that decision-making from experience is based on memory samples of previous outcomes. In decision-making from description, no memory sampling is required to estimate the expected outcomes.

Additional evidence of the impact of memory on decision-making was provided in Murty et al. (2016). In this study, participants engaged in multiple independent lottery trials. The experiment was divided into three phases: the reward phase, the decision

phase and the memory phase. In the reward phase, each lottery trial was associated with an outcome (gain or loss) and a house picture. In the decision phase, participants saw two house pictures from the lottery trials. Participants had to decide which lottery they wanted to re-engage by choosing one of the two house pictures. During the memory task, participants saw several house pictures. They had to recognise the lottery houses and recall the associated lottery outcome. The results suggest that participants re-engaged with high- than low-value lotteries more often. The memory of the lottery-outcome association predicted decisions to re-engage with lotteries. Murty et al. (2016) concluded that episodic memory influences adaptive decision-making.

In 2006, Iglesias-Parro and Gomes-Ariza offered an original approach to studying memory's impact on decision-making by using a retrieval practice paradigm. *Retrieval practice* is a learning technique that involves repeated retrievals of information from memory (Roediger & Karpicke, 2006b). The authors suggested that the inhibition of information during retrieval would reduce its accessibility and thus impact decisions where such information is required. In their experiment, participants studied six sets of fictitious candidates for a job opportunity. The authors induced retrieval practice by presenting a candidate's name along with an attribute stem. Half of the candidates' names was presented again during a retrieval phase. During the retrieval phase participants were asked to recall the attribute associated to the participant. This was followed by a decision-making phase measuring participants' preferences for good vs. poor candidates. The experiment ended with a cued recall phase of all candidates' names and attributes. Their results showed that practicing retrieval of relevant and irrelevant attributes for poor and good candidates influence

the final recall and participants' preference. Retrieval practice impairs the recall of positive attributes for good candidate. Moreover, retrieval practice led to a preference for unpractised good candidate over practiced good candidates. The authors concluded that retrieval practice biases decision-making.

A similar paradigm was used in Coman et al. (2013) offered an original approach to studying memory's impact on decision-making. The authors investigated the effect of memory on medical decision-making using a retrieval-induced forgetting paradigm. The experiment started with a learning phase during which participants learnt information about a fictitious disease and the advantages and disadvantages of four treatment options. The learning phase was followed by a retrieval phase. In the retrieval phase, they read a pamphlet presenting relevant or irrelevant advantages or disadvantages about the treatment options. Thus, the pamphlet omitted part of the information presented in the learning phase. The retrieval phase was followed by a cued recall task of treatments' advantages and disadvantages. At the end of the experiment, participants decided on the best treatment option for a fictitious friend. Their results show retrieval-induced forgetting for information omitted in the pamphlet. Moreover, the induced forgetting adversely impacted decision-making. Overall, the experiments reported above provide evidence of the impact of episodic memory on decision-making.

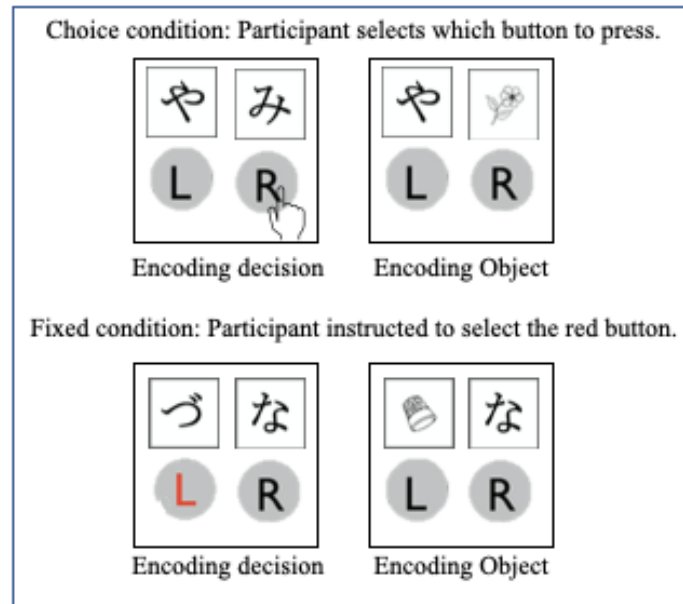
### 1.2.3 Causal evidence of the effect of on decision-making on episodic memory

In the previous section, I report results from experiments examining the influence of decision-making on memory using decision-making tasks that follow the definition of decision-making considered in this thesis (see Section 1.1.1). This was

done to distinguish research specific to the thesis's focus amongst the variety of evidence on the impact of memory on decision-making. However, this selection based on decision-making tasks could not be applied to experiments on the impact of decision-making on memory. Indeed, research on the relationship between the two processes seems to take a unidirectional focus and research on the impact of decision-making on episodic memory is scarce. To my knowledge, only one article from Murty et al. (2019) explicitly addresses this question. Murty et al. (2019) used a forgetting paradigm to investigate the role of decision-making on post encoding consolidation. In this experiment, participants saw two occluder screens illustrated with Japanese syllabaries - Hiraganas. Trials were divided into "choice" and a "fixed" condition. In the choice condition, participants chose one to remove one of the two occluder screens. In the fixed condition, the experimenter indicated which occluder screen should be removed. The removal of the occluder screen revealed a picture of an object that the participants had to memorise. Unbeknownst to the participant, the same picture was hidden behind both occluders. Hence, the choice task had no impact on the object to be encoded. An illustration of the task is presented in Figure 1. The experiment ended with an object recognition task at one-day delay. Murty et al.'s (2019) results suggest increased forgetting of the objects in the fixed compared to the choice condition. They concluded that decision-making impacts memory through post encoding consolidation.

## Figure 1

*Presentation of The Choice Encoding Task from Murty et al. (2019)*



*Note.* This figure illustrates Murty et al.'s 2019 choice encoding task. Participants have to press a button under the targeted occluder screen to remove it and display the object to be encoded. In the choice condition, participants were free to choose which of the two occluder screens to remove. In the control condition, the experimenter indicates which occluder screen must be removed. Adapted from "Decision-making increases episodic memory via postencoding consolidation", by V. P. Murty, S. DuBrow and L. Davachi, 2019, *Journal of Cognitive Neuroscience*, 31(9), 1308-1317.

Murty et al. (2019) is, to my knowledge, the first study to experimentally focus on the impact of decision-making on memory rather than the impact of memory on decision-making. However, their decision-making task fails to meet the definition of decision-making considered in the current thesis i.e., decision-making involves three steps: judgement, preference and choice. Murty et al.'s (2019) task does not involve

any judgment of the outcome related to the decision options. Their task was designed so that a participant's choice would have no impact on the object encoded. Thus, under the definition of decision-making adopted in this thesis, Murty et al. (2019) provide evidence of an effect of choice on memory rather than decision-making. Therefore, the impact of decision-making on episodic memory has not yet been examined in the literature.

Given the scarcity of evidence of the impact of decision-making on memory, research from related literature on reasoning and memory was considered. A causal effect of reasoning on memory representation can be found in the analogical reasoning literature. Analogical reasoning refers to the use of a well-known source domain to draw inferences on a comparable but less known target domain (Spellman & Holyoak, 1996). The process of identifying systematic correspondences between the source and the target is labelled as mapping. In a set of three experiments, Spellman & Holyoak (1996) investigates the impact of goal-driven analogy on analogical mapping. In their experiments, participants were asked to generate a mapping between analogues with ambiguous elements. Experiment 1 and 2 showed a production of many-to-one mappings when goal-irrelevant information was provided. Moreover, Experiment 3 showed that goal-irrelevant information impacted the resolution of ambiguous mapping elements on the basis of goal-relevant information. Thus, this study provided strong evidence that, for the same analogy, different goals generate different types of mapping. Blanchette and Dunbar (2002) further investigated the impact of analogical mapping. The authors tested the impact of source mapping on the memory for target information. In their experiments, participants were presented

with a description of the source analogue vs. irrelevant information (Experiment 1 and 2) or no information at all (Experiment 3). In all three experiments, the results suggest that false analogical inferences were retrieved for memory, even when a description of the source analogue was provided. Moreover, participants who read the description of the source were not able to differentiate between their inferences and the information that was actually presented in the description. Overall, studies on analogical reasoning show that different reasonings lead to different memory representations. By extension, the literature on analogical reasoning shows that high-level cognitive functions can impact memory.

#### 1.2.4 The study of episodic memory with forgetting and learning paradigms.

In a book chapter from 2018, Fellows reviewed *The Neuroscience of Human Decision-Making Through the Lens of Learning and Memory*. This review highlights the relevance of investigating the role of memory and learning in decision-making research. According to the author, decision-making creates a link between the memory of the past and future actions. To cite Fellows (2018): "decision-making is, by its very nature, future-oriented". This consideration of time and temporality is consensual in the literature on decision-making. Thus, investigating decision-making and episodic memory with forgetting and learning paradigms, as done in Iglesias-Parro and Gomez-Ariza (2006), Coman et al. (2013) and Murty et al. (2019), is especially relevant. In this section, I briefly present influential literature on forgetting and learning.

Empirical research on forgetting can be traced back to Ebbinghaus's seminal work (1885/1964). Ebbinghaus investigated forgetting on himself by memorising lists of

nonsense syllables and trying to learn the list again at multiple delay intervals. Ebbinghaus measured forgetting as a function of trials required to re-learn the entire list of syllables. Contemporary research on forgetting focusing on the loss of information over time follows Ebbinghaus's procedure (Roediger et al., 2010). Typical forgetting experiments involve comparing the retention of different groups of participants across different time intervals from immediately after learning information to several hours, days, weeks later. A forgetting curve is derived from plotting the retention (i.e., the amount of information retained) at each delay interval (Roediger et al., 2010). Rubin and Wenzel (1996) highlight that such a plot is actually a retention curve instead of a forgetting curve. However, the literature uses the two terms interchangeably. Research on forgetting curves suggests that forgetting is not linear but showed a rapid initial decrease and then tend to slow down to approach an asymptote (Baddeley, 1990; Della Sala, 2010).

Research on forgetting and learning are intrinsically related. The main focus of research on learning enhancement is the attenuation of long-term forgetting (Pashler et al., 2007). Techniques for learning enhancement include restudying, retrieval practice and spacing. The concept of restudying comes from research in educational psychology. It refers to the repeated reading of the material to be learned and is considered students' favourite learning technique (McDaniel et al., 2009). Despite its success amongst students, restudying is not the most effective technique for enhancing long-term retention. Research has shown that retrieval practice increases long-term retention compared to restudying (for a review, see Roediger & Karpicke, 2006b). Retrieval practice or the testing effect commonly refers to the benefit of

repeated retrievals of information from memory on retention (Roediger & Karpicke, 2006b). Although both concepts tend to be used interchangeably, retrieval practice specifically refers to the act of calling information to mind, and the testing effect is the benefit of retrieval practice on learning over re-reading or no exposure at all (Roediger & Butler, 2011). Retrieval practice can be paired with spacing to improve learning (Sodertstorm et al., 2016). The spacing or distributed practice effect refers to the benefit of separating study intervals on long-term retention over mass studying (for a review, see Cepeda et al., 2006). To conclude, using forgetting and learning research paradigms offer a suitable framework for investigating long-term episodic memory retention.

### **1.3 Overview of the present research**

This thesis investigates the relationship between decision-making and memory. More specifically, it aims to provide evidence of the impact of decision-making on episodic memory. In Chapter 2, Experiment 1 presents the development of an original experimental paradigm to test the retention of a standardised piece of prose under forgetting and when promoting learning. Experiment 2 compares the retention of the standardised prose piece with a moral dilemma story. Experiment 3 examines the frequency that different moral decision types are made from a set of dilemma stories to identify the moral dilemmas that lead to an even distribution of decision types (i.e., utilitarian vs. deontological). Chapter 3 uses the paradigm developed in Chapter 2 to study the impact of decision-making on episodic memory. In Experiment 4, I explore the effect of decision-making and moral decision types on memory retention after one day. Experiment 5 further investigates the influence of moral decision types on

retention for up to a one-month delay. In Chapter 4, Experiment 6 offers a partial replication of Experiment 5, and investigates the role for learning in the relationship between decision-making and memory.

## **Chapter 2: Development of a new experimental paradigm for investigating decision-making and memory.**

### **2.1 Introduction**

Decision-making and memory are thought to be inherently related (Fellows, 2018). Research has investigated this relationship at both neuroanatomical and cognitive levels. At the neuroanatomical level, evidence shows that the hippocampus and the medial prefrontal cortex interact to promote memory integration and decision-making (see Euston et al., 2012; Schlichting & Preston, 2017, for reviews). From a cognitive perspective, theories consider that decision-making is based on memory recollection, and the decision made leads to new mnemonic representations (Tremel, et al., 2018; Fellows, 2018). Generally, the relationship between decision-making and memory is considered to be reciprocal – both processes impact one another. The current chapter focuses on developing a research paradigm to investigate this relationship further. Ideally, the research paradigm should meet three requirements. Firstly, it should enable complex decision-making manipulation with distinct options for decision and a control group. Secondly, it should include prose content which will allow measuring the memory retention of the decision-making environment. Finally, it should offer a baseline performance that can be described under the retrieval practice paradigm.

The reciprocal relationship between decision-making and memory is theoretically assumed. However, empirical evidence mainly adopts a unidirectional perspective – focusing on how memory impacts decision-making. For instance, Madan and colleagues (Madan et al., 2014; Madan et al., 2017) investigated the relationship

between episodic memory and risky decisions. In Madan et al. (2014), participants were presented with pictures of four doors. Each door was associated with either a fixed gain (+20), a fixed loss (-20), a random gain (0 or +40) or a random loss (0, -40). Doors were presented two-by-two, and participants had to bet by selecting a door over multiple trials. In a surprise memory task, participants recalled the outcome associated with each door and estimated their frequency. The results showed that a higher performance at the gambling task (i.e, gain) was associated with better memory recall. Moreover, participants overestimated the frequency of the extreme outcomes. In a latter experiment, Madan et al. (2017) showed that this memory bias was observed only when decision outcomes were experienced rather than described. Overall, they concluded that memory bias for extreme outcomes influenced decision-making from experience. However, it must be noted that the decision-making task precedes the memory task in their experimental design. Therefore, their research might not provide evidence of how memory biases decision-making but rather how decision-making biases memory. Nevertheless, the authors did not consider such a possibility.

Another illustration of unidirectional investigation can be found in Coman et al. (2013). The authors focused on the impact of retrieval-induced forgetting on medical decision-making. Their experiment started with a learning phase where participants learned information about a fictitious disease and the advantages and disadvantages of four different treatments. This was followed by presenting a treatment pamphlet presenting either relevant or irrelevant information previously learnt about the medical treatments – the induced retrieval phase. This pamphlet omitted part of the information presented in the learning phase. This phase was followed by a cued recall of

treatments' information. After the cued recall task, participants decided on the best treatment option for a fictitious patient. Their results showed retrieval-induced forgetting for information that was omitted in the pamphlet. Moreover, the information presented in the pamphlet influenced decisions between treatment options – even when such information was irrelevant. The authors concluded that selective retrieval of medical information adversely impacted medical decision-making. Here, the authors did not consider whether medical decision-making may reciprocally impact the retention of information about treatment options. Such considerations are essential inasmuch as patients' recall of treatment information enhances treatment adherence (Dong et al., 2017).

Although empirical research has mainly focused on the impact of memory on decision-making, one study has taken the opposite approach – studying the impact of decision-making on memory. To my knowledge, Murty et al. (2019) is the first and only published study to focus on the effect of decision-making on memory using a forgetting paradigm. In this experiment, participants saw two occluder screens labelled with Hiragana characters. The occluder screens were presented side-by-side. Participants had to press a button to remove one of the occluder screens - the choice task. In the "choice" condition, participants chose which occluder screens to remove. In the control condition, called the "fixed" condition, the experimenter designated the occluder to remove. Once removed, the occluder screen revealed a hidden object picture. Unbeknownst to the participant, the same object picture was hidden behind both occluders. An illustration of the task is provided in Figure 1. Participants were instructed to memorise the object in the picture. After this encoding phase, participants

completed a post-encoding rating task where they rated how much they liked the Hiragana selected on a 5-point scale. The rating task was followed by a recognition task of the objects presented in the encoding task. Finally, participants completed a second recognition task after a 24-hour delay. The results showed reduced forgetting for object pictures in the choice condition over the fixed condition after 24 hours. Murty et al. (2019) concluded that decision-making enhances memory consolidation.

Although Murty et al.'s (2019) conclusion align with the literature on decision-making and memory, evidence remains limited. Indeed, their experimental elicitation of decision-making is questionable. In their experiment, Murty et al. (2019) elicited decision-making through the act of choosing between two Hiraganas. This task is somewhat arbitrary: the participants' choice has no specific outcome. As the authors explain: "there was no relationship between decisions and the underlying object image". I argue that choosing between two arbitrary options does not entail decision-making. As discussed in Chapter 1, decision-making is a process of three steps: judgement, preference and choice (Fischhoff & Broomell, 2020). Thus, making a choice is necessary but not sufficient to enact decision-making. Murty et al.'s (2019) experiment might involve preference since participants could select their preferred Hiragana. This assumption remains speculative as the authors do not report whether chosen Hiraganas were rated higher than designated ones. Nevertheless, Murty et al.'s (2019) design does not enact judgment (i.e., estimation of choices' possible outcomes) and thus cannot be considered a decision-making task. Additionally, there is an inconsistency in the designation of the authors' research paradigm. Murty et al. (2019) indicated that part of the data was previously reported in Murty et al. (2015).

Although the decision-making paradigm is the same across the two studies, Murty et al. (2015) do not label the task as such. In the earlier experiment, the authors refer to their task as a simple act of choosing that enacts a sense of agency over the learning experience. It is unclear why the same task and data are associated with different labels and research outcomes between Murty et al. (2015) and Murty et al. (2019). In sum, I suggest that the task presented above elicit an act of choice rather than decision-making.

Overall, research on the impact of decision-making on memory is scarce and methodologically limited. The lack of empirical evidence is likely due to the challenge of measuring the impact of decision-making on memory. Indeed, most available tasks allow measuring decision-making but not manipulating it experimentally. The current chapter addresses this challenge by developing a new experimental paradigm. To design this new paradigm, I have taken the forgetting approach presented in Murty et al. (2019). However, I have designed a decision-making task in line with the definition given in the general introduction. Experiment 1 focused on the retention of a standardised piece of prose under long-term forgetting and retrieval practice. Experiment 2 compares the retention of the standardised piece of prose with a moral dilemma story. Finally, Experiment 3 focuses on the percentage of different decision-making types provided for a set of popular dilemma stories.

## **2.2 Experiment 1: Investigating the Retention of a Piece of Prose Under Long-Term Forgetting and Retrieval Practice**

Since Ebbinghaus's (1885/1964) seminal work on memory and forgetting, a long-lasting interest has been addressed in developing mnemonic techniques. Amongst

these techniques, retrieval practice is one of the most effective strategies for enhancing retention (Racsmany et al., 2018). Retrieval practice, or the *testing effect*, commonly refers to the benefit of repeated retrievals of information from memory over restudying (Roediger & Karpicke, 2006b). Although both concepts tend to be used interchangeably, retrieval practice specifically refers to the act of calling information to mind, and the testing effect is the benefit of retrieval practice on learning over re-reading or no exposure at all (Roediger & Butler, 2011). Roediger and Karpicke (2006a) marked a turning point in the field by highlighting the benefit of retrieval practice under educationally relevant conditions. The authors suggested using prose materials and free recall without feedback to approximate the conditions of essay tests in education. In their experiment, participants restudied or practised retrieval on two short narratives about scientific topics (within-subject design). In the restudy condition, participants read the text over again. In the test condition, participants wrote down as much information they could recall. Their results showed that a single immediate test enhances long-term retention compared to restudying, even when no feedback is provided. They also found that repeated tests increased long-term retention over repeated restudying. However, the opposite effect was found when measuring short-term retention. Since then, the benefit of retrieval practice has been demonstrated across a range of educational materials in both real-world and laboratory settings (Horner, 2019).

Because of its specific relevance to the field, retrieval practice has been intensively studied in educational psychology (Roediger et al., 2011). Therefore, most studies compare the effect of retrieval practice to students' preferred learning strategy

- restudying (McDaniel et al., 2009). However, the current chapter aims to design a research paradigm for studying the impact of decision-making on episodic memory through long-term forgetting and learning. Thus, contrasting retrieval practice with restudying is not relevant to my thesis. Here, retrieval practice is used as a learning technique for enhancing long-term forgetting. Thus, the current experiments investigate the long-term forgetting of a standardised piece of prose with and without retrieval practice. Similar research is reported in Agarwal et al. (2008). In a set of two experiments, they examined the memory recall at a one-week delay following retrieval practice with no feedback vs. no retrieval practice (control condition). The results showed that the proportion of recall after a one-week delay was about three times higher under retrieval practice without feedback than under the control condition. More generally, research on the benefit of retrieval practice on forgetting over time shows that repeated retrieval “dramatically alters the shape of the forgetting curve” (Karpicke, 2017). The *forgetting curve* is characterised by a rapid initial decrease in retention that tends to slow down to approach an asymptote (Baddeley, 1997; Della Sala, 2010). Retrieval practice mitigates this initial decrease, and retention is sustained at a relatively close rate to initial learning (see Figure 2 for an illustration).

Although the benefit of retrieval practice on retention seems to be robust (Carpenter et al., 2008), there is a debate on whether the strength of the effect varies depending on the structure or complexity of the materials. This matter is especially relevant to the current research as it focuses on a piece of prose, a form of integrated materials with a certain level of complexity. Van Gog and Sweller (2015) suggest that most studies on retrieval practice used low *element interactivity* materials. They define

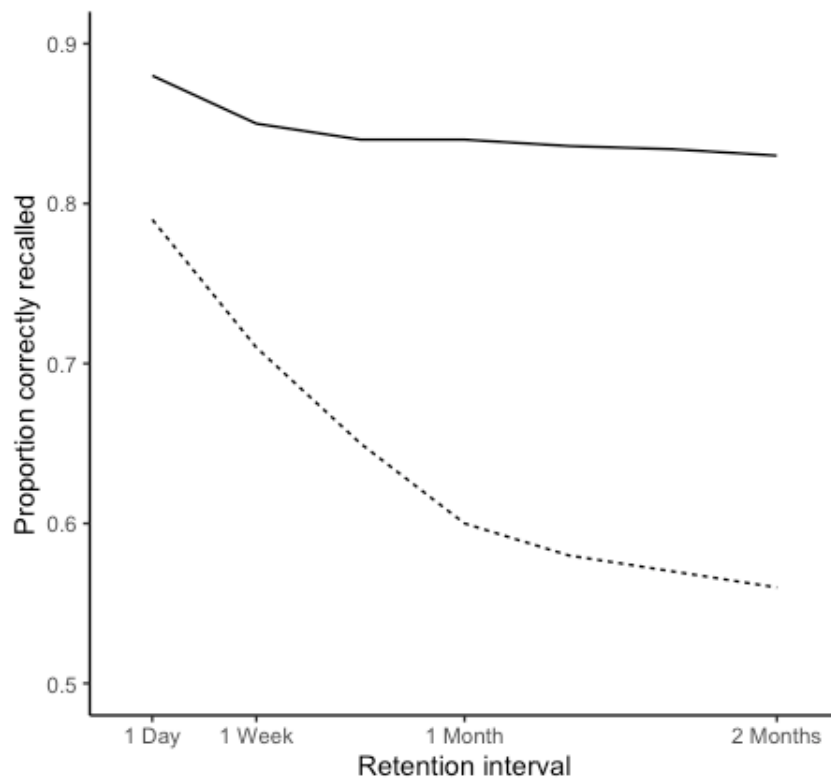
elements interactivity as the degree to which information elements are interrelated and must be processed simultaneously in working memory. The authors reviewed a collection of studies on retrieval practice and subjectively estimated whether the materials used were low or high in element interactivity. Materials assessed as high in element interactivity included examples of electrical circuits or text scenarios with deductive inferences. Materials assessed as low in element interactivity mainly entailed pairs of words or/and images. Van Gog and Sweller (2015) concluded that the complexity of the learning materials constitutes a limitation of the testing effect.

However, the conclusion above has been strongly criticised by Karpicke and Aue (2015). They argued that Van Gog and Sweller's (2015) subjective assessment of element interactivity was inconsistently applied across the reviewed studies. They also denounced the omission of studies that experimentally manipulated the complexity of the materials. Furthermore, Karpicke and Aue (2015) suggest that failures to observe a testing effect in the experiments cited were due to the type of retrieval task rather than the complexity of the materials. They report an experiment from Hinze and Wiley (2011) showing that using fill-in-the-blank rather than free recall tasks cancel the benefit of retrieval practice over restudying. Secondly, part of the experiments presented in Van Gog and Sweller (2015) use immediate massed retrieval practice. However, the literature has clearly established that retrieval practice is beneficial when some delay (e.g., days or weeks) is introduced between initial learning and retrieval tasks. Overall, evidence suggests that the recall of complex materials, such as a piece of prose, should benefit from retrieval practice as long as free recall tasks and spaced delay intervals are used.

For investigating the forgetting of a piece of prose under retrieval practice, I suggest using *Story C* from the logical memory subtest of the Wechsler Memory Scale (WMS-IV; Wechsler, 2009) as study materials. The use of Story C presents three advantages. To begin, the Logical Memory subtest measures verbal episodic memory using short narratives – i.e., pieces of prose. In addition, the WMS is one of the most popular standardised tools for assessing memory (Kent, 2013), especially the Logical Memory subtest (Morris et al., 2014). Therefore, I am familiar with the test and its administration. Finally, the task administration and scoring of the story recall are standardised, which reduces ad-lib decisions in developing the thesis materials. Overall, Story C offers a good starting point for exploring the forgetting and retrieval practice of a piece of prose.

**Figure 2**

*Proportion of retention under retrieval practice vs. normal forgetting, from Karpicke (2007).*



**Note.** This figure represents data from Hanawalt (1937) reported in Karpicke (2017). In this experiment, participants studied a set of drawings that they reprocessed from memory either immediately or after one day, one week, one month or two months. The straight line represents retention with repeated recall – retrieval practice. The dotted line represents retention with a single recall – forgetting. From “Retrieval-Based Learning: A Decade of Progress”, J.D. Karpicke, 2017, *Learning and Memory: A Comprehensive Reference*, (J.H. Byrne, Ed) 2nd edition, Volume 2, p.489 (<http://dx.doi.org/10.1016/B978-0-12-809324-5.21055-9>)

Studies on forgetting and retrieval practice usually involve a single within-subject design - half of the materials is used to practice retrieval, and the other half

measures forgetting. Such a design is appropriate when using lists of words or multiple narratives. However, the current research relies on only one piece of prose that cannot be randomly divided into two parts. Thus, the current research has two experiments. Experiment 1a investigates the forgetting of Story C with between-subject designs. I hypothesise that there will be an overall decrease in the retention of Story C over time. The retention should differ across each delay interval; that is, the retention at one day will be different from retention at one week and one month. Experiment 1b explores the retention of Story C under retrieval practice with a within-subject design. I hypothesise that there will be a decrease in retention over time (i.e., a main effect of delay intervals). However, the benefit of retrieval practice should be characterised by little or no difference in retention across each delay (i.e., no differences expected with pairwise comparisons). In other words, retention is expected to follow an overall negative trend but no clear inflexion point.

### 2.2.1 Experiment 1a: The Forgetting of a Piece of Prose.

#### **Method**

##### ***Sample size estimation***

A power analysis was run to estimate the required sample size for a given significance level of 0.05, a power of 0.80 and three experimental groups (one-day, one-week, and one-month delay interval). The value of  $f = 0.40$  for the predicted effect size was based on Rubin and Wenzel's (1996) literature review and meta-analysis of publications on forgetting curves. The authors report average  $r^2$  effect sizes depending

on the type of memory task, and the function used to describe the relationship between time and retention. When retention is measured with recall tasks and fitted to a linear function, the average effect size is  $r^2 = .71$ , corresponding to a large effect size according to Cohen's (1992) convention<sup>3</sup>. Thus, the expected  $f$  value was set at  $f = 0.40$ , corresponding to the minimum for large effect sizes. The power analysis estimates  $n = 21.10$ . Thus, participant recruitment aimed to reach at least 22 participants per group.

### ***Participants***

91 participants (31 men, 60 women) aged from 17 to 63 years ( $M_{\text{age}} = 24.60$ ,  $SD_{\text{age}} = 7.38$ ) were recruited among the students and staff of the University of Edinburgh via the University internal recruitment system, e-mails and word-to-mouth. No participant reported any cognitive or neurological condition. All participants provided written consent to take part in the study. The study received ethical approval (No 66-1819/1) from the PPLS ethics committee of the University of Edinburgh. Two fourth-year undergraduate students assisted me in participant recruitment and data collection.

### ***Materials***

#### Digit Span Tasks

The administration of the Digit Span-Forward and Digit Span-Backward subtests followed the standardised procedure of the Wechsler Intelligence Scale for

---

<sup>3</sup> According to Cohen (1992), small, medium and large effect sizes are respectively superior to 0.01, 0.09 and 0.25 for  $r^2$ s, and 0.10, 0.25 and 0.40 for  $f$ s.

Adults – IV (WAIS-IV, Wechsler, 2008). In both Digit Span-Forward and Digit Span-Backward, there are two sequences of digits per item. The experimenters read each sequence verbatim at the rate of one digit per second, dropping the voice slightly down on the last digit of the sequence. After being read a sequence, the participant must recall all digits in the same order for Digit Span-Forward and reverse order for the Digit Span-Backward. Each subtest is composed of eight items, resulting in sixteen sequences. In the Digit Span-Forward, the sequences go from two (item 1) to nine digits (item 8), increasing from one digit every item. In the Digit Span-Backward, the sequences start with two digits for items 1 and 2, then increase from one to seven digits (item 8). Each sequence is scored 0 if failed and 1 if correctly recalled. A subtest ends when both sequences from the same item are failed. The raw scores correspond to number of sequences correctly recalled (out of sixteen) for each subtest.

The use of the digit span task was twofold. Firstly, as working memory is known to impact memory retrieval and learning (Cubelli, 2010), digit span was administered to verify that experimental groups did not significantly differ in working memory abilities. Secondly, digit span was presented to participants as the main task of the experiment to avoid intentional rehearsal of the story between the delays.

#### Logical Memory Task

The administration of Story C from the Logical Memory subtest of the Wechsler Memory Scale -- fourth edition (WMS-IV; Wechsler, 2008) mainly followed the standardised procedure of the WMS-IV. Instructions were slightly adapted as only Story C was administered, and the recognition task was not included. The recognition task was not included in the procedure to prevent re-exposure to the study materials.

Instructions were also adapted as the experimenter did not read the story. To ensure regularity in reading speed, appropriate pronunciation, and no variation in the speech between the experimenters, the reading of Story C was pre-recorded. Story C was read by the Text-to-speech software, The Scottish Voice, provided by the University of Edinburgh (The Scottish Voice, 2011). I used the "Cerevoice Stuart" version 4.0.1, a Scottish male voice. The task started with the following instructions: "You are going to listen to a short story. Listen carefully and try to remember as much as you can". Then, the recording of Story C was displayed. Following the recording, the experimenter asked: "Can you tell me everything you remember about the story? Please recall everything you can, even if you are not sure." For delayed recall, participants were asked: "Can you tell me everything you remember about the story you heard yesterday/ last week/ last month? Please recall everything you can, even if you are not sure." No time limit was set for the participant to answer. The recall task ended when the participants informed the experimenter that they could not recall more.

Story C is divided into 25 "ideas" to be recalled. Each idea is scored 1 if correctly recalled and 0 if not recalled. The raw score corresponds to the total number of ideas correctly recalled out of 25.

### ***Design***

A between-subjects experimental design was used. The between-subjects independent variable was the time interval for delayed recalls. Since forgetting is a curvilinear process (Rubin & Wenzel, 1990), three delay intervals were used for a more thorough picture of retention over time. The intervals were selected to increase in duration (i.e., spaced) from initial learning and to match with typical calendar units:

one day, one week, one month. The dependent variable was the total number of ideas from Story C correctly recalled out of 25. The design also included three additional variables to control potential biases: Story C immediate recall, Digit Span-Forward and Digit Span-Backward scores.

### ***Procedure***

The experiment was divided into two sessions – initial learning and delayed recall. The study was first presented to the participant as an experiment on working memory with the Digit Span as the main task to prevent intentional rehearsal of the story between the two sessions. The Logical Memory task was presented as a distractor task.

The first session started with the Digit Span-Forward and the Digit Span-Backward. They were followed by the Logical Memory task. The first session ended with collecting socio-demographic information and planning the follow-up sessions. Participants were then randomly allocated to one of the three conditions: one-day delayed recall, one-week delayed recall or one-month delayed recall.

### **COVID-19 Notice**

The first session initially took place in a laboratory room of the Psychology Department. Tasks were administered in a paper-and-pencil format. Follow-up sessions were conducted via phone calls.

A second round of recruitment had to be carried out during the COVID-19 outbreak to reach the required sample size (i.e., at least twenty-two participants per group). Thirty-three participants were recruited in this second round of data collection.

Since all on-campus and in-person testing was interrupted, the setting of the first session had to be adapted. All precautions were taken, so online adaptation of the first session was as similar as possible to the on-campus sessions. Instead of meeting in person, participants met with the experimenter during an online meeting on Microsoft Teams or Facebook Messenger. All tasks were computerised to set up an online version of the experiment. Participants were indicated that the experiment started once they had followed the URL provided by the experimenter. In the online version of the procedure, the instruction was not read by the experimenter but written and displayed on participants' computer screens. Sequences of the digit spans were not read by the experimenter but read by the text-to-speech software used to record the reading of Story C instead. These adjustments were made so the administration of the tasks would not be disrupted by potential lags or connection issues between the experimenter and the participant.

### **Statistical analyses**

The analysis script and a package containing the anonymised dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdbf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdbf7ec8b03ea1ad66).

Following the literature on forgetting, hypotheses were tested using analyses of variance – one-way ANOVA (Carpenter et al., 2008). For cases where ANOVA's assumptions were not met, Kruskal-Wallis' rank-sum tests nonparametric equivalent were used. Firstly, I verified that the three delay groups did not differ in Story C immediate recall, Digit Span-Forward score, and Digit Span Backward score. Then, I tested the impact of delay intervals on the retention of Story C. T-tests with Bonferroni correction were used for pairwise comparisons. Finally, an exploratory analysis was

run to verify the potential impact of the experimenters (undergraduate students vs. PhD student) and the switch from in-person, paper-and-pencil to online computerised administration. It must be noted that the experimenters and administration conditions were matched. In other words, the undergraduate students used the paper-and-pencil version of the tasks while I administrated the computerised versions. To test the impact of the variations mentioned above on retention, I added an interaction term between the experimenter variable and the delay variable to the ANOVA model. Statistical analyses were performed in RStudio, R version 4.1.0 (2021-05-18). An R Markdown document reporting the R script and R outputs of the statistical analysis is available in Appendix 1.

## Results

### *Story C immediate recall*

#### Descriptive statistics

Descriptive statistics of Story C immediate recall for each delay group can be found in Table 1.

**Table 1**

*Immediate Recall of The Story C for Each Delay Group*

Delay	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Day	29	13.14	4.86	2.00	14.00	20.00
Week	35	11.23	3.98	4.00	12.00	18.00

Delay	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Month	27	12.15	4.74	0.00	13.00	18.00

*Note.* The Story C recall is scored over 25 ideas. *N*, *M*, *SD*, *Min*, *Median*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

#### Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(2,88)=0.10$   $p=.91$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W=0.94$ ,  $p<.001$ . Thus, assumptions of the one-way ANOVA were not met. The Kruskal-Wallis rank-sum test revealed no significant difference in Story C immediate recall between the delay groups,  $\chi^2(2)=3.57$ ,  $p=.17$ .

### ***Digit Span-Forward***

#### Descriptive statistics

Descriptive statistics of the digit span forward scores for each delay group can be found in Table 2.

**Table 2**

*Digit Span-Forward Scores for Each Delay Group*

Delay	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Day	29	10.14	2.89	4.00	10.00	16.00

Delay	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Week	35	10.60	1.74	6.00	10.00	15.00
Month	27	10.59	1.78	8.00	11.00	16.00

*Note.* The Digit Span-Forward is scored over 16 sequences of digits. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

#### Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test showed homoscedasticity,  $F(2, 88) = 2.88$ ,  $p = .06$ . The Shapiro-Wilk's normality test revealed that residuals are not normally distributed,  $W = 0.10$ ,  $p = .007$ . Thus, assumptions of the one-way ANOVA were not met. The Kruskal-Wallis rank sum test suggested no significant differences in Digit Span-Forward scores between the delay groups,  $\chi^2(2) = 1.12$ ,  $p = .57$ .

### ***Digit Span-Backward***

#### Descriptive statistics

Descriptive statistics of the digit span forward scores for each delay group can be found in Table 3.

#### **Table 3**

*Digit Span Forward Scores for Each Delay Group*

Delay	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Day	29	9.69	2.30	5.00	9.00	15.00
Week	35	9.26	2.06	6.00	9.00	14.00
Month	27	9.30	2.02	6.00	9.00	15.00

*Note.* The Digit Span-Backward is scored over 16 sequences of digits. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

#### Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test revealed homoscedasticity,  $F(2, 88) = 0.21$ ,  $p = .81$ . The Shapiro-Wilk's normality test suggested that residuals are not normally distributed,  $W = 0.95$ ,  $p = .001$ . Thus, the assumptions for the one-way ANOVA were not met. The Kruskal-Wallis rank sum test showed no significant differences in backward digit span scores between the delay groups,  $\chi^2(2) = 1.11$ ,  $p = 0.57$ .

### ***Effect of delay on the retention of story C***

#### Descriptive statistics

Descriptive statistics for the number of ideas correctly recalled for each delay group is available in Table 4.

#### **Table 4**

*Number of Ideas of Story C Correctly Recalled Per Delay Group*

Delay	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Day	29	10.59	4.59	0.00	12.00	18.00
Week	35	6.49	3.33	0.00	7.00	12.00
Month	27	6.00	3.41	0.00	6.00	13.00

*Note.* The Story C recall is scored over 25 ideas. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

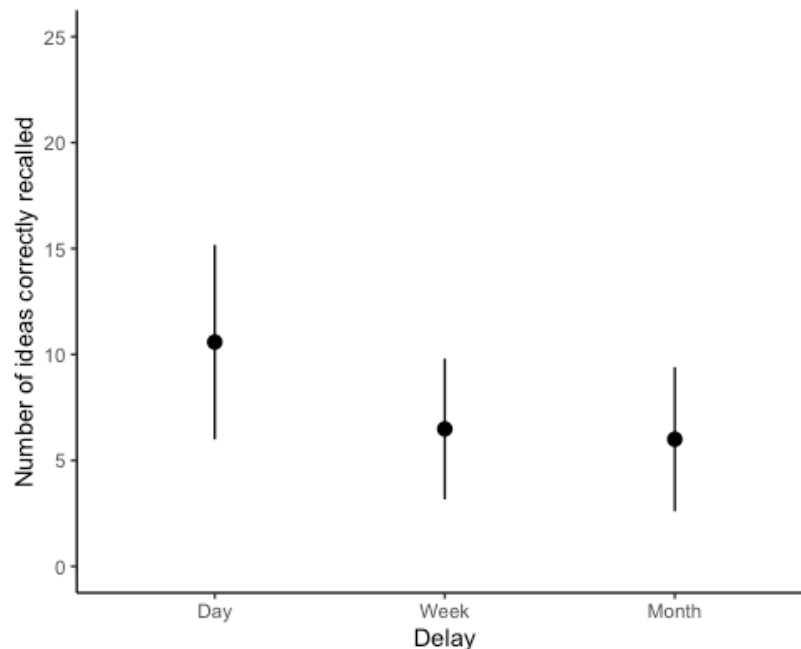
#### Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested homoscedasticity,  $F(2, 88) = 1.45$   $p = .24$ . The Shapiro-Wilk's normality test showed that residuals are normally distributed,  $W = 0.97$ ,  $p = .05$ . Thus, the assumptions of the one-way ANOVA were met.

The ANOVA revealed a significant main effect of delay groups on the retention of Story C,  $F(2, 88) = 12.86$ ,  $p < 0.001$ ,  $f = 0.54$ . The results are illustrated in Figure 3. Pairwise comparisons suggested a significant difference in retention between the one day and one week delay group  $p < .001$ . A significant effect was also found between one day and one month delay groups  $p < 0.001$ . However, no significant difference in retention was found between the one week and one month delay intervals,  $p = 1$ .

**Figure 3**

*Number of ideas correctly recalled for each delay interval*



*Note:* The dots represent the means, and the straight lines represent the standard deviations.

***Exploratory analysis: effect of administration conditions on Story C retention***

To verify whether having different experimenters and different administration conditions had an impact on the effect of delay on retention, a parameter with an interaction term was added to the ANOVA model of the previous analysis. Here, I would remind the reader that the experimenters and administration conditions were matched – the undergraduate students used the paper-and-pencil version of the experiment, while I used the computerised version of the tasks. Thus, only one interaction parameter was added to the ANOVA model.

## Descriptive statistics

Descriptive statistics for the number of ideas correctly recalled for each delay condition is available in Table 5.

**Table 5**

*Number of Ideas of Story C Correctly Recalled Per Experimenter and Delay Group*

Experimenter	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
PhD student						
Day	10	6.70	4.69	0.00	6.00	13.00
Week	14	4.29	3.07	0.00	4.00	10.00
Month	9	4.89	4.08	0.00	4.00	13.00
Undergraduate students						
Day	19	12.63	3.00	7.00	13.00	18.00
Week	21	7.95	2.65	1.00	8.00	12.00
Month	18	6.56	2.99	0.00	7.00	10.00

*Note.* The Story C recall is scored over 25 ideas. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

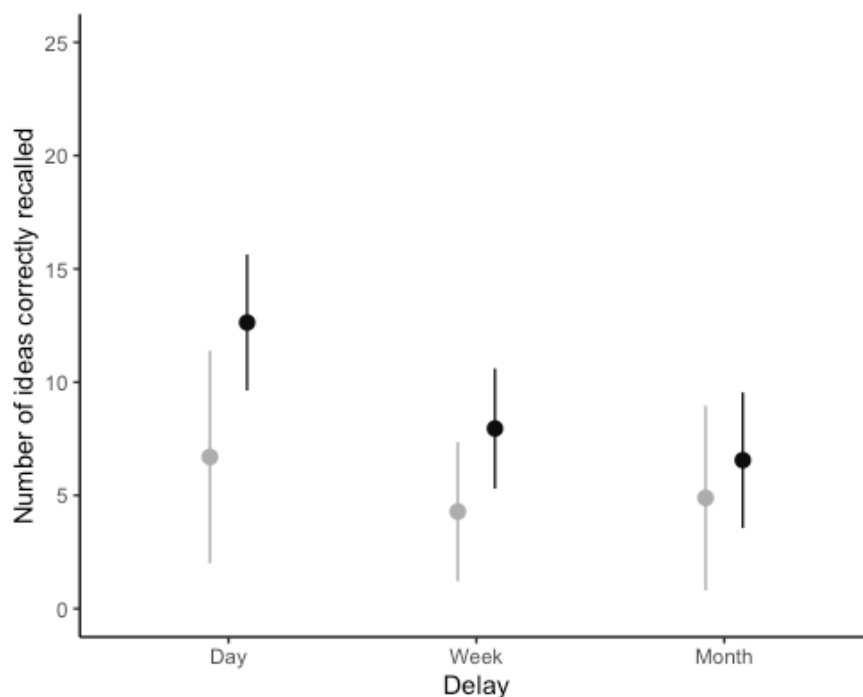
## Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test for homogeneity of variance was not significant,  $F(5, 85) = 1.32, p = .26$ . The Shapiro-Wilk's test was not significant,  $W = 0.10, p = .82$ . The

ANOVA model showed a main effect of delay,  $F(2, 85) = 17.33$ ,  $p < .001$ ,  $f = 0.54$ , and a main effect of the experimenter,  $F(1, 85) = 28.30$ ,  $p < .001$ ,  $f = 0.48$ . Participants who completed the experiment with the undergraduate students ( $M = 9.05$ ,  $SD = 3.83$ ) had a higher retention than participants who did the experiment with the PhD student ( $M = 5.18$ ,  $SD = 3.91$ ). However, no significant interaction was found between the delay and experimenter variables,  $F(2, 85) = 2.67$ ,  $p = .07$ . An illustration of the results can be found in Figure 4.

#### Figure 4

Number of Ideas of Story C Correctly Recalled as a Function of Experimenters and Delay Groups



Note: The dots represent the means, and the straight lines represent the standard deviations. The left-sided light grey lines and dots represent data collected by the

PhD student. The right-sided dark grey lines and dots represents data collected by the undergraduate students.

## **Discussion**

This experiment aimed to investigate the long-term forgetting of Story C (WMS-IV, Weschler, 2009). Forgetting was measured through the retention of the story after one day, one week and one month. The results suggest that the retention of Story C decreases with time. However, this decrease in retention might not follow a linear trend as no significant difference was found between the one week and one month delay intervals. Such a finding is coherent with the literature describing the rate of retention over time as a rapid initial decrease that tends to slow down to approach an asymptote (Baddeley, 1990; Della Sala, 2010).

A main limitation of the current experiment is the impact of changes in testing conditions. Changes were made to cope with the Scottish government advice following the lockdown in March 2020. An exploratory analysis revealed that participants tested with the paper-and-pencil version by the undergraduate students performed better than participants who completed the computerised version with me. However, this effect did not interact with the impact of delay on retention. Thus, the change in experimental conditions did not influence the effect of the variable of interest (i.e., delay interval).

## 2.2.2 Experiment 1b: The retention of a Piece of Prose. Under Retrieval Practice with Incidental Free Recall

### **Method**

#### ***Participants***

For consistency with Experiment 1a, I aimed to recruit approximately the same number of participants into each group as Experiment 1a (i.e., between 27 and 35).

29 participants (21 men, 8 women) aged from 18 to 60 years ( $M_{\text{age}} = 29.90$ ,  $SD_{\text{age}} = 11.90$ ) were recruited among the students and staff of the University of Edinburgh via e-mails and word-to-mouth. No participant reported any cognitive or neurological condition. All participants provided written consent to take part in the study. The study received ethical approval (No 66-1819/1) from the PPLS ethics committee of the University of Edinburgh.

#### ***Materials***

The materials used in Experiment 1b were the same as Experiment 1a. The Story C from the Logical Memory subtest of the WMS-IV (Weschler, 2009) measured episodic memory retention. The Digit Span-Forward and Digit Span-Backward subtests of the WAIS-IV (Weschler, 2008) was used as a lure to avoid intentional rehearsal of Story C. Thus, the Digit Span-Forward and Digit Span-Backward subtests were presented as the main tasks of the experiment. The Story C recall task was presented as a distractive task.

### ***Design***

A within-subjects experimental design was used. The within-subjects independent variable was the time interval for delayed recall. The same delay intervals as for Experiment 1a were used - one day, one week, and one month. The dependant variable was the total number of ideas from Story C correctly recalled.

### ***Procedure***

The first session was identical to the first session of Experiment 1a. At the end of the first session, the date and times of follow-up sessions via phone calls were planned. Participants were aware of the number of sessions involved. However, they did not know which sessions involved the Story C recall task. Retrieval practice of Story C was induced through the free incidental recall of the story at one day (session 2), one week (session 4), and one month delay (session 5). The instruction for delayed recall was: "Can you tell me everything you remember about the story you heard yesterday/ last week/ last month? Please recall everything you can, even if you are not sure." No time limit was set for the participant to answer. The recall task ended when the participants informed the experimenter that they could not recall more. The third session involving the Digit Span-Forward and Digit Span-Backward only was added as a ploy at a three-day delay. This procedure was meant to reduce the chances of participants expecting the Story C recall task during the last session at one week delay (session 4). During session 3, participants completed the digit span only. At the end of the last session, the purpose of the study was revealed, and the experiment was explained. No participant reported knowing they were being deceived

about the study's primary aim. A summary of the procedure can be found in Table 6.

**Table 6**

*Testing Protocol of Experiment 1b*

Testing Session	1 <sup>st</sup> session	2 <sup>nd</sup> session	3 <sup>rd</sup> session	4 <sup>th</sup> session	5 <sup>th</sup> session
Delay	Immediate	One day	Three days	One week	One month
Tasks	Digit span	Digit span	Digit span	Digit span	-
	Recall task	Recall task	-	Recall task	Recall task

*Note.* The testing protocol was not described to participants to avoid intentional rehearsal of the story between sessions.

COVID-19 notice

As in Experiment 1a, a second round of recruitment for Experiment 1b was conducted during the first lockdown in Scotland (March 2020). See the COVID-19 notice section of Experiment 1a for detailed information on how the first session was modified. Follow up sessions were not impacted by the lockdown as they were conducted via phone calls.

**Statistical analyses**

The analysis script and a package containing the anonymised dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdbf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdbf7ec8b03ea1ad66).

The hypotheses were tested using the Friedman's test non-parametric equivalent of the one-way ANOVA for repeated measures. First, I tested the impact of delay intervals on the retention of story C. Then, Wilcoxon's rank-sum tests with

Bonferroni corrections were used for pairwise comparisons. A 2x3 mixed ANOVA was run to verify whether the variation in the administration condition<sup>4</sup> (computerised vs. paper-and-pencil) had an impact on the effect of delay intervals (one day, one week, and one month). Statistical analyses were performed in RStudio, R version 4.1.0 (2021-05-18). An R Markdown document reporting the R script and R outputs of the statistical analysis is available in Appendix 1.

## Results

### *Effect of delay intervals on the retention of story C*

#### Descriptive statistics

The descriptive statistics of the retention for each delay interval are presented in Table 7.

**Table 7**

*Number of Ideas of Story C Correctly Recalled or Each Delay Interval*

Delay	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Day	10.41	4.27	1.00	11.00	16.00
Week	10.17	4.34	2.00	11.00	18.00
Month	8.76	3.95	2.00	9.00	17.00

<sup>4</sup> It must be noted that I did not receive assistance for the recruitment and data collection of Experiment 1b.

*Note.* The Story C recall is scored over 25 ideas. *M*, *SD*, *Min*, and *Max* refer to mean, standard deviation, minimum, and maximum respectively.

#### Inferential statistics

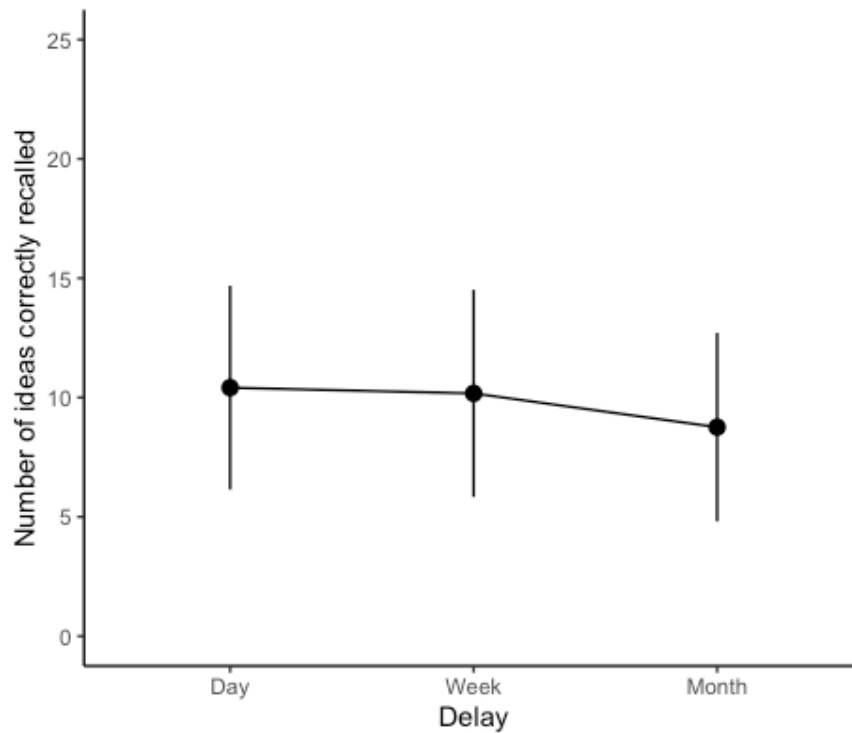
Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Mauchly's test was not significant,  $W= 0.91$ ,  $p= .29$ , indicating that variances are homogeneous. Shapiro Wilk's normality test suggests that the residuals are not normally distributed  $W= 0.97$ ,  $p= .03$ . Given that the assumptions for the one-way repeated measures ANOVA<sup>5</sup> were not met, a Friedman's test was run. The results indicated a significant main effect of delay intervals on retention,  $\chi^2(2)= 13.13$ ,  $p= .001$ ,  $W= 0.23$ . However, pairwise comparisons with Wilcoxon rank test revealed no significant difference in retention between the one day and one week delay ( $p= 1$ ), the one day and one month delay ( $p= 0.25$ ), as well as one week and one month ( $p= 0.41$ ). Results are illustrated in Figure 5.

---

<sup>5</sup> The results of the repeated measured ANOVA are available in Appendix 1 for information.

**Figure 5**

*Number of Ideas Correctly Recalled Per Delay Interval*



*Note:* The dots represent the means, and the vertical straight lines represent the standard deviations.

***Exploratory analysis: The effect of variations in the first session's tasks administration***

**Descriptive statistics**

Descriptive statistics for the number of ideas correctly recalled for each delay condition is available in Table 8.

**Table 8**

*Number of Ideas of Story C Correctly Recalled as a Function of Administration*

*Conditions and Delay Intervals*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Computerised	10					
Day		9.20	6.11	1.00	9.50	16.00
Week		8.80	6.23	2.00	8.50	18.00
Month		7.60	5.38	2.00	8.00	17.00
Paper-and-pencil	19					
Day		11.05	2.91	5.00	12.00	15.00
Week		10.89	2.88	5.00	12.00	15.00
Month		9.37	2.95	4.00	9.00	16.00

*Note.* The Story C recall is scored over 25 ideas. *M*, *SD*, *Min*, and *Max* refer to mean, standard deviation, minimum, and maximum respectively.

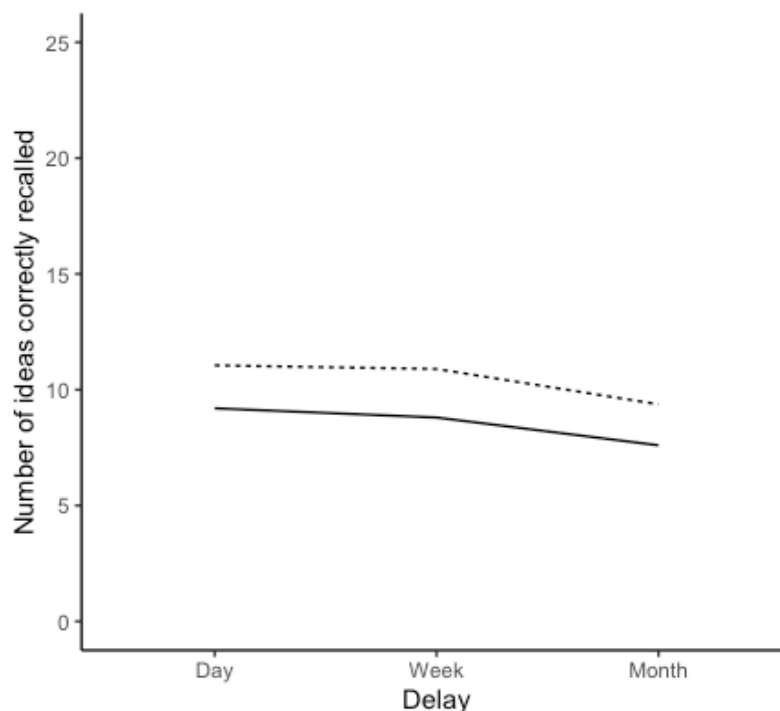
#### Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Mauchly's test was not significant,  $W= 0.91$ ,  $p= .29$ , indicating that variances are homogeneous. The Shapiro Wilk's normality test suggests that the residuals are normally distributed  $W= 0.98$ ,  $p= .14$ . The mixed-model ANOVA showed a main effect of delay intervals,  $F(2, 54)= 9.88$ ,  $p< .001$ ,  $\eta^2G= 0.03$ . However, they

were no significant main effect of administration,  $F(1, 27) = 1.49, p = .23$ , and no interaction effect,  $F(2, 54) = 0.09, p = .91$ . An illustration of the results is available in Figure 6.

### Figure 6

*Number of Ideas of Story C Correctly Recalled as a Function of Delay Intervals and Administration Conditions*



*Note.* The straight line represents recall when the first session involved computerised online tasks. The dotted line represents recall when the first session involved in-person, paper-and-pencil tasks.

### Discussion

This experiment aimed to investigate the retention of Story C under retrieval practice. Retrieval practice was induced using incidental free recall of the story after

one day, one week, and one month. The results showed a main effect of delay on retention over time. Interestingly, no decrease in retention was not found when comparing each interval with one another. This finding corroborates my prediction that forgetting levels off when practising retrieval. Finally, it must be noted that the switch from a paper-and-pencil to a computerised version of the task for the first session of the experiment did not impact the results.

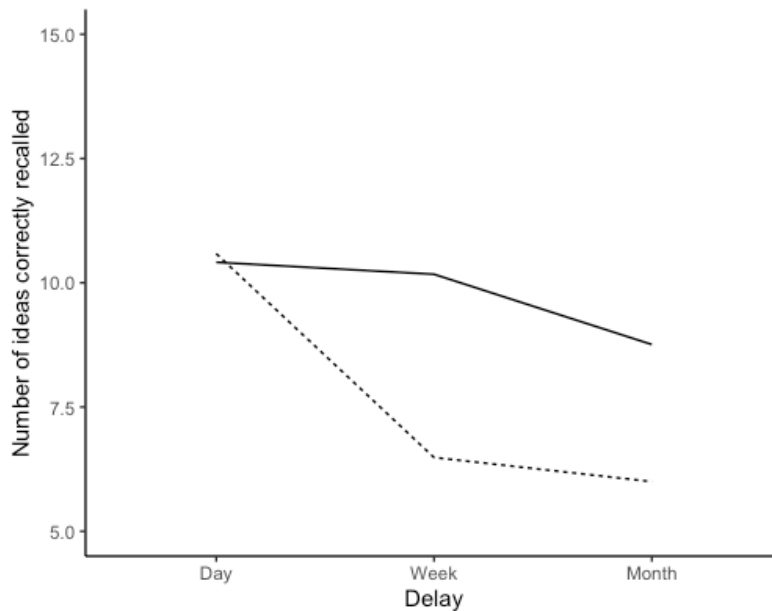
### 2.2.3 Experiment 1 Discussion

The study aimed to explore the impact of forgetting and learning on the retention of a piece of prose. Two experiments were conducted. Experiment 1a focused on forgetting by observing the retention of Story C (WMS-IV, Weschler, 2009) across three delay intervals - one day, one week, and one month, using a between-subjects design. Experiment 1b examined how learning impacts the long-term retention of Story C when using retrieval practice using a within-subject design.

Together, the two experiments provided findings in line with the prediction from the literature (Figure 2). A summary of the findings is illustrated in Figure 7. On the one hand, the forgetting of Story C is characterised by a rapid decrease in retention flooring in the long run. On the other hand, practising retrieval leads to better and more steady story retention over time. Thus, the paradigms used in this study are suitable to investigate the forgetting and learning of a short narrative.

**Figure 7**

*Retention of Story C Under Forgetting vs. Retrieval Practice*



Note. The straight and dashed lines represent retrieval practice and forgetting respectively.

### **2.3 Experiment 2: The Long-Term Retention of a Dilemma Story Under Retrieval Practice**

The present chapter presents the development of a new experimental paradigm to study the impact of decision-making on memory. Previous research on the topic (Murty et al., 2019) rely on a decision-making task that does not meet the definition of decision-making presented in this thesis. To address this issue, I suggest measuring the impact of decision-making on memory through the learning and forgetting of a moral dilemma story. Moral dilemmas are short stories involving a moral conflict calling upon opposing ethical reasons (Christensen & Gomila, 2012). Experiment 1 presents the forgetting of a standardised short story without (Experiment 1a) and with retrieval

practice (Experiment 1b). In this experiment, the study materials were Story C from the Logical Memory subtest of the Wechsler Memory Scale -- fourth edition (WMS-IV, Wechsler, 2008). Experiment 1 characterised the long-term forgetting of Story C and how such forgetting was impacted by retrieval practice. The current experiment compares the retention of a moral dilemma story to the retention of Story C under retrieval practice. Such comparison informs on whether dilemma stories observe typical learning effects on forgetting.

A measure of dilemma story recall is reported in Arbuckle and Harsany (1985). In this experiment, the authors investigate the impact of age differences in the recall of moral dilemmas as a function of recall instructions. The research focused on the effects of instructions and age but not on moral dilemma stories. Therefore, Arbuckle & Harsany's (1985) findings are not directly relevant to my experiment, despite relying on similar paradigms (i.e., recalling dilemma stories).

Moreover, this thesis focuses explicitly on forgetting and learning of a dilemma story. To what extent dilemma stories are subject to typical retention and learning is unknown. Therefore, before measuring the effect of decision-making in a moral dilemma on the retention of the dilemma story, I suggest controlling for a potential impact of the dilemma story per se when no-decision-making is asked.

### 2.3.1 Method

#### **Participants**

41 participants (9 men, 32 women) aged from 19 to 65 years ( $M_{\text{age}}=31.92$ ,  $SD_{\text{age}}= 13.48$ ) were recruited among the students and staff of the University of

Edinburgh. 19 participants were recruited as part of Experiment 1b. Seven participants withdrew from the study as they could not attend follow up sessions. Their data were removed for the statistical analysis. No participant reported any cognitive or neurological condition. All participants provided written consent to take part in the study. The study received ethical approval (No 347-1718/13) from the PPLS ethics committee of the University of Edinburgh.

## **Materials**

### ***Digit Span Tasks***

In this experiment, the administration of the digit span was used as a lure to avoid intentional rehearsal of the story. The administration of the Digit Span-Forward and Digit Span-Backward subtests followed the standardised procedure of the Wechsler Intelligence Scale for Adults – IV (WAIS-IV, Wechsler, 2008). In both Digit Span-Forward and Digit Span-Backward, there are two sequences of digits per item. The experimenters read each sequence verbatim at the rate of one digit per second, dropping the voice slightly down on the last digit of the sequence. After being read a sequence, the participant must recall all digits in the same order for Digit Span-Forward and reverse order for the Digit Span-Backward. Each subtest is composed of eight items, resulting in sixteen sequences. In the Digit Span-Forward, the sequences go from two (item 1) to nine digits (item 8), increasing from one digit every item. In the Digit Span-Backward, the sequences start with two digits for items 1 and 2, then increase from one to seven digits (item 8). Each sequence is scored 0 if failed and 1 if correctly recalled. A subtest ends when both sequences from the same item are

failed. The raw scores correspond to number of sequences correctly recalled (out of sixteen) for each subtest.

### **Story C**

The administration of Story C from the Logical Memory subtest of the WMS-IV (Wechsler, 2009) was adapted from the standardised procedure of the WMS-IV. Instructions referring to the recognition task were ignored. The recognition task was not included in the procedure to prevent re-exposure to the study materials. To ensure regularity in reading speed, appropriate pronunciation, and no variation in the speech between the experimenters, the reading of Story C was pre-recorded. Story C was read by a Text-to-speech software, The Scottish Voice, provided by the University of Edinburgh (The Scottish Voice, 2011). The story was read by a Scottish male voice - "Cerevoice Stuart" version 4.0.1. The task started with the following instructions: "You are going to listen to a short story. Listen carefully and try to remember as much as you can". Then, the recording of Story C was displayed. Following the recording, the experimenter asked: "Can you tell me everything you remember about the story? Please recall everything you can, even if you are not sure." For delayed recall, participants were asked: "Can you tell me everything you remember about the story you heard yesterday/ last week/ last month? Please recall everything you can, even if you are not sure." No time limit was set for the participant to answer. The recall task ended when the participants informed the experimenter that they could not recall more. Story C was divided into 25 "ideas" to be recalled. Each idea was scored 1 if correctly recalled and 0 if not recalled. The raw score corresponds to the total number of ideas correctly recalled out of 25.

### ***Dilemma story***

A new dilemma story was invented based on the logic of the Trolley Dilemma (Foot, 1967/2002). See Section 1.1.3 for a presentation of the Trolley Dilemma. The story opposed two conflicting reasons - a utilitarian and a deontological moral reason. The former suggested causing harm to a minority for the sake of the majority. The latter called upon refusing to cause any harm independently of the greater benefit. The story was developed to approximate the lexical characteristics of Story C of the logical memory subtest. The lexical characteristics of the two stories is available in Table 9. The text was divided into 25 ideas to be recalled, copying the scoring form of Story C. The Dilemma Story recall scoring form is available in Appendix 2. By drawing on the Logical Memory subtest of the WMS-IV (administration, story, and scoring), I have aimed to optimise psychometric qualities and allow comparisons between retention rates. The reading of the dilemma story was also recorded with “Cerevoice Stuart” version 4.0.1

**Table 9**

*Comparison Between the Lexical Characteristics of The Dilemma Story and The Story C*

Lexical characteristics	Dilemma story	Story C
Total ideas	25	25
Total sentences	5	5
Mean sentence length	17.2	17.2
Total words	86	86
Total syllables	132	124
Mean syllables per words	1.53	1.44
Flesch reading ease index*	59.93	64
Flesch-Kincaird index**	9.17	9

*Note.* \*This test rates text on a 100-point scale. The higher the score, the easier it is to understand the document. The formula for the Flesch Reading Ease score is:  $206.835 - (1.015 \times ASL) - (84.6 \times ASW)$ , where ASL = *average sentence length* and ASW = *average number of syllables per word*.

\*\*This test rates text on a U.S. school grade level. For example, a score of 8.0 means that an eighth-grader can understand the document. Most documents aim for a score of approximately 7.0 to 8.0. The formula for the Flesch-Kincaird Grade Level score is:  $(.39 \times ASL) + (11.8 \times ASW) - 15.59$

## **Design**

A mixed design was employed. The experiment included one between-subjects independent variable. This variable is referred to later as the story type and corresponds to the story participants heard - Story C or the dilemma story. The design also included one within-subject variable, delay intervals with four levels: immediate, one day, one week and one month. The dependent variable was the number of ideas from the story correctly recalled.

## **Procedure**

The experimental procedure strictly follows the retrieval practice procedure of Experiment 1b. The experiment was divided into five sessions – one initial learning session, one lure session, and three retrieval practice sessions. The study was first presented to the participant as an experiment on working memory with the Digit Span-Forward and the Digit Span-Backward were presented as the main tasks of the experiment. The story recall was presented as a distractive task. The lure session involving the Digit Span- Forward and the Digit Span-Backward only aimed to participants' adherence to the study instructions and avoid intentional rehearsal of the story between sessions.

The first session started with the Digit Span-Foward and the Digit Span-Backward. This was followed by the story recall task. The first session ended with collecting socio-demographic information and planning the follow-up sessions. A summary of the testing protocol is presented in Table 6. Participants were aware of the number of sessions involved. However, they did not know which sessions involved the recall task.

At the end of the last session, the purpose of the study was revealed, and the experiment was explained. No participant reported knowing they were being deceived about the main aim of the study.

### 2.3.2 Statistical analysis

The analysis script and a package containing the anonymised dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdbf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdbf7ec8b03ea1ad66).

A 2x4 mixed ANOVA was used to compare retention as a function of delay intervals and story types. Pairwise comparisons were performed using t-tests with Bonferroni corrections. Note that part of the dataset included in this analysis, data for the Story C condition, were also reported and analysed separately in Experiment 1b.

### 2.3.3 Results

#### **Descriptive Statistics**

Descriptive statistics of story recall are available in Table 10.

**Table 10**

*Number of Ideas Correctly Recalled as a Function of Story Type and Delay Intervals*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Story C	19					
Immediate		13.42	2.43	9.00	14.00	17.00
Day		11.05	2.91	5.00	12.00	15.00
Week		10.89	2.88	5.00	12.00	15.00
Month		9.37	2.95	4.00	9.00	16.00
Dilemma	15					
Immediate		10.47	3.70	7.00	9.00	18.00
Day		8.47	4.17	3.00	8.00	19.00
Week		7.87	3.66	2.00	8.00	16.00
Month		6.27	3.01	1.00	6.00	11.00

*Note.* The story recall is scored out of 25 ideas. *N*, *M*, *SD*, *Min* and *Max* represent group size, mean, standard deviation, minimum and maximum respectively.

### **Inferential statistics**

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Mauchly's test was not significant,  $W= 0.81$ ,  $p= .27$ , indicating that

variances are homogeneous. The Shapiro Wilk's normality test suggests that the residuals are normally distributed  $W= 0.98, p= .14$ .

The mixed-model ANOVA revealed a significant main effect of story type,  $F(1,32) = 8.22, p < .001, \eta_G^2 = 0.18$ . A main effect of the delay intervals was also observed,  $F(3,96) = 45.82, p < .001, \eta_G^2 = 0.18$ . No significant interaction effect was found between the story type and the delay interval variables,  $F(3,96) = 0.21, p = 0.89$ .

The pairwise comparison with Bonferroni method suggests that the total recall at one-week ( $p= .02$ ) and one-month ( $p>.001$ ) delay intervals significantly differ from the Immediate session. All other comparisons did not reach significance (see Table 11).

**Table 11**

*Adjusted P Values of Pairwise Comparisons with Bonferroni Method on The Four - Levels of The Delay Interval Variable*

	Immediate	Day	Week
Day	.06	-	-
Week	.02	1.00	-
Month	<.001	.15	.40

#### 2.3.4 Discussion

The present experiment aimed to investigate the retention of a dilemma story. The retention of the dilemma story under retrieval practice was compared to the

retention of Story C. The results showed that participants who listened to the dilemma story recalled less than participants who listened to Story C. The difference in retention between the dilemma story and Story C is likely due to the dilemma being a more complex narrative. Indeed, the Flesch reading ease index (Table 9), the descriptive statistics and comments made by participants during the debriefing of the experiment (not recorded nor transcribed) supports this idea. Indeed, the Flesch reading ease index indicates that Story C is easier to understand than the dilemma story. Also, participants in the dilemma story condition reported difficulties understanding the story. Noteworthy, pairwise comparisons have shown an absence of difference in recall between the day, week and month sessions. Retrieval practice explains the absence of memory decay over time.

However, the absence of interaction between the story type and the delay intervals indicates that the benefit of retrieval practice on retention is equivalent across the two stories. Overall, the dilemma story leads to lower retention than Story C, but it does not show a different rate of forgetting under retrieval practice.

#### **2.4 Experiment 3: A Comparison of Moral Decision Types in a Moral Dilemmas Set.**

The previous experiment suggested that retrieval practice influences the retention of a dilemmas story to the same extent as a standardised piece of prose. However, the dilemma story designed for Experiment 2 does not seem to be the most adequate for studying forgetting. The Flesch reading ease index and participants' comments indicated that the dilemma story was difficult to understand. For

investigating the impact of decision-making on memory, dilemmas with a simple structure, easily understandable, should be used.

Moreover, to study the impact of decision-making on memory, the dilemma story should allow comparisons between retention following a utilitarian and deontological moral decision. Thus, the current experiment aims to identify a moral dilemma story that equally spread participants across the two types of moral decisions - a ratio approximating 1:2 of utilitarian and deontological decisions.

One straightforward solution to finding an accessible dilemma is to use another dilemma from published moral dilemma sets. Using dilemmas repeatedly studied in the literature should ensure their understandability. Greene et al. (2001) published a set of moral dilemmas widely used in judgment and decision-making research (Christensen & Gomila, 2012). This dilemma set was adapted by Moore et al. (2008), who proposed a set with more balance characteristics across dilemmas (e.g., equal number of deaths, comparable phrasing, identical length). Following personal communications, A. Moore provided me with the ratio of moral decision types (i.e., utilitarian vs. deontological) for each dilemma obtained by Moore et al. (2008) but not reported in their results (A. Moore, personal communication, September 24, 2018). Ratios of moral decision types for a set of dilemmas were reported in Harrison's et al. (2008). Thus, the current experiment compares decision type ratios of moral dilemmas from Greene et al. (2008) and Harrison et al. (2008).

## 2.4.1 Method

### Participants

111 participants (79 women, 31 men, 1 other) aged from 19 to 70 years ( $M_{\text{age}}=29.00$ ,  $SD_{\text{age}}= 8.20$ ) were recruited through online advertisement. No participant reported any cognitive or neurological condition. All participants provided their informed consent. The study received ethical approval (No 30-1819/7) from the PPLS ethics committee of the University of Edinburgh. Participants received a small honorarium as compensation for their participation.

### Materials

Three dilemmas from Moore et al. (2008) were pre-selected based on the ratios of moral decision types previously obtained (Appendix 3) and their lexical characteristics (Appendix 4): the *Bike Week*, the *Nobel Prize*, and the *Vaccine*. One dilemma from Harrison et al. (2008) with a fairly even endorsement rate was also selected: the *Best Friends*. The dilemmas mentioned above were originally presented in a written format and mainly contain gist information. To be used as an episodic memory task in the following experiments, the dilemma should instead be read and include peripheral details to be recalled – as per stories from the Logical Memory subtest of the WMS-IV (Wechsler, 2009). Therefore, two modified versions for each of the aforementioned dilemmas were created. These modified versions differed in terms of administration (audio vs written) and story content (original vs with added peripheral details). The final set contained sixteen stimuli – four different versions of

each dilemma story. The written versions of the dilemmas (original vs with added peripheral details) are provided in Appendix 5.

## Procedure

An online survey was designed using QualtricsXM. The survey started with an information and consent form on which participants agreed by clicking on “Next”. Participants were exposed to and answered each dilemma in only one of their four versions – written or audio, original or with added details. The survey ended with demographic questions.

### 2.4.2 Results

Moral decision ratios for each version of the dilemmas are reported in percentage in Table 12.

**Table 12**

*Percentage of Endorsement for Each Dilemma*

Decision	The Bike Week				The Nobel Prize			
	Original		Detailed		Original		Detailed	
	Audio	Written	Audio	Written	Audio	Written	Audio	Written
	Deontological	72	64	78	58	50	71	53
Utilitarian	28	36	22	42	50	29	47	31
Decision	The Vaccine				The Best Friends			
	Original		Detailed		Original		Detailed	
	Audio	Written	Audio	Written	Audio	Written	Audio	Written
	Deontological	55	84	56	98	54	12	93

Utilitarian	45	16	44	2	46	88	7	89
-------------	----	----	----	---	----	----	---	----

---

The results suggest that the *Nobel Prize* original audio version leads to the most even proportion of moral decision types with 50% of deontological and utilitarian decision provided. The detailed audio version of the *Nobel Prize* had the second most even proportion of moral decision types with 53% of deontological and 47% of utilitarian decisions.

### 2.4.3 Discussion

The objective of this experiment study was to find an appropriate dilemma story for studying the impact of decision-making on memory. To do so, I selected from the existing literature four dilemmas with the most even endorsement rates reported (Moore et al., 2008; Harrison et al., 2008). Each dilemma was modified so there were versions with added details and audio versions, mimicking the format of the WMS-IV Logical Memory subtest (Wechsler, 2009).

The objective of this experiment was to find a dilemma story dividing participants evenly across moral decision types. To do so, I selected four dilemmas from the literature which previously led to an even ratio across moral decision types (Moore et al., 2008; Harrison et al., 2008). I created a modified version with added details and audio display of each dilemma to mimic the administration format of the WMS-IV Logical Memory subtest (Wechsler, 2008).

The Nobel Prize dilemma seemed to be the most suitable. The audio version I have designed from Moore et al.'s (2008) has perfectly even percentages of decision types – 50% for each option. Its audio version with added details was the closest to

being entirely balanced with 53% and 47% for deontological and utilitarian decisions, respectively. Therefore, the detailed audio version of the Nobel Prize dilemma appears to be the best fitting materials for my research purposes.

## **2.5 Chapter 2 Discussion**

The current chapter presented the development of a new experimental paradigm to study the impact of decision-making on memory. Experiment 1 illustrated the long-term forgetting and the benefit of retrieval practice on the retention of a standardised piece of prose. Experiment 2 compared the benefit of retrieval practice on retention between the piece of prose and the moral dilemma story. Experiment 3 described the percentage of moral decision types provided for a set of moral dilemmas. All findings considered, this chapter led to designing a unique experimental paradigm. The following experiments will study the impact of decision-making on episodic memory through the forgetting and retrieval practice of a modified version of the Nobel Prize dilemma.

## **Chapter 3: The impact of decision-making on long-term episodic memory**

### **3.1 Introduction**

Research on decision-making and memory considers that both functions interact and mutually influence one another. Research has investigated this relationship at both neuroanatomical and cognitive levels. At the neuroanatomical level, evidence suggests that interactions between the hippocampus and the medial prefrontal cortex promote memory integration and decision-making (see Euston et al., 2012; Schlichting & Preston, 2017, for reviews). From a cognitive perspective, theories consider that decision-making is based on memory recollection, and the decision made leads to new mnemonic representations (Tremel et al., 2018; Fellows, 2018). However, empirical research mainly provides evidence of one side of the relationship - the impact of memory on decision-making. The reciprocal assumption has received less support. To our knowledge, only Murty et al. (2019) have attempted to provide empirical evidence of the impact of decision-making on memory.

In Murty et al. (2019), participants are presented with two occluder screens illustrated with Japanese syllabaries - Hiraganas. In the "choice" condition, participants decided which occluder screens to remove. In the "fixed" condition, the experimenter indicated which occluder screen participants should select for removal. This selection phase was followed by an encoding phase. In the encoding phase, the removal of the occluders screen revealed a hidden object image. The same image was hidden behind both occluder screens. Participants were asked to memorise the object on the image. A post-encoding rating task followed where participants rated how much they liked the Hiragana selected on a 5-point scale. The rating task was

followed by a recognition task of the objects presented in the encoding task. Murty et al.'s (2019) procedure is illustrated in Figure 1. The experiment ended with the completion of a second recognition task after a 24-hour delay. The results showed reduced forgetting for object pictures in the choice condition over the fixed condition after 24 hours. Murty et al. (2019) concluded that decision-making enhances episodic memory via post-encoding consolidation.

Although Murty et al. (2019) support current theories on the impact of decision-making on memory, I argue that evidence remains limited. As presented in Section 1.1.1, decision-making is the process of three steps: judgement, preference and choice. Decision-making tasks usually elicit these three steps by presenting decision options with conflicting outcomes. Murty et al.'s (2019) experimental paradigm fails to meet this definition in several ways. First, the experimental design involves *perceived* choice only; each choice option is associated with the same outcome. Indeed, the participant can choose the right or left screen, but both will reveal the same object. Moreover, the task does not involve any judgment of potential choice outcomes. Therefore, their experimental design does not involve judgments or preferences but only choices. Overall, Murty et al. (2019) showed that choice-making enhances long-term episodic memory, but the effect of decision-making on memory remains to be verified.

This chapter reports a conceptual replication of Murty et al.'s (2019) experiment. To address the criticism raised on Murty et al.'s (2019) task, I developed a new experimental paradigm. The previous chapter (Chapter 2) presents three experiments testing the adequacy of different tasks. Experiment 1 compared the long-term

forgetting and retrieval practice of a standardised piece of prose. Experiment 2 compared the long-term retention of the standardised piece of prose and a dilemma story. Experiment 3 compared the percentage of moral decision types for a set of moral dilemmas. Overall, the previous chapter suggests that the impact of decision-making on episodic memory can be investigated through the forgetting and retrieval practice of a modified version of the *Nobel Prize* (adapted from Moore et al., 2008).

The *Nobel Prize* dilemma (see Appendix 5) is a hypothetical short story presenting two conflicting moral reasons. The reader and a fellow researcher have made a discovery. The discovery can either be used for good or evil purposes. The colleague wants to sell the discovery to a buyer who will use it as a weapon. The reader can let him sell the discovery or kill him in a setup scenario where his death will sound like an accident. On the one hand, killing the colleague involves considering the costs and benefits and is consistent with utilitarian principles. According to Greene's dual-process model (Greene et al., 2001, 2004, 2008), utilitarian judgments are associated with controlled cognitive processes. Greene et al. (2008) showed a selective interference with utilitarian judgments under cognitive load. On the other hand, letting the colleague sell the discovery is consistent with the deontological principle as it advocates the "not to kill" principle (Christensen et al., 2014). Such a judgment is associated with an automatic emotional response (Greene et al., 2001, 2004, 2008).

This original design aims to provide new insights into the relationship between decision-making and memory. Considering Murty et al.'s (2019) findings and current theories (Tremel et al., 2018; Fellows, 2018), making a decision should enhance

memory retention. Moreover, the dual-process model (Greene et al., 2001, 2004, 2008) states that making a utilitarian decision requires further cognitive processing of the story content. Therefore, I hypothesise that making a utilitarian decision should further enhance long-term retention compared to a deontological decision. Such findings would provide new empirical evidence of the impact of decision-making on memory.

Experiment 4 presents a conceptual replication of Murty et al. (2019) in the context of moral decision-making. The experiment investigates the impact of decision-making and moral decision types on the retention of a dilemma story. This research expands previous findings by considering whether distinct moral decision types can impact episodic memory differently.

Experiment 5 focuses on the impact of decision-making types on long-term forgetting. Murty et al. (2019) suggest that research should focus on the impact of decision-making on memory consolidation. The authors found reduced forgetting rates across an immediate to a one-day time interval in their study. Nonetheless, more investigation is required to establish an impact of decision-making on forgetting rates. Since forgetting is a curvilinear process (Bogartz et al. 1990; Rubin & Wenzel, 1996), a measurement of retention with three-time intervals at least is required to observe differences in forgetting curves. My experiment addresses this issue by studying the impact of moral decision types on long-term forgetting.

### 3.2 Experiment 4: The Impact of Decision-Making on Memory Retention after A Day

Tversky and Kahneman's (1973) *availability heuristic* states that the accessibility of instances, when retrieved from memory, influences judgments and decisions. This theory disseminated the idea that memory and decision-making are inherently related. For half a century, research has provided abundant evidence of how memory influences decision-making (Redish & Mizumori, 2015; Madan et al. 2014; Johnson & Busemeyer, 2010; Dougherty et al., 2003). However, investigations of the impact of decision-making on memory have been somewhat neglected. To my knowledge, the only research on the topic suggest that decision-making reduces forgetting (Murty et al., 2019). The current experiment investigates whether a similar effect stands in the context of moral decision-making.

Moral dilemmas are very popular in judgement and decision-making research (Christensen & Gomila, 2012). Moral dilemmas are short pieces of prose eliciting social decision-making. One of the most famous moral dilemmas is the *Trolley dilemma* (Foot, 1967/2002). In this dilemma, a runaway trolley is heading for five railway workers. The driver of the trolley has the option to divert it to another track where only one worker will be killed. The decision is whether to sacrifice one person to save more people. The narrative form of the dilemmas offers materials for assessing memory recall. Therefore, dilemmas enable manipulating decision-making and measuring memory recall together. As such, using dilemma stories are especially appropriate for studying the impact of decision-making on memory.

To this end, a unique experimental design combining moral decision-making and memory paradigms was created. The design was devised throughout three experiments that are presented in Chapter 2. Such a design represents an innovative method for studying decision-making and memory. Within this novel approach, hypotheses are drawn from previous research and contemporary theories on decision-making and memory. According to Murty et al. (2019), decision-making enhance memory via post encoding consolidation. According to the dual-process model of moral decision-making (Greene et al., 2004, 2008), making a utilitarian decision result from a deeper level of cognitive processing). Considering that a deeper level of processing leads to better retention ( Craik & Lockhart, 1972), making utilitarian decision should increase the memory retention of the dilemma. Thus, I hypothesise that (1) making a decision improves memory retention and (2) such a benefit will be larger when making a utilitarian decision compared to a deontological decision.

### 3.2.1 Method

#### **Participants**

104 participants were recruited via the Department of Psychology participant pool and the general public using word-of-mouth and online advertisement. Volunteers who self-reported memory difficulties due to cognitive or neurological conditions were excluded from the experiment. The demographic characteristics of the sample are provided in Table 13. The study was approved by the University of Edinburgh - School of Philosophy, Psychology and Language Sciences Ethics Committee (No 24-1920/16). Participants gave informed consent in accordance with the British

Psychological Society Code of Ethics and Conduct. Participants received course credits or compensate small honorarium for their participation. A volunteer research assistant assisted me in the recruitment and data collection.

**Table 13**

*Socio-Demographic Characteristics of The Sample*

<i>N</i>	Age					Education					Gender		
						1 = High School	2= Undergraduate	3= Postgraduate	4= Other	NA	1= Female	2= Male	3= Other
	<i>M</i>	<i>SD</i>	Min	Median	Max	1	2	3	4	NA	1	2	3
104	25.90	6.42	17	24	61	2	49	43	8	2	80	23	1

*Note.* *N*, *M*, *SD*, Min, Max and NA refer to sample size, mean, standard deviation, minimum, maximum and missing data respectively. Missing data are either due to a non-classifiable input or to the item not being answered.

**Materials**

The administration of the tasks was computerised using OpenSesame. For all tasks, the instructions were written on the computer screen. To ensure there was regularity in reading speed, appropriate pronunciation and no variation in speech, the reading of the digit span items, and the dilemma story was recorded via a text-to-speech software (The Scottish Voice, 2011). The Cerevoice Stuart version 4.0.1, a Scottish male voice, was used. The readings were recorded and added to the computerised procedure. Audio recordings were presented through headphones placed and adjusted for volume before the experiment started. Participants provided

their answers by typing on a numeric keyboard or speaking aloud into the headphones' microphone, depending on the task.

### ***The Digit Span Tasks***

A computerized version of the Digit Span-Forward and Digit Span-Backward was developed. The administration of the Digit Span-Forward and Digit Span-Backward subtests was adapted from the standardised procedure of the Wechsler Intelligence Scale for Adults – IV (WAIS-IV, Wechsler, 2008). In both Digit Span-Forward and Digit Span-Backward, there are two sequences of digits per item. The sequences of digits were read by TheScottishVoice text-to-speech software (The Scottish Voice, 2011). After being read a sequence, the participant must recall all digits in the same order for Digit Span-Forward and in reverse order for the Digit Span-Backward. Each subtest is composed of eight items, resulting in sixteen sequences. In the Digit Span-Forward, the sequences go from two (item 1) to nine digits (item 8), increasing from one digit every item. In the Digit Span-Backward, the sequences start with two digits for items 1 and 2, then increase from one to seven digits (item 8). Each sequence is scored 0 if failed and 1 if correctly recalled. A subtest ends when both sequences from the same item are failed. The raw scores correspond to the number of sequences correctly recalled (out of sixteen) for each subtest.

The use of the digit span task was twofold. Firstly, as working memory is known to impact decision-making (Gansler et al., 2011), and memory retrieval (Cubelli, 2010), the digit span was administered to verify that experimental groups did not significantly differ in their working memory abilities. Secondly, digit span was presented to

participants as the main task of the experiment to avoid intentional rehearsal of the dilemma story during the delay periods.

### ***The Moral Dilemma Task***

Audio recording of the modified Nobel Prize Dilemma was used. For scoring purposes, the dilemma story was divided into 27 ideas to be recalled (see Appendix 6). The scoring followed that of the logical memory subtest of the Wechsler Memory Scale-IV (WMS-IV; Wechsler, 2009). Four external judges, PhD students in Psychology, who were native-English speakers, were involved in the conception of the scoring form. An initial version of the scoring form, dividing the dilemma story into logical “ideas” (i.e., scorable units), was presented to the judges. They were invited to indicate whether they agreed with each scorable unit. In cases of disagreement, comments along with suggested changes to the scorable units were encouraged. The scoring form was adapted following the judges’ advice.

### **Design**

A between-subjects design was used. The independent variable was the decision-making with three levels: no-decision, deontological and utilitarian. The dependent variable was the number of items correctly recalled from the dilemma story.

### **Procedure**

The experiment started with the completion of the computerised version of the Digit Span-Forward and Digit Span-Backward. The standardised instructions of the WAIS-IV were written on the computer screen. After that, the audio recording of the

first digit sequence was displayed at a rate of one digit per second. Then, the participant had to recall the digit sequence by typing the digits in the same order on the numeric keyboard. This was repeated for each digit sequence the Digit Span-Forward. Then, the Digit Span-Backward followed using the same procedure as above, but the digits were recalled in the reverse order that they were presented.

The moral dilemma task was performed after the completion of the digit span tasks. To mimic the procedure of the Logical Memory subtest of the WMS-IV (Wechsler, 2008), participants were given the following instructions: "You are going to listen to a short story. Listen carefully and try to remember as much as you can". Then, participants listened to the audio recording of the modified Nobel Prize dilemma.

The narrative of the story was the following: *"You and a fellow researcher, Professor Joe Grant, have discovered a powerful new energy source that is cheap, safe, and clean. You realize that this could lead to the elimination of pollution and poverty around the world. However, your colleague wants to sell this discovery to the US army, which will turn it into a powerful new weapon. You can prevent this by releasing propane gas in the lab, so that when he turns on his computer, it will cause an explosion within two seconds. This will also cause the death of your colleague. Everyone will think that it was just a lab accident, and the discovery will not be sold to those who might create a weapon out of it."*

Following the presentation of the dilemma story, participants were randomly allocated to the "decision" or "no-decision" (control) condition. In the decision condition, participants were asked to answer the question: "What are you going to do?" Below the question, participants were presented with the following two options: "You

release the gas” or “You let your colleague sell the discovery”. They were then allocated to the utilitarian condition if they answered, “You release the gas” or to the deontological condition if they answered, “You let your colleague sell the discovery”. The presentation order of the decision options was randomised to avoid order effects.

In the no-decision condition, participants were asked to answer an unrelated question: “Which colour do you prefer?” Below the question, participants were presented with the following two options in a randomised order: “You prefer the colour blue” or “You prefer the colour red”. This question was asked so to involve making a choice without enacting decision-making.

After answering one of the questions above, participants were given the instructions for memory recall: “Can you say everything you can remember about the story? You can say things even if you are not sure about the details or order”. Participants answered by speaking aloud. Answers were recorded and saved under an anonymised participant identification number. The dilemma story was divided into 27 ideas (Appendix 6). The retention of the story was measured through the number of ideas correctly recall. The first session ended with the collection of socio-demographic details. Participants were contacted via phone calls a day later for an incidental delayed free recall of the dilemma story. The instruction for delayed recall was: “Can you tell me everything you remember about the story you heard yesterday/ last week/ last month? Please recall everything you can, even if you are not sure.” No time limit was set for the participant to answer. The recall task ended when the participants informed the experimenter that they could not recall more. The recording and scoring procedures were the same as for the first session. The follow up session

ended with a debriefing. The purpose of the experiment was explained, and participants' questions were answered.

### 3.2.2 Statistical Analyses

The analysis script and a package containing the anonymised dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdf7ec8b03ea1ad66).

Inter-rater reliability was assessed using Cohen's  $\kappa$  for a random sample of ten participants' post-decision recall of the story.

Participants over 50 years old ( $n = 4$ ) were removed from the analyses as they were not homogeneously distributed across the three decision groups. Therefore, 100 participants were included in the analyses.

One-way analyses of variance (ANOVA) were used to verify that the three groups did not differ in their immediate recall of the dilemma story or in the Digit Span-Forward and Digit Span-Backward. For cases where ANOVA's assumptions were not met, Kruskal-Wallis' rank-sum tests nonparametric equivalent were used.

### **Contrasts analyses**

Although ANOVA is probably the most popular statistical tool in Psychology research (Zhou & Skidmore, 2017), it may not be the most adequate to test my research hypotheses. ANOVA is an omnibus F-test measuring whether a significant part of the variation of dependant variable can be attributed to the variability among the experimental conditions' means. In the current experiment, the ANOVA would test for a main effect of decision on the story recall. However, testing for an overall effect of decision would not answer my research question. Indeed, my hypotheses target

specific assumptions on pairwise relationships between the three levels of the decision variable – i.e., decision differs from no-decision, and utilitarian differs from deontological. If I were to carry an ANOVA, I should first see whether there is any difference in the mean response across any of the three conditions. Then, if significance is found only, to carry out post-hoc pairwise comparisons. Such analyses would involve four significance tests (one ANOVA and three pairwise comparisons). Among these four tests, only one of them would be directly related to one of my hypotheses - the comparison between the utilitarian and deontological group. A main effect of decision would not directly test my hypotheses as a significant effect could arise from any differences between the three decision groups. Likewise, the hypothesis of a difference in retention when making a decision and no-decision could not be directly tested. Therefore, performing an ANOVA omnibus F-test followed by pairwise comparisons is not the most appropriate analysis for testing my hypotheses and significance tests for focused questions is preferred.

Contrasts analyses are significance tests for focused questions that allow one to evaluate specific predictions (Rosenthal & Rosnow, 1985). A contrast is defined as a linear combination of means whose multipliers or weights ( $\lambda$ ), add to zero (Rosenthal et al., 1999). The  $\lambda$  weights are set based on the research question of interest and must sum to zero ( $\sum\lambda=0$ ). The comparisons may include all condition means or some means only – excluded means will be weighted with  $\lambda_i=0$  (Rosenthal et al., 1999). By reducing the number of comparisons run compared to ANOVA followed by post-hoc multiple comparisons, contrasts analyses reduce the risk of Type-I errors and increase statistical power (Brauer & McClelland, 2005).

Therefore, contrasts analyses were performed to test my hypothesis. The statistical hypotheses and contrasts are presented in Figure 8.

### Figure 8

#### *Statistical Hypotheses and Contrasts*

The tests compare the mean of the following groups:

$\mu_1$ : No-decision

$\mu_2$ : Deontological decision

$\mu_3$ : Utilitarian decision

**Research question 1: Does making any decision as compared to no-decision enhance the recall of the dilemma story?**

$$H_0: (\mu_2 + \mu_3) / 2 - \mu_1 = 0 \text{ or } H_0: -1 * \mu_1 + 1/2 * \mu_2 + 1/2 * \mu_3 = 0$$

$$H_1: (\mu_2 + \mu_3) - \mu_1 > 0$$

Contrast 1:  $\lambda_1 = -1$ ;  $\lambda_2 = +0.5$ ;  $\lambda_3 = +0.5$

**Research question 2 : Does making a utilitarian decision as compared to a deontological decision enhance the recall of the dilemma story?**

$$H_0: \mu_3 - \mu_2 = 0 \text{ or } H_0: 0 * \mu_1 - 1 * \mu_2 + 1 * \mu_3 = 0$$

$$H_1: \mu_3 - \mu_2 > 0$$

Contrast 2:  $\lambda_1 = 0$ ;  $\lambda_2 = -1$ ;  $\lambda_3 = +1$

*Note:* both contrast 1 and 2 are based on a directional alternative hypothesis (greater) so that the p-value will be computed in one direction only – right sided.

Although focused question tests and omnibus tests with post-hoc analyses would rarely lead to different conclusion (see Appendix 1 for a comparison between the ANOVA results with and without contrasts) mean comparisons with focused

questions provides more conceptual clarity and improves statistical power (Rosenthal, Rosnow and Rubin, 1999). Indeed, the planned contrasts focused on my specific hypotheses and halve the number of tests run compared to an ANOVA omnibus F-test with pairwise multiple comparisons.

Statistical analyses were performed in RStudio, R version 4.1.0 (2021-05-18). An R Markdown document reporting the R script and R outputs of the statistical analysis is available in Appendix 1.

### 3.2.3 Results

#### **Inter-rater reliability of the story recall scoring.**

To assess inter-rater reliability of the story recall, two judges scored the participants' answers: one volunteer research assistant and me. The volunteer research assistant was informed of the research's aim. Both judges scored the ten participants' post-decision recall of the 27 items of the story. Two hundred and seventy (27 x 10) items were scored in total. Each item was score as 1 if "recalled", and 0 if "not recalled". The judges' agreement was 91.1%. Cohen's  $\kappa$  reveal an almost perfect agreement  $\kappa=0.82$  (95% CI, 0.75 to 0.89),  $p<0.05$ .

#### **Post-Decision Recall**

##### ***Descriptive statistics***

To verify that decision groups had equivalent initial encoding, recall of the dilemma was measured immediately after the decision was made (i.e., 'post-decision recall'). Descriptive statistics of post-decision recall are available in Table 14.

**Table 14***Descriptive Statistics of The Post-decision Recall for Each Decision Group*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
No-decision	33	12.12	5.84	2.00	13.00	21.00
Deontological	35	13.03	6.03	1.00	14.00	22.00
Utilitarian	32	12.53	4.61	6.00	11.50	22.00

*Note.* The dilemma story recall is scored over 27 ideas. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

### ***Inferential statistics***

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(2,96)=0.75$   $p= .47$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.97$ ,  $p= .04$ . Assumptions for the one-way ANOVA were not met. The Kruskal-Wallis rank-sum test shows no significant difference in post-decision recall between the decision groups,  $\chi^2(2)= 0.66$ ,  $p= .717$ .

## **Digit Span-Forward**

### ***Descriptive statistics***

Descriptive statistics of Digit Span-Forward scores across decision groups are available in Table 15.

**Table 15***Descriptive Statistics of Digit Span-Forward per Decision Groups*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
No-decision	33	9.45	1.64	6.00	10.00	12.00
Deontological	35	10.21	1.74	6.00	10.00	13.00
Utilitarian	32	10.25	1.72	7.00	10.00	14.00

*Note.* The Digit Span-Forward is scored over 16 sequences. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

***Inferential statistics***

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(2,96)=0.02$   $p= .89$ . The Shapiro-Wilk's normality test showed that residuals are normally distributed,  $W= 0.98$ ,  $p= .09$ . The one-way ANOVA revealed no significant differences in mean Digit Span-Forward score between the decision groups,  $F(2, 96)= 2.28$ ,  $p= .11$ .

**Digit Span-Backward*****Descriptive statistics***

Descriptive statistics for Digit Span-Backward scores are available in Table 16.

**Table 16***Descriptive Statistics of The Digit Span-Backward for Each Decision Group*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
No-decision	33.00	9.18	1.69	3.00	9.00	11.00
Deontological	35.00	9.50	1.44	6.00	10.00	11.00
Utilitarian	32.00	9.62	1.13	7.00	10.00	11.00

*Note.* The Digit Span-Backward is scored over 16 sequences. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

### ***Inferential statistics***

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(2,96)=1.26$   $p= .29$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.90$ ,  $p< .001$ . Assumptions of the one-way ANOVA were not met. The Kruskal-Wallis rank-sum test revealed no significant differences found in mean Digit Span-Backward scores between decision groups,  $\chi^2(2)= 1.04$ ,  $p= .60$ .

### **Dilemma story recall after one day**

#### ***Descriptive statistics***

All descriptive statistics for the total number of items correctly recalled after a one-day delay are provided in Table 17.

**Table 17**

*Descriptive Statistics for The Total Number of Items Correctly Recalled After a Day as a Function of Decision*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
No-decision	33	10.84	5.91	0.00	11.00	22.00
Deontological	35	11.11	5.53	0.00	12.00	20.00
Utilitarian	32	10.75	4.40	4.00	10.00	19.00

*Note.* The story recall is scored over 27 ideas. *M*, *SD*, *Min*, and *Max* represent mean, standard deviation, minimum and maximum respectively.

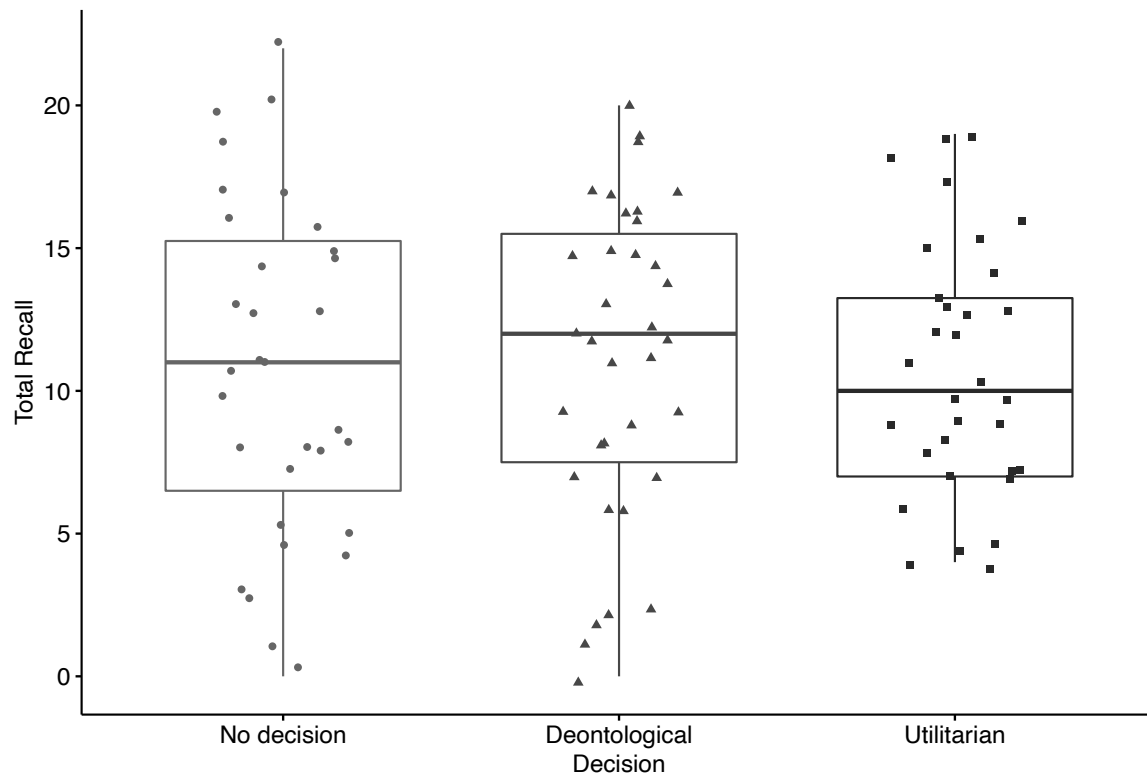
### ***Inferential Statistics***

Contrast analyses were used to test for an effect of decision-making groups on the retention of the dilemma story after one day. The contrast coding is available in Figure 8. Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested homoscedasticity,  $F(2, 96) = 1.50$ ,  $p = .23$ . The Shapiro-Wilk's normality test showed that residuals are normally distributed,  $W = 0.98$ ,  $p = .25$ .

The first contrast revealed no significant differences in mean total recall between making any decision and no-decision,  $t(96) = 0.08$ ,  $p = .47$  (95%CI: -1.81,  $+\infty$ ), right-sided. The second contrast comparing the utilitarian and the deontological group was also not significant,  $t(96) = -0.28$ ,  $p = .61$  (95%CI: -2.53,  $+\infty$ ), one-sided. The distribution of total recall for each experiential condition is illustrated in Figure 9.

**Figure 9**

*Retention of The Dilemma Story After One Day Per Decision Group*



*Note:* the boxplots present for each experimental group: data points, the median, the upper (Q3) and lower (Q1) quartiles, the interquartile range (IQR), the highest (Q3 + 1.5\*IQR) and lowest (Q1-1.5\*IQR) when excluding potential outliers.

## **Exploratory analysis**

### ***Ideas Recall Analysis***

Considering the absence of an effect of decision-making on the mean number of items recalled after one day, I examined my data using a more qualitative approach. While there is no overall difference between the quantity of ideas recalled by each decision group, could the actual remembered items differ? In other words, after

examining the amount of information that each group remembered regarding the dilemma story, further consideration should be given to the items that each group actually remembered. While I do not have the power to carry out that many statistical comparisons across the three groups (i.e., one for each ideas so 27 comparisons total), a descriptive analysis can be performed.

Figure 10

*Ideas Delayed recall Occurrences for Each Decision Group.*

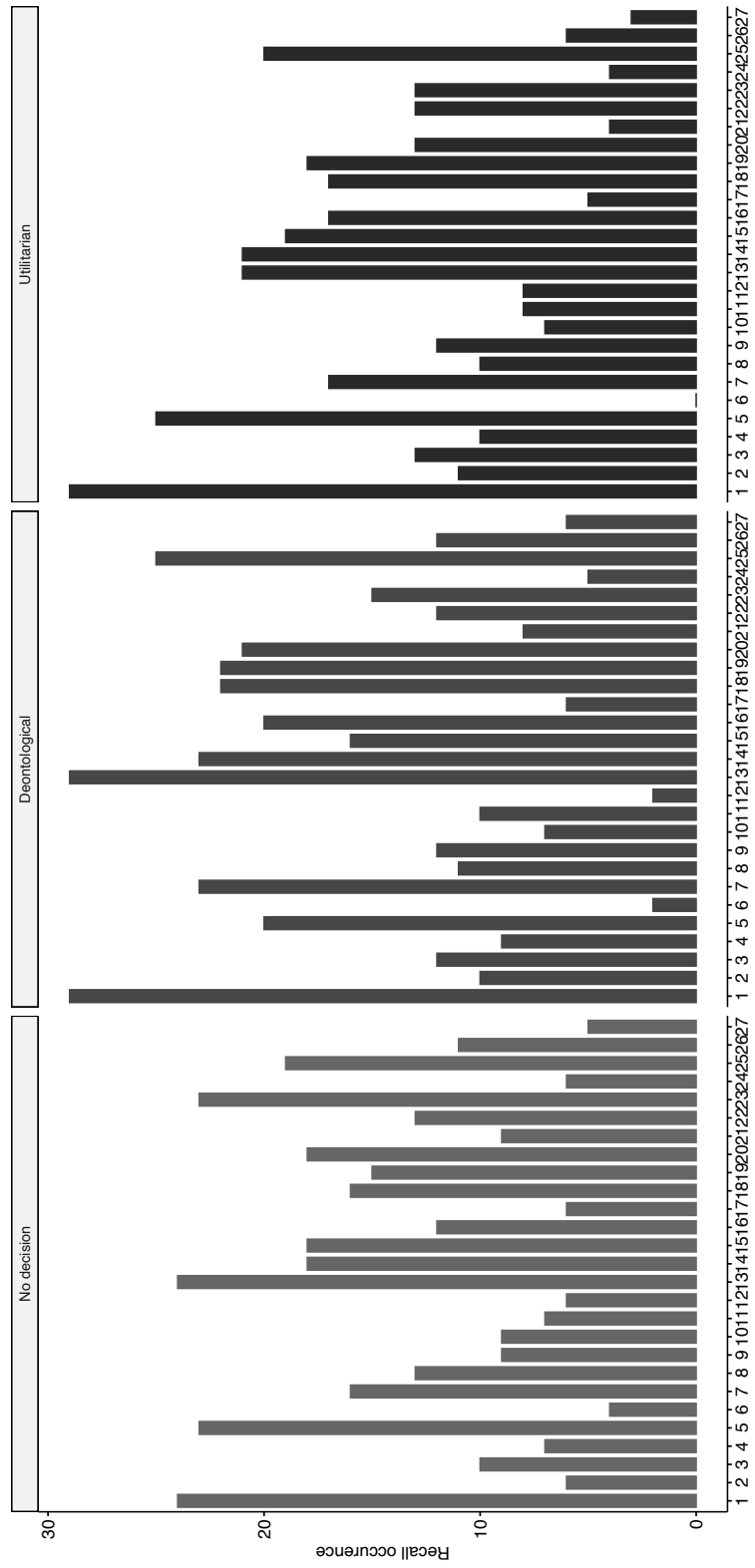


Figure 10 illustrates the recall occurrence of each idea of the story per decision. Visual inspection of Figure 10 suggests that all groups show similar item recall patterns for the dilemma story.

### ***Forgetting Rate Analysis***

In Murty et al. (2019), forgetting is estimated through a proportional difference in recognition memory across tests. For a more adequate comparison of my results to Murty et al. (2019), I computed forgetting rates based on Murty et al.'s (2019) ratio: (immediate recall score – delayed recall score)/ immediate recall score.

Descriptive statistics, diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(2,94) = 1.25$   $p = .29$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W = 0.82$ ,  $p < .001$ . Assumptions for the one-way ANOVA were not met. The Kruskal-Wallis rank-sum test shows no significant difference in post-decision recall between the decision groups,  $\chi^2(2) = 0.73$ ,  $p = .70$ .

### 3.2.4 Discussion

This study aimed to investigate the effect of decision-making on memory. To this end, I used a unique experimental design combining long-term episodic memory retention and moral dilemma paradigms. After listening to a dilemma story, two-thirds of the participants were asked to decide on one of the dilemma options (utilitarian or deontological). Decision-making was not induced for the remaining third of the participants who answered an unrelated question instead (no-decision). The results

showed similar retention of the dilemma story after one day across all three experimental conditions.

It was somewhat surprising to not find any differences in long-term episodic memory retention after making vs. not making a decision. It contradicts a long-running assumption about the relationship between decision-making and memory. Traditionally, it is believed that decision-making and memory have a reciprocal influence on one another (Fellows, 2018; Tremel et al., 2018). However, while the effect of memory on decision-making has been intensively studied (Redish & Mizumori, 2015, Madan et al., 2014; Johnson & Busemeyer, 2010; Dougherty et al., 2003), there is a scarcity of empirical evidence on whether decision-making impacts memory. To my knowledge, only Murty et al. (2019) investigated this direction of the effect. Yet, my experiment did not support their findings.

Observing similar story retention after making a utilitarian or deontological decision was also against my predictions. The dual-process model of decision-making (Greene et al., 2001, 2004, 2008) suggests that making a utilitarian involves a deeper level of cognitive control. Hence, making a utilitarian decision should enhance the retention of the story. However, my experiment did not show any differences in retention as a function of moral decision types.

Before discussing the implications of my results, one may question their reliability. Thus, I will remind the reader of the measures taken to optimise statistical power. To reduce type-II error – i.e., failing to reject a false null hypothesis, I followed the  $n > 30$  rule of thumbs (Roscoe, 1975). The rule state that a sample of at least 30 observations per group is adequate for most behavioural research (Memone et al.,

2020). This rule is based on the central limit theorem, which suggests that a sample mean approaches the population mean as the sample increases. To reduce the risk of type-I error – i.e., rejecting a true null hypothesis, I conducted planned, focus hypotheses testing with contrasts analyses (Brauer & McClelland, 2005). Finally, one-sided hypotheses testing was used, which also improves statistical power. I was able to do so as theories on decision-making and memory make explicit predictions on the expected direction of the effect. Although I cannot rule out that the absence of effect may be due to a type-II error, I acknowledge that measures were taken to reduce this likelihood.

If current findings are not due to statistical flaws (e.g., lack of power), could they be due to experimental glitches? It could be argued that the control condition (i.e., answering an unrelated question instead of the dilemma) lacks validity. Indeed, examples of moral dilemmas are well-known to the public. Participants might anticipate the task and engage in decision-making independently of the instructions provided. In other words, simple exposure to the dilemma likely initiates the process of decision-making, even when not explicitly prompted. In such a case, decision-making would be engaged in the control condition, making it not valid in controlling for the effect of decision-making.

The limitation discussed above might explain the absence of difference in retention between the no-decision and the decision conditions. However, it cannot account for the absence of difference between moral decision types. This specific aspect of the null effect could be due to an inadequacy of my measurement of retention. In line with Murty et al. (2019), I have measured retention after a one-day

delay. However, such a stand-alone measurement is not adequate to capture the essence of forgetting. Considering forgetting as a loss of information over time (Roediger et al., 2010) is probably the most consensual approach to studying forgetting. This approach offers an empirical measurement of forgetting using forgetting curves derived from plotted retention at different time intervals. One must note that the common practice is to call these forgetting curves, but they are actually retention curves (Roediger et al., 2010). Although the terms forgetting and retention tend to be used interchangeably, the former refers to the process, whereas the latter is a way to measure the process. Traditionally, forgetting curves are presented as an exponential decrease of retention over time (see Rubin & Wenzel, 1996, for a review). A design with only two measurement times fails to account for the curvilinearity of forgetting. However, the exponential nature of forgetting is not the only reason that there is a need for at least three delay intervals. Indeed, in the previous experiment, retention is supposedly driven by two parameters: decision-making and time - in a forgetting paradigm, time is an inherent predictor. For any empirical hypothesis to be refutable, an experimental design must include  $N+1$  measurement times for each independent variable or predictor (Bogartz, 1990). Therefore, at least three delay intervals are required to investigate whether decision-making impacts long-term forgetting.

Let us now consider that my results are genuine – not due to statistical or experimental issues. How to explain the divergence from Murty et al. (2019)? Explanations for our conflicting results may lie in a critical difference between the two experimental paradigms. To appreciate how Murty et al.'s (2019) paradigm

substantially differs from mine, we must consider the approach of decision-making adopted in this thesis. Decision-making is the process of judging, preferring and choosing one option across several possibilities. I used a moral dilemma to enact decision-making as it fully fits this definition. Contrastingly, Murty et al.'s task (2019) matches only one aspect of the definition above: it involves choosing between two options. In Murty et al.'s experiments, participants chose between two occluder screens. By selecting an occluder screen, the participant revealed the item to be remembered. However, their choice had no impact on the item that was revealed. In other words, participants encoded the same item regardless of the occluder screen they chose.

Moreover, the authors indicated that part of their data was previously reported in Murty et al.'s (2015). In their 2015 paper, Murty et al. refer to their experimental design as "a simple act of choosing". They suggest that perceived agency influences decision-making, but they do not directly refer to their paradigm as a decision-making one. However, when they used the same paradigm again in Murty et al. 2019, they presented it as a decision-making paradigm. It is unclear why the same paradigm is granted different concepts from one experiment to another. This might suggest a change in the authors' perspective on the definition of decision-making. Still, they seem to acknowledge the debatable validity of their paradigm when qualifying it as "arbitrary decision-making". However, I argue that decision-making cannot be arbitrary since it involves judging between decision outcomes. Hence, I believe Murty et al. (2015, 2019) have more likely provided evidence of the impact of perceived agency on memory rather than decision-making.

In summary, the present experiment suggests that decision-making in moral dilemmas does not impact memory retention. The discussion highlights two potential limitations in the reliability of the results. The absence of difference between the no-decision and the decision conditions could be due to a failure to control the decision-making process's elicitation. This issue would be intrinsic to moral dilemma paradigms and cannot be directly addressed. Yet, such an issue cannot explain the absence of difference in retention between the utilitarian and the deontological conditions. An inadequate forgetting paradigm may alternatively cause this. In Experiment 5, I address the latter limitation by investigating differences in retention between utilitarian and deontological decisions in a three delay-intervals forgetting paradigm.

### **3.3 Experiment 5: The Impact of Moral Decision Types on Long-Term Forgetting Up to One Month Delay**

The previous experiment aimed to conceptually replicate Murty et al.'s (2019) findings on the impact of decision-making on forgetting. To this end, I created a novel experimental paradigm combining decision-making and memory retention. I used a moral dilemma story to elicit both decision-making and memory recall. Participants were either asked to decide on one of the two dilemma options (deontological vs utilitarian) or answer an unrelated question (control condition). Then, the retention of the story was measured with incidental free recall at one day delay. The results did not provide evidence for the impact of decision-making or moral decision types on retention after a one-day delay. Firstly, deciding on the dilemma did not influence retention. Secondly, similar retention patterns of the story were observed regardless of the decision condition. These results dissent from previous empirical findings (Murty

et al., 2019) and theories on the relationship between decision-making and memory (Fellows et al., 2018; Tremel et al., 2018).

Two main limitations must be considered as potential explanations for these antithetical results. First, the results could be due to the decision-making paradigm. It could be argued that the control condition (i.e., answering an unrelated question instead of the dilemma) lacks validity. Indeed, examples of moral dilemmas are well-known to the public. Participants might anticipate the task and engage in decision-making independently of the instructions provided. In other words, simple exposure to the dilemma is likely sufficient to initiate the process of decision-making, even when not explicitly prompted. In such a case, the control condition would defeat its purpose. Nevertheless, the suggestion above does not account for all my results. Indeed, it does not seem fair to assume that utilitarian and deontological decisions arise from the same process. The utilitarian and deontological decision groups should have necessarily engaged in different decision-making processes. Still, the two groups observe similar patterns of retention. Thus, an alternative account for the absence of an effect must be considered.

Perhaps the null results are due to a limitation in the retention paradigm. Although I used the same retention paradigm as Murty et al. (2019), this paradigm might lack validity. Indeed, forgetting is a curvilinear process (Rubin & Wenzel, 1990) influenced by time. Thus, multiple delay intervals, at least three (Bogartz, 1990), are required to observe a difference in forgetting over time. To address these limitations, the present experiment examines retention over three delay intervals as a function of moral decision types.

To summarise, the absence of effects reported in Experiment 4 could be due to limitations in the decision-making or the retention paradigms adopted. The current experiment is a partial replication attempting to address these limitations. Since I cannot guarantee a condition where decision-making is not engaged, it seems preferable to focus on the effect of moral decision types—differences between utilitarian and deontological decisions. Furthermore, I measure retention at three delay intervals - one day, one week and one month. Overall, two delay intervals are added to the retention paradigm, and one decision-making condition is dropped.

The benefits of this new design are twofold. First, removing one decision-making condition reduces the number of comparisons as the measurement time increases. Avoiding numerous comparisons is preferable when considering statistical power. Additionally, the computation of retention curves allows me to replicate and expand the previous findings. Overall, Experiment 5 aims to partially replicate and further investigate the results obtained in Experiment 4.

### 3.3.1 Method

#### **Participants**

136 participants were recruited via the Department of Psychology participant pool and the general public using word-of-mouth and online advertisement. Volunteers who self-reported memory difficulties due to neurological or neurodevelopmental conditions were excluded from the experiment. The demographic characteristics of the sample are provided in Table 18. The study was approved by the University of Edinburgh - School of Philosophy, Psychology and Language Sciences Ethics

Committee (No 24-1920/11). Participants gave informed consent in accordance with the British Psychological Society Code of Ethics and Conduct. Participants received course credits or were compensated with a small honorarium for their participation.

**Table 18**

*Socio-Demographic Characteristics of The Sample from Experiment 5*

<i>N</i>	Age					Education					Gender	
						1= Undergraduate					2= Postgraduate	
	<i>M</i>	<i>SD</i>	Min	Median	Max	1	2	3	4	NA	1	2
136	25.60	8.37	17	24	58	67	42	23	3	1	105	31

*Note.* *N*, *M*, *SD*, Min, Max and NA refer to sample size, mean, standard deviation, minimum, maximum and missing data, respectively. Missing data are either due to a non-classifiable input or the item not being answered.

## Materials

The administration of the tasks was pen-and-paper based. By doing so, I tried to get more insight into participants' engagement with the decision-making task. I expected that participants would be less likely to give a random answer if the experimenter asked the dilemma question rather than presenting it on a computer. More engagement with the task might increase its validity.

### ***Digit Span-Forward and Digit Span-Backward***

The administration and scoring of Digit Span-Forward and Digit Span-Backward subtests of the WAIS-IV (Wechsler, 2008) mainly followed Experiment 4.

The following adjustments were made to fit the paper-and-pencil administration. The experimenter read the digit sequence to the participants. Then, the participant had to recall the digit verbally.

In both experiments, the Digit Span-Forward involved recalling the digits in the same order as read by the experimenter. In the Digit Span-Backward, the digits had to be recalled in reverse order. Each subtest was composed of eight items, resulting in sixteen items for the digit span subtest. Each digit sequence was scored 0 if failed and 1 if correctly recalled. A subtest ended when both sequences from the same item were failed.

### ***The Moral Dilemma Task.***

The dilemma story used in both experiments is a modified version of the Nobel Prize dilemma (Moore et al., 2008). The story content is available in Appendix 5. The administration and scoring of the moral dilemma task mainly followed Experiment 4. Few adjustments were made. The reading of the dilemma story was done using the audio recording from Experiment 4. Since the task administration was paper-and-pencil based, the experimenter directly asked the dilemma question (i.e., “Do you let your colleague sell the discovery or do you release the gas?”). Participants answered the experimenter verbally. The scoring form of the dilemma story recall is provided in Appendix 6. Results from the previous experiments suggested the scoring of the story recall had high inter-rater reliability. Therefore, it was adequate to use the same materials and procedure.

## **Design**

A between-subjects design with two independent variables was used. The moral decision type independent variable had two levels: deontological and utilitarian. The delay interval independent variable had three levels: one day, one week, and one month. The dependent variable was the number of ideas correctly recalled from the dilemma story.

## **Procedure**

### ***Initial testing session***

The experiment started with the Digit Span-Forward and the Digit Span-Backward. The digit span tasks were followed by the Moral Dilemma task. The instructions provided were the following: "You are going to listen to a short story. Listen carefully and try to remember as much as you can". Then, participants listened to the audio recording of the Nobel Prize dilemma story. Following the reading of the dilemma story, participants were asked: "What are you going to do?" "Do you release the gas?" or "Do you let your colleague sell the discovery?". The presentation order of the decision options was randomised to avoid order effects. This was followed by the recall of the dilemma story. The following instructions were provided: "Can you say everything you can remember about the story? You can say things even if you are not sure about the details or order". The post-decision recall of the story was recorded and saved under an anonymised participant identification number. The first session ended with the collection of socio-demographic details, and an appointment was given for the follow-up session. As in Experiment 4, the delayed recall was incidental, so

participants were not informed that they would have to recall the dilemma story. This was done to avoid intentional rehearsal of the story.

### ***Follow-up session***

For delayed recall, participants were randomly assigned to one of the three delay intervals - one day, one week or one month. Follow up sessions were conducted in person, in the same testing room as for the first session. The instruction for delayed recall was “Can you tell me everything you remember about the story you heard yesterday/ last week/ last month? Please recall everything you can, even if you are not sure.” No time limit was set for the participant to answer. The recall task ended when the participants informed the experimenter that they could not recall more. The delayed recall of the story was recorded and saved under the participant identification number used in the first session. The follow-up session ended with a debriefing. The purpose of the experiment was explained, and participants’ questions were answered.

### 3.3.2 Statistical Analyses

The analysis script and a package containing the anonymised dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdbf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdbf7ec8b03ea1ad66).

Nine participants were not included in the analysis as they did not attend the follow-up session. Hence, 127 participants were included in the analysis.

T-tests for independent samples were used to verify whether the two decision groups differ in post-decision recall of the dilemma story or in Digit Span-Forward and Digit Span-Backward scores. Mann-Whitney-Wilcoxon tests were used when assumptions for the t-test were not met.

Experiment 4 did not find an effect of decision groups on the recall of the dilemma story. Considering the absence of effect in Experiment 4, using an omnibus statistical approach to detect any possible effects seems appropriate. Assumptions were not met for the most common types of analyses - mixed-model ANOVA or multiple linear regression models with t-test on the regression coefficients (see Appendix 1). Failing to meet assumptions for parametric tests impacts the reliability of p values and confidence intervals (CI). Thus, I have applied a bootstrapping method a multiple regression model as recommended in Little (2013). Bootstrapping is a method for determining the significance of the sample results. Bootstrapping is achieved by resampling with replacement from the original sample. This method does not rely on assumptions about the population distribution like  $t$  or  $F$  families (Erceg-Hurn et al., 2013). Significance in mean difference is considered by interpreting the 95% confidence intervals – the result is significant at a 5% confidence level when the 95%CI for slope estimates do not include 0. 10'000 bootstrap samples were computed as suggested in Hesterberg (2015).

Akaike information criterion (AIC) and Bayesian information Criterion (BIC) model selection were used to distinguish between models describing the relationship between delay, decision-making and the dilemma story recall. Three models were compared. A specification of the three models at test is provided in Table 19.

**Table 19***Experiment 5 Models' Specification*

<p>Model 0 (M0): Null model</p>	$M0: ItemRecall = \beta_0 + \beta_1 * (DelayWeek)$ $+ \beta_2 * (DelayMonth) + \epsilon$ <p><i>Where one day is the reference level.</i></p>
<p>Model 1 (M1): Alternative model</p>	$M1: ItemRecall = \beta_0 + \beta_1 * (DelayWeek)$ $+ \beta_2 * (DelayMonth)$ $+ \beta_3 * (DecisionUtilitarian) + \epsilon$ <p><i>Where one day is the reference level for the Delay variable,</i> <i>And deontological is the reference level for the Decision variable.</i></p>
<p>Model 2 (M2): Interaction model</p>	$M2: ItemRecall = \beta_0 + \beta_1 * (DelayWeek)$ $+ \beta_2 * (DelayMonth)$ $+ \beta_3 * (DecisionUtilitarian)$ $+ \beta_4 * (DelayWeek) * (DecisionUtilitarian)$ $+ \beta_4 * (DelayMonth) * (DecisionUtilitarian) + \epsilon$ <p><i>Where one day is the reference level for the Delay variable,</i> <i>And deontological is the reference level for the Decision variable.</i></p>

Note: According to the null model, the number of items correctly recalled is predicted by the delay condition only. The effect of delay on memory retention is known and assumed from the literature. Therefore, it was included in the null model and was not the target of our comparisons.

### 3.3.3 Results

#### Post-Decision Recall

To ensure that moral decision types did not impact initial encoding, the recall of the dilemma was assessed immediately after the decision was made (i.e., 'post-decision recall').

#### *Descriptive statistics*

Descriptive statistics of post-decision recall for each moral decision type is provided in Table 20.

**Table 20**

*Descriptive Statistics of Post-decision Recall for Each Decision Type*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Deontological	64	12.42	4.96	1.00	13.00	21.00
Utilitarian	72	14.04	4.80	0.00	15.00	24.00

*Note.* The dilemma story recall is scored over 27 ideas. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

#### *Inferential statistics*

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(1,134) = 0.39, p = .54$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W = 0.98, p = .02$ . Assumptions for the t-test were not met. The

Mann-Whitney-Wilcoxon test revealed no significant difference in post-decision recall between the utilitarian and the deontological decision type,  $W= 1858.5$ ,  $p= .05$ .

## Digit Span-Forward

### *Descriptive statistics*

Descriptive statistics of Digit Span-Forward for each moral decision type is provided in Table 21

**Table 21**

*Descriptive Statistics of Digit Span-Forward for Each Decision Type*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Deontological	64	10.19	2.32	4.00	10.00	15.00
Utilitarian	72	9.99	2.09	6.00	10.00	15.00

*Note.* The Digit Span-Forward is scored over 16 sequences. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

### *Inferential statistics*

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(1,134)= 0.58$ ,  $p= .45$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.97$ ,  $p= .01$ . Assumptions for the t-test were not met. The Mann-Whitney-Wilcoxon test revealed no significant difference in post-decision recall between the utilitarian and the deontological decision type,  $W= 2408.5$ ,  $p= .65$ .

## Digit Span-Backward

### *Descriptive statistics*

Descriptive statistics of Digit Span-Backward for each moral decision type is provided in Table 22.

**Table 22**

*Descriptive Statistics of Digit Span-Backward for Each Decision Type*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Deontological	64	9.12	2.25	2.00	9.00	16.00
Utilitarian	72.00	8.92	2.26	5.00	9.00	15.00

*Note.* The Digit Span-Backward is scored over 16 sequences. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

### *Inferential statistics*

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(1,134) = 0.25$ ,  $p = .62$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W = 0.95$ ,  $p < .001$ . Assumptions for the t-test were not met. The Mann-Whitney-Wilcoxon test revealed no significant difference in post-decision recall between the utilitarian and the deontological decision type,  $W = 2507$ ,  $p = .37$ .

## Dilemma story delayed recall

### *Descriptive statistics*

All descriptive statistics for the total number of items correctly recalled at delay for each time intervals and decision type are available in Table 23.

**Table 23**

*Descriptive Statistics for The Total Number of Items Correctly Recalled as a Function of Moral Decision Type and Delay*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Deontological						
Day	21	10.14	4.96	0.00	12.00	18.00
Week	21	8.52	4.76	1.00	9.00	16.00
Month	20	4.55	5.70	0.00	3.00	19.00
Utilitarian						
Day	25	11.64	3.98	0.00	12.00	19.00
Week	20	9.85	2.87	4.00	10.00	16.00
Month	20	5.00	5.65	0.00	2.50	16.00

*Note.* The recall of the dilemma story is scored over 27 ideas. *N*, *M*, *SD*, *Min*, and *Max* represent sample size, mean, standard deviation, minimum and maximum respectively.

### ***Inferential statistics***

AIC and BIC were used to distinguished between the three models describing the relationship between delays, decision-making and the story recall. The decision model had lower AIC (601.31) and BIC (615.53) than the interaction model (AIC= 605.02, BIC= 624.94). Overall, the null model had lowest AIC (601.25) and BIC (612.63).

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested homoscedasticity,  $F(2, 125) = 2.96$   $p = .08$ . The Shapiro-Wilk's normality test showed that residuals are normally distributed,  $W = 0.96$ ,  $p = .001$ .

The bootstrapped regression model suggests a significant difference in story recall between the one-day and one-month delay groups. All other comparisons were not significant (see Table 24). Results are illustrated in Figure 11.

**Table 24**

*Bootstrapped Regression Model of Story Recall Predicted by Moral Decision Types and Delay Intervals.*

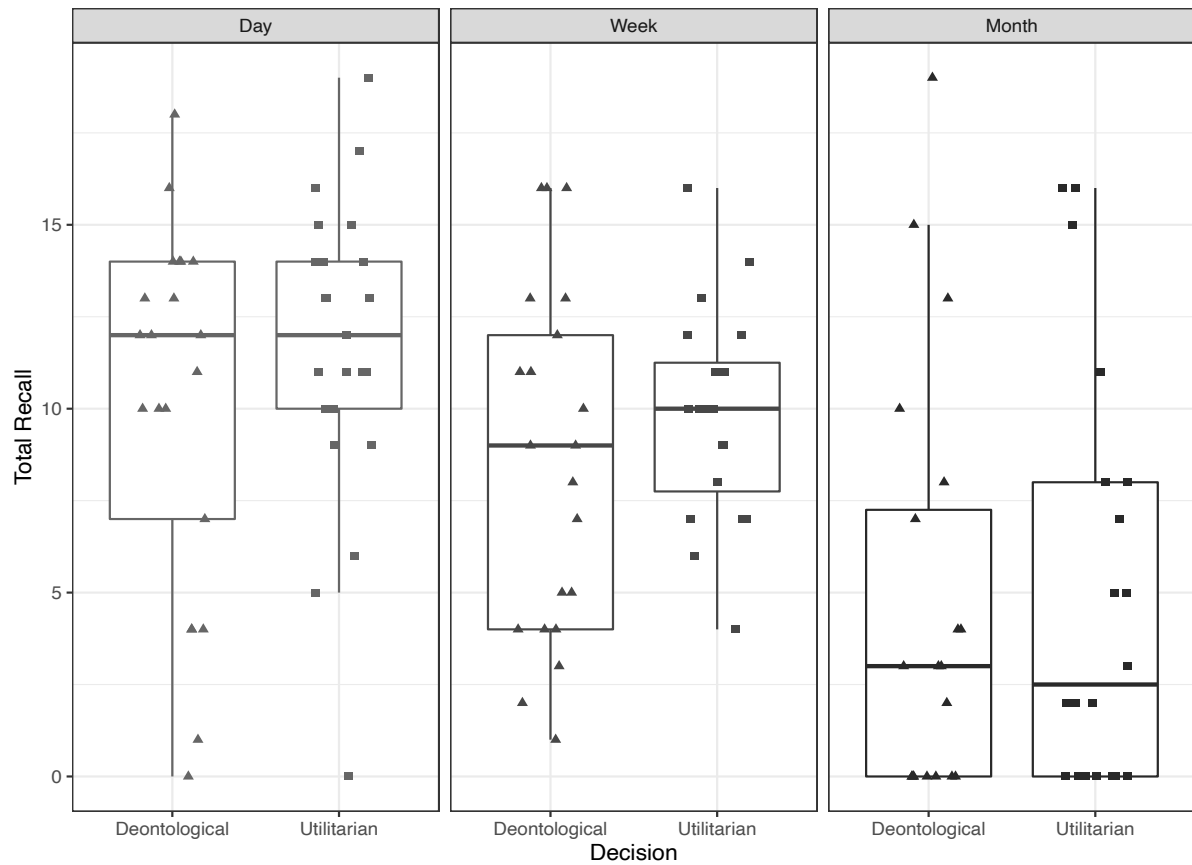
Parameter	Average bootstrap estimate	95%CI	
		LL	UL
Intercept: Day	10.14	7.95	12.18
Week	-1.62	-4.57	1.36

Parameter	Average bootstrap estimate	95%CI	
		LL	UL
Month	-5.59	-8.72	-2.16
Utilitarian	1.50	-1.12	4.15
Week:Utilitarian	-0.17	-3.75	3.43
Month:Utilitarian	-1.05	-5.49	3.33

*Note.* CI, LL and UL, respectively stands for confidence interval, lower limit and upper limit.

**Figure 11**

*Total of Items Correctly Recalled Per Delay and Decision Group*



Note: the boxplots present for each experimental group: data points, the median, the upper (Q3) and lower (Q1) quartiles, the interquartile range (IQR), the highest (Q3 + 1.5\*IQR) and lowest (Q1-1.5\*IQR) when excluding potential outliers.

### ***Exploratory analysis***

#### Item recall analysis

Such as it was done for in Experiment 4, I have conducted a visual exploration of item recall occurrences. Figure 12 illustrates the recall occurrence of each item per

decision. The bar charts show that decision groups have similar recall patterns of the items of the dilemma story and these patterns seems consistent over time.



### 3.3.4 Discussion

Experiment 5 aimed to investigate the moral decision types on memory further. In this new experiment, I focused on comparing the retention of the dilemma story following a deontological or a utilitarian decision. The retention of the dilemma was measured after a one-day, one-week or one-month delayed recall. Including three delay intervals in the study design allowed me to observe the forgetting curves of the dilemma story. Following the dual-process model of decision-making (Greene et al., 2001, 2004, 2008), I hypothesised that making a utilitarian decision would enhance long-term retention compared to a deontological decision. Experiments 4 and 5 suggest the opposite, that is, retention is not impacted by the type of decision made. Firstly, I will consider the potential limitations of these results. Secondly, I will discuss the implications of these results on a theoretical level.

In Experiment 4, I have highlighted limitations in the decision-making and the forgetting paradigms. In the present experiment, both paradigms have been adapted to reach an adequate compromise. The no-decision condition was abandoned as listening to a dilemma story may inevitably engage a process of decision-making. Therefore, I focused on the two conditions inherently resulting from different processes of decision-making – the deontological and utilitarian decisions. Removing one condition from the decision-making variable prevented a drastic increase in comparisons across groups since the forgetting paradigm was improved by adding two delayed recall intervals. Delayed recalls were added at one-week and one-month delay. In sum, the consequences of adding two conditions to the delay variable were attenuated by removing one condition from the decision variable.

Nevertheless, adjusting the forgetting paradigm brought some challenges. The addition of two delay conditions doubled the number of experimental groups – three groups in Experiment 4 against six in Experiment 5. Although I did increase my sample size accordingly, I could not meet the  $n > 30$  rule of thumbs (Roscoe, 1975). As participants were randomly allocated toward the three delay conditions, potential volunteers were refrained by the prospect of a one-month commitment.

The inferential statistics included a bootstrapped regression model. This omnibus approach (i.e., detecting any broad range of departures from the null) is statistically more lenient than the focused, planned, one-tailed comparisons used in Experiment 4. Compared to a focused approach, an omnibus approach reduces the risk of type-II error – i.e., failing to reject a false null hypothesis.

Although such a conclusion only partially answers the research question, it does encourage one to reconsider the relationship between decision-making and memory. Facing the paucity of empirical investigations on the impact of decision-making on memory, I extended my literature review to include the literature on judgment and memory, which provides potential explanations for my results. First, I must clarify how decision-making and judgement differ. On the one hand, decision-making refers to the entire process of choosing a course of action (Hastie, 2001). Hence, decision-making entails actions (Krueger, 2011). On the other hand, judgments refer to a step in the decision-making process where the outcomes of possible options are estimated (Fischhoff & Broomell, 2020). Judgments both precede and enable decisions (Krueger, 2011). This distinction is consensual in the literature on judgment and decision-making.

Unlike the decision-making literature, the judgment literature is less categorical and more pragmatic when considering the relationship between judgments and memory. Hastie and Park (1986) considered various explanatory models on the relationship between judgments and memory. The authors characterised explanatory models according to two criteria. The first criterion is the direction of the relationship between judgment and memory. Both processes can be assumed to be either independent or unidirectionally related – memory causes judgment and vice versa. The second criterion is the stage of memory involved in the relationship – encoding vs. retrieval. Because each model has received supportive empirical evidence, the relationship between judgments and memory cannot be unequivocally characterised.

Hastie and Park (1986) suggested that divergence in the experimental tasks adopted can explain the heterogeneity of models and results on the relationship between judgment and memory. According to the authors, a key distinction in the experimental conditions is whether the judgement task is memory-based or online. They consider a task memory-based when judgments necessarily require retrieving concrete evidence from long-term episodic memory. The authors suggest that a memory-based judgment can be produced by surprising the subject with a 'novel' judgement, i.e., a judgement that is not expected nor anticipated. Memory-based judgment tasks are opposed to online judgment tasks where the judgment and the information related to that judgement are concomitant. From this distinction, the authors hypothesised that observing a relationship between judgments and memory depends on whether the judgment task is online or memory-based. In a set of four experiments, Hastie and Park (1986) measured the relationship between memory and

judgement using a judgments rating task (on a 10-point scale) and a recall task (proportion of information correctly recalled). Their results showed a relationship between judgment rating and memory recall only when two conditions were met. Firstly, subjects must be forced to rely on evidence stored in long-term episodic memory to make their judgment. Secondly, no previous related judgment should be available in long-term episodic memory. Failure to meet these two conditions would lead to the absence of a relationship between judgment and memory.

Park and Hasite's research (1986) suggest that the relationship between memory and judgment is not unequivocal but judgment-context dependent. Could this assumption be valid for the entire decision-making process? My results revealed no impact of moral decision types on long-term episodic memory. Could this finding be attributed to the decision-making paradigm involved? Let us assume that the online vs. memory-based dichotomy pertains to decision-making tasks. One of the reasons I chose a moral dilemma paradigm was to guarantee that participants would not have made any similar decisions in the past. In the case of the Nobel Prize dilemma, it is unlikely that any participants have ever decided whether to cause the death of a colleague. Also, the experiment was presented as a memory task. The decision-making component was not presented to the participants in the briefing. Thus, the design was meant to be memory-based: participants could not recall previous similar decisions, and they should process the content of the dilemma story to reach a decision.

However, the decision-making task might not be as memory-based as intended. Indeed, in my research paradigm, participants must decide on the dilemma

immediately after listening to it. A decision is made while the dilemma story is still processed in working memory. Furthermore, the popularity of moral dilemmas may provide participants with previous knowledge on their inclination toward utilitarian or deontological decisions. In such a case, my research paradigm would fail to meet the two conditions required to observe an impact of decision-making on memory. Consequently, it is unclear where the current paradigm stands within the online vs memory-based dichotomy suggested by Hastie and Park (1986).

### **3.4 Chapter 3 Discussion**

This chapter reports Experiments 4 and 5 investigating the impact of decision-making on long-term episodic memory. In Experiment 4, I studied the impact of decision-making and moral decision types on the retention of a moral dilemma story. The results showed that making a decision on the dilemma did not impact its retention. The results also suggested that making different moral decisions lead to similar retention of the dilemma story. This second finding was further investigated in Experiment 5, where retention was measured up to a one-month delay. Experiment 5 showed that the forgetting of the dilemma story was not impacted by the moral decision made. Altogether, the two experiments encourage reconsidering the relationship between decision-making and memory.

The present findings contradict several aspects of the literature on decision-making and memory. Firstly, one of my hypothesis was based on the dual-process theory of decision-making (Greene et al., 2004, 2008). According to this theory, utilitarian decisions involve a deeper level of cognitive processes. Considering that a deeper level of processing leads to better retention ( Craik & Lockhart, 1972), I

assumed that making a utilitarian decision should lead to greater retention of the dilemma story. Yet, the absence of difference in retention following a deontological or utilitarian decision suggests that both decisions involve a similar level of processing. One may argue that deontological decision-making is guided by a greater emotional response compensating for reduced controlled cognitive processing. Indeed, research has shown that emotional arousal improves memory retention (for a review, see Reisberg & Heuer, 2004). However, this argument would also contradict the dual-process theory. Greene et al. (2004) suggest that the greater level of processing assumed in utilitarian decision-making includes emotional processes. In other words, both deontological and utilitarian decision-making involves emotional processing. However, the earlier solely involves an automatic emotional response, and the latter involves more controlled emotional and non-emotional cognitive processes. Thus, the dual-process theory does not present a difference in emotional vs cold processing but rather in the magnitude of cognitive processing.

Furthermore, these findings differ from published results on the impact of decision-making on forgetting (Murty et al., 2019). However, Murty et al.'s (2019) and the current research vary in two critical aspects. First, I will briefly remind the reader of the distinctions between the decision-making tasks adopted. Then, I will discuss the differences in the memory paradigms and further develop their impact on the findings.

Murty et al. (2019) elicited decision-making with a perceived choice task. The authors considered the choice as 'perceived' since participants were seemingly asked to choose between two options, but both entailed the same outcome. Such a task does not fit my definition of decision-making in two ways. Firstly, choice-making is not

equivalent to decision-making but is a component of the decision-making process. Secondly, decision-making entails judgment, that is, the consideration of the outcomes associated with decision options. If the options do not differ in outcomes, no judgment is required to make a choice. In contrast, I have specifically chosen a moral dilemma task as it fits my definition of decision-making in eliciting judgment, preference, and choice. Overall, the two decision-making tasks fundamentally differ.

My experiments also vary from Murty et al. (2019) in the memory paradigm adopted. While I measured retention using a free incidental recall task, Murty et al. (2019) measured retention using an intentional recognition memory task. In my experiments, participants had to tell me everything they could remember about the story. No feedback nor cues were provided. Also, the delayed recall task was incidental. Participants were not informed that they would have to recall the story during the follow-up session. The aim was to avoid intentional rehearsal and additional exposure to the study materials. In contrast, Murty et al.'s (2019) recognition task presents participants with two items - one old and one new. Participants had to recognise the item they had already seen. Using a recognition task to measure immediate retention enables further exposure to the materials between initial encoding and delayed recognition. By exposing participants to the study materials again, they are offered an additional opportunity to learn them. Furthermore, Murty et al. (2019) do not say whether participants were informed of the delayed recognition task. Since the authors do not mention deception in their research protocol nor any tasks that do not involve the targeted materials, participants were likely aware of the delayed recall task. Informing participants of the delayed recall task may encourage them to apply

learning techniques such as intentionally retrieving the study materials between delay intervals. Thus, it cannot be guaranteed that perceived agency has directly enhanced retention. It could also be the case that perceived agency had impacted participants' learning strategies. Such an effect would enhance long-term retention in a way that could not be distinguished from the effect of the perceived agency. Thereof, Murty et al.'s (2019) forgetting paradigm may encompass confounding effects of learning.

The criticism formulated above brings me to consider the role of learning in the relationship between decision-making and memory. The literature on decision-making assumes a strong relationship with learning. This relationship is explained in detail by Fellows (2018). According to the author, the relationship between decision-making and memory is circular - our decisions influence our memory representations, and these representations will give later decisions. In other words, the outcome of the decision made will be encoded in memory to guide the predictions of outcomes for future decisions. These iterations between decision-making and memory are considered an essential aspect of associative learning.

Patient studies have provided further evidence of the role of learning in decision-making. Research in patients with VMPFC and orbitofrontal damage shows that decision-making impairments stem from a deficit in reversal learning skills (Fellows & Farah, 2005). In 1994, Damasio described a patient's impairment in decision-making skills following vmPFC damage as unable to "learn from his mistake". Damasio also refers to decision-making impairment as myopia for the future. This terminology was used to describe the patient's condition where, despite having experienced the outcomes of poor decisions, they could not associate these decisions

with their adverse outcomes. Evidence on the relationship between decision-making and reversal learning are reviewed in Clark et al. (2004). Overall, research suggests that learning might play a critical role in the relationship between decision-making and memory.

To summarise, the current chapter discussed the impact of decision-making and moral decision type on long-term episodic memory. The two experiments reported above did not provide evidence of this effect. The literature on judgment and memory was considered to explain the absence of an effect. By extrapolation, I suggested that the impact of decision-making on memory might only be observed under specific experimental designs - i.e., memory-based tasks. Additionally, I discussed the role of learning in the relationship between decision-making and memory. I suggested that learning might be an essential aspect of the relationship between decision-making and memory. Therefore, future research on decision-making and memory should control the potential role of learning.

#### **Chapter 4: The role of learning in the relationship between decision-making and memory.**

Theories on the relationship between decision-making and memory unequivocally state that the two processes have a reciprocal influence on one another (Fellows, 2018; Tremel et al., 2018). However, the impact of decision-making on memory has yet to be fully characterised. In Chapter 2, a new experimental paradigm combining decision-making and memory tasks was developed. This paradigm was used in Experiment 4 and Experiment 5 presented in Chapter 3. Chapter 3 investigates the impact of decision-making and episodic memory using a moral dilemma. The moral dilemma story presents a fictitious scenario where a decision must be made between a utilitarian and a deontological option. In Experiment 4, a control condition was induced by replacing the decision-making question with an unrelated question. After a one-day delay, the retention of the dilemma story was measured. Two hypotheses were formulated: 1. Making a decision on the dilemma will enhance the retention of the story; 2. Making a utilitarian decision will lead to better retention of the story than a deontological decision. The results contradicted both hypotheses. The current chapter reconsiders the previous findings by investigating whether learning influence the effect of decision-making on memory.

The results from Experiment 4 contradict previous results from Murty et al. (2019). In their experiment, Murty et al. (2019) report enhanced retention of study materials when decision-making is induced. On the contrary, the results of Experiment 4 suggest that decision-making does not impact memory retention. To explain the divergence between Murty et al.'s (2019) results and mine, the differences between

each experimental paradigm were considered. First, the two studies use different memory paradigms. Murty et al. (2019) use an intentional immediate recognition task before the delayed recall. Using a recognition task enable additional learning of the study materials. To avoid confounding effects of learning, I controlled participants' encoding of the dilemma story with an incidental free recall task. The difference in results between Murty et al. (2019) and I could suggest that the impact of decision-making on memory is modulated by learning.

The role of learning in the relationship between decision-making and memory has been previously theorised. Fellows (2018) suggests that decision-making impacts memory through associative learning. When a decision is made, the selected option and its related outcome are associated. This association will then be updated in memory to guide future decisions. A similar hypothesis is formulated in Tremel et al. (2018), who assume that learning and retrieving decisions' outcomes guide future decisions. The present research proposes to experimentally test the role of learning in the relationship between decision-making and memory.

It must be noted that the current research does not question the relationship between decision-making and learning. Decision-making and learning are indubitably related. Compelling evidence from neurobiological (for a review, see Ni & Li, 2021) and clinical studies (Scholl & Klein-Flugge, 2018) provide an in-depth characterisation of this relationship. Here, I will investigate whether learning modulates the impact of decision-making on memory.

Another point of divergence between Murty et al.'s (2019) and my experiments was decision-making elicitation. I evoked decision-making using a moral dilemma

task. Murty et al. (2019) described their task as arbitrary decision-making, imbuing participants with a sense of agency over their environment. However, since decision-making encompasses judgment (Fischhoff & Broomell, 2020), a decision-making task cannot be arbitrary. Thus, I argued that Murty et al.'s (2019) experimental task does not elicit decision-making but rather perceived agency.

Considering the two main limitations of Murty et al.'s (2019), I suggest that their results illustrate the impact of perceived agency on learning. The effect of perceived agency on learning is known in the literature as the 'self-directed learning effect'. This effect refers to the increase in learning when allowing some control over the learning experience (Coverdale & Nairne, 2019). Studies have shown that even minimal control such as item location (Voss et al., 2011), 3D manipulation of the items (Harman et al., 1999), or timing of the stimuli onset (Markant et al., 2014) can improve learning. Additionally, research has found that choosing which items to encode enhances their retention compared to unchosen items (Kornell & Metcalfe, 2006; Coverdale et al., 2019). Some refer to this latter phenomenon as 'the self-choice effect' (Watanabe & Soraci, 2004). Overall, the literature provides abundant evidence that making choices enhances learning.

Could the benefit of an act of choosing on learning, or the self-choice effect, apply in the context of decision-making? On the one hand, self-choice paradigms may share similar aspects with decision-making paradigms. Firstly, a part of the decision-making process involves an act of choosing (see Section 1.1.1). Secondly, the self-choice effect might be induced by the process of judgment. Judgment is the initial step of decision-making, where the outcomes of possible options are estimated (Fischhoff

& Broomell, 2020). Research has shown that, in self-choice paradigms, choices are made to maximise learning outcomes - e.g., allocating more study time to an item considered more difficult to remember (Kornell & Metcalfe, 2006). This idea was further investigated by Katzman & Hartley (2020). The authors studied the impact of utility value on the benefit of choice by manipulating the degree to which each option was consequential to earn a reward. Their results showed that the benefit of choice was observed only when choice options had the greatest utility. These findings suggest that the self-choice effect is related to a process of utility judgment. In sum, a self-choice effect could be expected in decision-making paradigms as they overlap with self-choice paradigms in terms of choice and judgement.

On the other hand, the role of judgement in the self-choice effect is debated. A self-choice effect can be observed in studies using incidental memory recall (Nairne et al., 2007; Cunningham et al., 2011). However, if participants are not informed that they will need to recall items, their choice cannot be guided by the intention to maximise learning outcomes. In other words, if the self-choice effect was due to utility judgment, such an effect could not be observed in incidental recall tasks. In 2019, Coverdale & Nairne conducted two experiments to identify the determinant of the self-choice effect. The authors investigated whether the self-choice effect is due to the sole act of choosing or utility judgment. In the earlier assumption, the mnemonic effect would be simply due to choice's enactment. Whether the choice made is useful to the encoding context would not impact the mnemonic effect. In the latter assumption, the self-choice effect requires more than choice's enactment to be observed. The effect would depend on making an appropriate judgment of each option's utility to the context

of encoding. Thus, the mnemonic effect would depend on utility judgment. In their experiments, participants were presented with fictitious survival scenarios and were asked to choose between two items - the more or the less useful for survival. Their results suggest that congruity is not responsible for the mnemonic benefit of choice. The authors concluded that the memory benefit is due to the choice's enactment.

Overall, it is unclear whether the benefit of self-choice on learning and memory could be observed in a decision-making paradigm. The current chapter will address this question by investigating the role of learning in the impact of decision-making on memory. The relevance of this investigation is two-fold. Firstly, it addresses the question raised by the previous chapter - given the strong theoretical assumption that decision-making impacts memory, does learning play a critical role in this relationship? Secondly, it allows considering the extent to which choosing and deciding are two distinctive processes. This question is of paramount importance since the literature on the self-choice effect, and self-directed learning uses choice-making and decision-making interchangeably (Markant et al., 2014, p1212; Murty et al., 2016, p1; Coverdale & Nairne, 2019; p1312; Dubrow et al., 2019, p1965, Katzman & Hartley, 2020, p3). Using a decision-making paradigm, will I observe similar effects are in self-choice paradigms? To answer the questions above, I conducted a longitudinal study comparing the difference in the retention of a dilemma story as a function of decision-making and learning.

#### **4.1 Experiment 6: The Effect of Decision-Making on Long-Term Forgetting as a Function of Learning**

The literature on decision-making and memory considers that the two processes influence one another. Although the impact of decision-making on memory is theoretically accepted, few empirical results support this assumption. Chapter 2 reports Experiments 4 and 5 investigating the impact of decision-making and moral decision types on episodic memory. Decision-making and memory were studied using a moral dilemma story. Decision-making types were elicited by making a deontological or a utilitarian decision on the moral dilemma. A utilitarian decision focus on maximising the benefits and minimising the costs across affected individuals (Mill, 1861/1998). A deontological decision follows sensitivity to rights and duties (Kant, 1785/2001). In Experiment 4, a control condition was induced by replacing the dilemma question with an unrelated question. Memory was assessed through the retention of the dilemma story in both experiments.

Experiment 4 and Experiment 5 lead to surprising results. Experiment 4 revealed no impact of decision-making on retention after a one-day delay, regardless of the moral decision type. Experiment 5 replicated and expanded these results as similar retention across moral decision types were observed after a one-month delay. Therefore, I concluded that decision-making in moral dilemmas does not affect episodic memory. Altogether, my results defy the long-running assumption of a reciprocal relationship between decision-making and memory. The current experiment will further investigate this relationship by considering the role of learning. Firstly, I will discuss theories and empirical results suggesting that decision-making affects

memory through learning. Secondly, I will consider potential limitations and alternative assumptions.

To explore the role of learning in the relationship between decision-making and memory, I suggest using a retrieval practice paradigm. Retrieval practice is a learning technique based on the testing effect (Roediger & Karpicke, 2006b). The testing effect refers to the greater benefit of retrieving information from memory over restudying the same information. Retrieval practice paradigms have been used in the literature to study the impact of memory on decision-making. For instance, Iglesias-Parro and Gomes-Ariza (2006) studied the effect of retrieval practice on decision-making. The authors hypothesised that the inhibition of information during retrieval would make it less accessible and thus impact decisions where such information is required. In their experiment, participants studied six sets of fictitious candidates for a job opportunity. Retrieval practice was induced by presenting a candidate's name along with an attribute stem. Part of candidates' names was presented during the retrieval phase, and participants recalled the associated attribute. This was followed by a decision-making phase measuring participants' preferences for good vs. poor candidates. The experiment ended with a cued recall phase of all candidates' names and attributes. Their results have shown that the practice of relevant and irrelevant attributes for poor and good candidates influence the final recall and participants' preference. The authors concluded that retrieval practice biases decision-making. Another example of a retrieval practice paradigm in decision-making research is Coman et al. (2013). The authors explored the impact of retrieval-induced forgetting on decision-making. In their experiment, the participants were taught about the advantages and disadvantages of

treatment options for a fictitious disease. Participants read treatment pamphlets during the retrieval phase, presenting some treatments' irrelevant or relevant advantages and disadvantages. After a final cued recall, participants chose the best treatment option for a fictitious friend. Their results showed that participants preferred the mentioned or unmentioned treatment depending on whether advantages or disadvantages were accessible in the pamphlet. The authors concluded that retrieval-induced forgetting could adversely affect decision-making. Altogether, these findings suggest that the retrieval practice is a suitable paradigm for studying the relationship between episodic memory and decision-making.

The present research focuses on the role of learning in the effect of decision-making on memory. I suggest comparing the impact of decision-making on memory as a function of retrieval practice. Considering the literature above, I hypothesise that retrieval practice heightens decision-making's impact on memory. Moreover, the dual-process model of moral decision-making (Greene et al., 2004, 2008) suggests that making a utilitarian decision implies deeper processing of the dilemma story than a deontological decision. Considering that a deeper level of processing leads to better retention ( Craik & Lockhart, 1972), it is expected that the benefit of retrieval practice on memory will be of a greater extent when making a utilitarian decision as compared to a deontological decision.

As highlighted above, the literature on decision-making and memory allows reasonable predictions on the effects expected in a retrieval practice paradigm. However, my previous experiments (see Chapter 2) and related literature on judgment and memory bring me to ponder the assumptions formulated above. Judgment and

decision-making are related inasmuch as judgment is the initial step of the decision-making process. The literature on judgment and memory reports inconsistencies in findings. Hastie and Park (1986) suggest that the relationship between judgment and memory is found in memory-based designs but not online designs. On the one hand, memory-based tasks surprise participants with a novel judgment that cannot be anticipated, automated, or approximated by former similar judgments. On the other hand, in online judgement tasks, a judgment is made while information is still being processed. The information for the judgment situation has not yet been stored in long-term episodic memory. According to the authors, this online vs. memory-based distinction dictates whether a relationship between judgment and memory will be observed. In a set of four experiments, the authors measured the relationship between memory and judgement using a judgment rating task (on a 10-point scale) and a recall task (proportion of information correctly recalled). The results showed that judgment rating and memory recall were only related in memory-based tasks but not online tasks.

Let us assume that the online vs. memory-based dichotomy applies to the broader decision-making process. In such a case, the expected effect of decision-making on memory depends on the task used. However, it is not clear where my experimental design stands in the online vs. memory-based dichotomy. On the one hand, the task was intended to be memory-based. Participants were not informed of the decision-making component of the experiment. They should be surprised by the decision-making task. On the other hand, the decision phase directly follows the listening to the dilemma story. Thus, decision-making and the treatment of the story

information are concomitantly engaged. Furthermore, moral dilemmas are well-known by the general public, which may provide participants with previous knowledge on their inclination to make utilitarian or deontological decisions. Therefore, the task might rather be online than memory-based decision-making. In that case, an absence of the effect of decision-making on memory would be predicted.

In sum, this experiment considers two scenarios characterising the impact of decision-making on memory. One possibility is that learning guides the impact of decision-making on memory. Decision-making shows an effect on memory only when learning is induced. In such a case, decision-making will only impact episodic memory under the retrieval practice condition. In other words, retrieval practice would interact with decision-making and moral decision types to impact memory retention. Another possibility is that the experimental design involves online decision-making. Thus, decision-making would not impact memory and promoting learning would not influence this effect.

#### 4.1.1 Method

##### **Participants**

A total of 185 participants were recruited via the Department of Psychology participant pool and the general public using word-of-mouth and online advertisement. Volunteers who self-reported memory difficulties due to neurological or neurodevelopmental conditions were excluded from the experiment. The socio-demographic characteristics of the sample are provided in Table 25. The study was approved by the University of Edinburgh - School of Philosophy, Psychology and Language Sciences Ethics

Committee (No 24-1920/11). Participants gave informed consent in accordance with the British Psychological Society Code of Ethics and Conduct. Participants received course credits or were compensated with a small honorarium for their participation. Among the 185 participants, 104 were initially recruited as part of Experiment 4 (one-day delayed recall) and were called for a one-week and one-month follow-up session to compose the retrieval practice group of the current experiment. Joining the data collection of two experiments within one recruitment pool and procedure was necessary to capitalise on financial and human resources and achieve a reasonable sample size. Two volunteer research assistants (fourth-year undergraduate students) assisted me in participant recruitment and data collection.

**Table 25**

*Socio-Demographic Characteristics of The Sample*

<i>N</i>	Age					Education					Gender		
	<i>M</i>	<i>SD</i>	Min	Median	Max	1	2	3	4	NA	1	2	3
185	24.80	8.41	17	22	61	4	106	63	9	3	143	41	1

*Note.* *N*, *M*, *SD*, Min, Max and NA refer to sample size, mean, standard deviation, minimum, maximum and missing data respectively. Missing data are either due to a non-classifiable input or to the item not being answered.

## **Materials**

The administration of the tasks was computerised using OpenSesame. For all tasks, the instructions were written on the computer screen. For regularity in reading speed, appropriate pronunciation and no variation in speech, the reading of the digit span items, and the dilemma story were recorded via TheScottishVoice text-to-speech software (The Scottish Voice, 2011). The Cerevoice Stuart version 4.0.1, a Scottish male voice, was used. Audio recordings were presented through headphones placed and adjusted for volume before the experiment started. Depending on the task, participants answered by typing on a numeric keyboard or speaking aloud into the headphones' microphone.

### ***The Digit Span Tasks***

A computerised version of the Digit Span-Forward and Digit Span-Backward subtests of the WAIS-IV (Wechsler, 2008) was developed. In both Digit Span-Forward and Digit Span-Backward, there are eight items with two sequences of digits. Each subtest has sixteen sequences in total. The sequences of digits were read by TheScottishVoice text-to-speech software (The Scottish Voice, 2011). After listening to a sequence, the participant must recall all digits in the same order for Digit Span-Forward and reverse order for the Digit Span-Backward. In the Digit Span-Forward, the sequences go from two (item 1) to nine digits (item 8), increasing from one digit every item. In the Digit Span-Backward, the sequences start with two digits for items 1 and 2, then increase from one to seven digits (item 8). Each sequence is scored 0 if failed and 1 if correctly recalled. A subtest ends when both sequences from the same

item are failed. The raw scores correspond to the number of sequences correctly recalled for each subtest.

The use of the digit span task was twofold. Firstly, as working memory is known to impact decision-making (Gansler et al., 2011), memory retrieval and learning (Cubelli, 2010), digit span was administered to verify that experimental groups did not significantly differ in their working memory abilities. Secondly, digit span was presented to participants as the main task of the experiment to avoid intentional rehearsal of the dilemma story during the delay periods.

### ***The Moral Dilemma Task***

An audio recording of a modified Nobel Prize Dilemma was used. For scoring purposes, the dilemma story was divided into 27 ideas to be recalled (see Appendix 6). The scoring was based on the logical memory subtest of the Wechsler Memory Scale-IV (WMS-IV; Wechsler, 2009). Each idea was scored 1 if correctly recalled and 0 if not recalled.

### **Design**

The experiment used a two-factor between-subjects design. The first independent variable was decision-making and had three levels: no-decision, deontological, and utilitarian. One-third of the participants were randomly allocated to the no-decision group. The remaining two-thirds were deontological or utilitarian, depending on their answer to the dilemma question. The second independent variable was retrieval practice and has two levels: with or without. Participants recruited as part of Experiment 4 were allocated to the retrieval practice condition. The remaining

participants composed the without-retrieval practice group. The dependent variable was the number of correctly recalled ideas from the dilemma story.

## **Procedure**

### ***Initial Testing Session***

The experiment started with the computerised version of the Digit Span-Forward, followed by the Digit Span-Backward. The standardised instructions of the WAIS-IV (Wechsler, 2008) were written on the screen. After that, the audio recording of the first digit sequence was displayed at a rate of one digit per second. Then, the participant had to recall the digit sequence by typing the digits on the numeric keyboard. Digits had to be recalled in the same order for the Digit Span-Forward and reverse order for the Digit Span-Backward. A subtest ends when both sequences from the same item are failed.

The moral dilemma task was performed after completing the digit span tasks. The procedure was based on the logical memory subtest of the WMS-IV (Wechsler, 2009). Participants were given the following instructions: "You are going to listen to a short story. Listen carefully and try to remember as much as you can". Then, participants listened to the audio recording of the modified Nobel Prize dilemma story.

The narrative of the story was the following: "You and a fellow researcher, Professor Joe Grant, have discovered a powerful new energy source that is cheap, safe, and clean. You realise that this could lead to the elimination of pollution and poverty around the world. However, your colleague wants to sell this discovery to the US army, which will turn it into a powerful new weapon. You can prevent this by

releasing propane gas in the lab, so that when he turns on his computer, it will cause an explosion within two seconds. This will also cause the death of your colleague. Everyone will think that it was just a lab accident, and the discovery will not be sold to those who might create a weapon out of it.”

Then, participants in the decision condition were asked to answer the question: “What are you going to do?” Below the question, participants were presented with the following two options in a randomised order: “You release the gas” or “You let your colleague sell the discovery”. Participants in the no-decision condition were asked to answer the question: “Which colour do you prefer?” Below the question, the following two options were presented in a randomised order: “You prefer the colour blue” or “You prefer the colour red”.

After answering one of the questions above, participants were given the instructions for memory recall: “Can you say everything you can remember about the story? You can say things even if you are not sure about the details or order”. Participants answered by speaking aloud. Answers were recorded and saved under an anonymised participant identification number. The first session ended with the collection of socio-demographic details, and one or three appointments were given for follow-up sessions.

### ***Follow-up Sessions***

Follow-up sessions were conducted via phone calls. Participants in the retrieval practice condition were contacted at a one-day, one-week, and one-month delay. Participants in the condition without-retrieval practice were only called after a one-month delay. In both conditions, participants were not informed that the follow-up

sessions consisted in recalling the dilemma story. This was done so participants in the without-retrieval practice conditions would not engage in learning strategies. Although participants in the retrieval practice condition were also not informed of delayed recalls, this should not interfere with the efficacy of retrieval practice (Smith et al., 2013). The instruction for delayed recall was the following: "Can you tell me everything you remember about the story you heard yesterday/ last week/ last month? Please recall everything you can, even if you are not sure." No time limit was set for the participant to answer. The recall task ended when the participants informed the experimenter that they could not recall more. The final session ended with a debriefing. The purpose of the experiment was explained, and participants' questions were answered.

### ***COVID-19 Notice***

The experiment was split into two to four sessions. The first session initially took place in the laboratory of the Psychology Department, University of Edinburgh. Participants completed the tasks on the computer of the laboratory room.

Following the COVID outbreak, on-campus in-person testing was interrupted. Consequently, the setting of the first session had to be adapted. Since the experimental tasks were computerised, I set up an online version of the experiment. All precautions were taken, so the first online session was similar to the first on-campus session. Participants met with the experimenter during an online meeting when they were given general instructions. Then they started the experiment by following a URL.

Follow-up sessions remained unchanged as they were conducted via phone calls.

#### 4.1.2 Statistical analyses

The analysis script and a package containing the anonymised dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdf7ec8b03ea1ad66).

Participants over 50 years-old ( $n = 7$ ) were removed from the analyses as they were not homogeneously distributed across the six groups. Participants who failed to attend the follow-up session for the one-month delayed recall ( $n = 13$ ) were also removed from the analyses. In total, 165 participants were included in the analyses.

Multiple linear regressions were used to verify whether the groups differ in post-decision recall of the dilemma story or Digit Span-Forward and Digit Span-Backward scores. When assumptions were not met, I have applied a bootstrapping method on multiple regression models, as recommended in Little (2013). Bootstrapping is a method for determining the significance of the sample results. It is achieved by resampling with replacement from the original sample. Therefore, it does not rely on assumptions about the population distribution like  $t$  or  $F$  families (Erceg-Hurn et al., 2013). Significance is considered by interpreting the 95% confidence intervals – the result is significant at a 5% confidence level when the 95%CI for slope estimates do not include 0. When a bootstrap method was used, 10'000 bootstrap samples were computed as suggested in Hesterberg, (2015).

The principal statistical analysis test whether the impact of decision-making on memory changes when controlling for learning – here elicited with retrieval practice. A bootstrapping method was applied to the multiple regression model because the

assumptions for the parametric procedure were not met. Contrast coding was also applied to the decision-making parameters. Contrasts analyses are significance tests for focused questions that allow one to evaluate specific predictions (Rosenthal & Rosnow, 1985). My contrasts focus on two comparisons<sup>6</sup>. The first contrast focus on the impact of decision-making. It compares retention when making any decision (deontological and utilitarian) vs. no-decision. The second contrast focus on the difference between moral decision types. It compares retention when making a deontological decision vs. a utilitarian decision. By applying contrast coding to the regression parameters, I focus only on comparisons relevant to my research questions. Focused comparisons increase statistical power.

The statistical model is the following:

$$\begin{aligned}
 ItemRecall = & \beta_0 \\
 & + \beta_1 * (WithRetrievalPractice) \\
 & + \beta_2 * (AnyDecision\ vs.\ NoDecision) \\
 & + \beta_3 * (Utilitarian\ vs.\ Deontological) \\
 & + \beta_4 * (WithRetrievalPractice) * (AnyDecision\ vs.\ No\ Decision) \\
 & + \beta_5 * (WithRetrievalPractice) * (Utilitarian\ vs.\ Deontological) + \epsilon
 \end{aligned}$$

Where *Without-retrieval Practice* is the reference level for the *Retrieval Practice* variable.

The hypotheses at test are presented in Figure 13.

---

<sup>6</sup> The same contrasts coding was used in Experiment 4. See Section 3.2.2 for a detailed presentation of contrasts analyses.

**Figure 13***Research Questions and Statistical Hypotheses of Experiment 6*

**Research question 1: Does the impact of decision-making on the retention of the dilemma story depends on retrieval practice?**

$$H_0: \beta_4 = 0$$

The interaction between retrieval practice and decision-making is equal to zero.

$$H_1: \beta_4 \neq 0$$

The interaction between retrieval practice and decision-making is not equal to zero.

**Research question 2: Does the impact of moral decision types on the retention of the dilemma story depends on retrieval practice?**

$$H_0: \beta_5 = 0$$

The interaction between retrieval practice and moral decision types is equal to zero.

$$H_1: \beta_5 \neq 0$$

The interaction between retrieval practice and moral decision types is not equal to zero.

Akaike information criterion (AIC) and Bayesian information Criterion (BIC) model selection were used to distinguish between models describing the relationship between retrieval practice, decision-making and the dilemma story recall. Three models were compared. A specification of the three models at test is provided in Table 26.

Table 26

*Experiment 6 Models' Specification*

Model 0 (M0): Null model	$M0: ItemRecall = \beta_0$ $+ \beta_1 * (WithRetrievalPractice) + \epsilon$
Model 1 (M1): Decision model	$M2: ItemRecall = \beta_0$ $+ \beta_1 * (WithRetrievalPractice)$ $+ \beta_1 * (WithRetrievalPractice)$ $+ \beta_2 * (AnyDecision vs. NoDecision)$ $+ \beta_3 * (Utilitarian vs. Deontological) + \epsilon$
Model 2 (M2): Interaction between retrieval practice and decision model	$M3: ItemRecall = \beta_0$ $+ \beta_1 * (WithRetrievalPractice)$ $+ \beta_2 * (AnyDecision vs. NoDecision)$ $+ \beta_3 * (Utilitarian vs. Deontological)$ $+ \beta_4 * (WithRetrievalPractice) * (Any Decision vs. No Decision) + \epsilon$ $+ \beta_5 * (WithRetrievalPractice) * (Utilitarian vs. Deontological) + \epsilon$

*Note:* Without-retrieval practice is the reference level for the retrieval practice condition. The effect of retrieval practice on memory retention is known and assumed from the literature. Therefore, it was included in the null model and was not the target of my comparisons.

### 4.1.3 Results

#### Post-Decision Recall

##### *Descriptive statistics*

Descriptive statistics of post-decision recall per decision and retrieval practice groups are available in Table 27.

**Table 27**

*Descriptive Statistics of Post-Decision Recall for Each Decision and Retrieval Practice Group*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Without RP						
No-decision	27	13.70	5.78	7.00	12.00	24.00
Deontological	25	13.08	4.83	2.00	14.00	22.00
Utilitarian	21	13.81	5.29	3.00	15.00	22.00
With RP						
No-decision	29	12.38	5.99	2.00	13.00	21.00
Deontological	32	12.65	6.19	1.00	14.00	22.00
Utilitarian	31	12.74	4.53	6.00	12.00	22.00

*Note.* The recall of the dilemma story is scored over 27 ideas. *N*, *M*, *SD*, *Min*, and *Max* represent sample size, mean, standard deviation, minimum and maximum respectively.

##### *Inferential statistics*

Diagnostic plots for the distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(5,158) = 0.25$ ,  $p = .46$ . The Durbin-Watson for the independence of the residual was not significant,  $D-$

$W= 1.99, p= .96$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.98, p= .006$ . Assumptions for the multiple linear regression were not met. The bootstrapped model revealed that all comparisons were not significant (Table 28).

**Table 28**

*Bootstrapped Model of Post-Decision Recall as a Function of Decision-making and Retrieval Practice*

Parameter	Average bootstrap estimate	95%CI	
		LL	UL
Intercept: Without RP	13.70	11.59	15.92
With RP	-1.32	-4.41	1.71
Any Decision - No Decision	-0.62	-3.51	2.23
Utilitarian - Deontological	0.11	-3.00	3.17
With RP: Any Decision - No Decision	0.89	-3.39	5.06
With RP: Utilitarian - Deontological	0.26	-3.84	4.43

*Note.* RP, CI, LL and UL, respectively stands for retrieval practice, confidence interval, lower limit and upper limit

## **Digit Span-Forward**

### ***Descriptive statistics***

Descriptive statistics of Digit Span-Forward per decision and retrieval practice groups are available in Table 31.

**Table 29**

*Descriptive Statistics of Digit Span-Forward for Each Decision and Retrieval Practice*

*Group*

Decision	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Without RP						
No-decision	27.00	10.15	1.66	7.00	10.00	13.00
Deontological	25.00	10.36	1.68	8.00	10.00	14.00
Utilitarian	21.00	9.71	2.37	4.00	9.00	14.00
With RP						
No-decision	29.00	9.52	1.64	6.00	10.00	12.00
Deontological	32.00	10.10	1.74	6.00	10.00	13.00
Utilitarian	31.00	10.23	1.75	7.00	10.00	14.00

*Note.* The Digit Span-Forward is scored over 16 items. *N*, *M*, *SD*, *Min*, and *Max* represent sample size, mean, standard deviation, minimum and maximum respectively.

### ***Inferential statistics***

Diagnostic plots for the distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(5,158)= 0.89, p= .49$ . The Durbin-Watson for the independence of the residual was not significant,  $D-W= 2.30, p= .05$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.99, p= .52$ . Assumptions for the multiple linear regression were met. The multiple regression model revealed that all comparisons were not significant (Table 30).

**Table 30**

*Multiple Regression Model of Digit Span as a Function of Decision-making and Retrieval Practice*

Parameter	Coefficient	SE	95% CI	<i>t</i> (158)	<i>p</i>
Intercept: Without RP	10.15	0.35	(9.47, 10.83)	29.38	< .001
With RP	-0.63	0.48	(-1.58, 0.32)	-1.31	.19
Deontological	0.21	0.50	(-0.77, 1.20)	0.43	.67
Utilitarian	-0.43	0.52	(-1.47, 0.60)	-0.83	.41
With RP: Deontological	0.37	0.68	(-0.98, 1.71)	0.54	.59
With RP: Utilitarian	1.14	0.70	(-0.24, 2.52)	1.64	.10

*Note.* RP, SE, CI, correspond to retrieval practice, standard error and confidence interval respectively.

### **Digit Span-Backward**

#### ***Descriptive statistics***

Descriptive statistics of Digit Span-Backward per decision and retrieval practice groups are available in Table 31.

**Table 31**

*Descriptive Statistics of Digit Span-Forward for Each Decision and Retrieval Practice*

*Group*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Without RP						
No-decision	27.00	9.56	1.40	6.00	10.00	12.00
Deontological	25.00	9.28	1.74	6.00	10.00	12.00
Utilitarian	21.00	9.71	1.19	8.00	9.00	12.00
With RP						
No-decision	29.00	9.17	1.77	3.00	9.00	11.00
Deontological	32.00	9.48	1.46	6.00	10.00	11.00
Utilitarian	31.00	9.58	1.12	7.00	10.00	11.00

*Note.* The Digit Span-Backward is scored over 16 items. *N*, *M*, *SD*, *Min*, and *Max* represent sample size, mean, standard deviation, minimum and maximum respectively.

### ***Inferential statistics***

Diagnostic plots for the distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance  $F(5,158)= 1.24$ ,  $p=.29$ . The Durbin-Watson for the independence of the residual was not significant,  $D-W= 1.95$ ,  $p=.69$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.95$ ,  $p< .001$ . Assumptions for the multiple linear regression were not met. The bootstrapped model revealed that all comparisons were not significant (Table 32).

**Table 32**

*Bootstrapped Regression Model of Digit Span-Backward Predicted by Retrieval Practice and Decision-Making*

Parameter	Average bootstrap estimate	95%CI	
		LL	UL
Intercept: Without RP	9.56	9.00	10.07
With RP	-0.38	-1.24	0.43
Any Decision - No Decision	-0.28	-1.12	0.60
Utilitarian - Deontological	0.16	-0.56	0.91
With RP: Any Decision - No Decision	0.59	-0.61	1.78
With RP: Utilitarian - Deontological	0.25	-0.78	1.31

*Note.* RP, CI, LL and UL, respectively stands for retrieval practice, confidence interval, lower limit and upper limit.

## **Delayed recall**

### ***Descriptive statistics***

Descriptive statistics for the number of ideas correctly recall after a one-month delay are available in Table 33.

**Table 33**

*Descriptive Statistics for The Total Number of Items Correctly Recalled as a Function of Decision and Retrieval Practice*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
<b>Without RP</b>						
No-decision	27	4.11	3.07	0.00	4.00	13.00
Deontological	25	4.04	3.59	0.00	3.00	14.00
Utilitarian	21	6.10	4.07	0.00	7.00	14.00
<b>With RP</b>						
No-decision	29	9.62	5.59	0.00	9.00	22.00
Deontological	32	8.91	5.09	0.00	9.00	17.00
Utilitarian	31	8.97	4.44	2.00	9.00	19.00

*Note.* The recall of the dilemma story is scored over 27 ideas. RP, *N*, *M*, *SD*, *Min*, and *Max* represent retrieval practice, sample size, mean, standard deviation, minimum and maximum, respectively.

### ***Inferential statistics***

Diagnostic plots for the distribution of residuals are available in Appendix 1. The Levene's test suggested heteroscedasticity  $F(5,159)= 3.08$ ,  $p < .001$ . The Durbin-Watson for the independence of the residual was not significant,  $D-W= 1.80$ ,  $p= .19$ . The Shapiro-Wilk's normality test showed that residuals are not normally distributed,  $W= 0.99$ ,  $p= .39$ . Assumptions for the multiple linear regression were not met. The bootstrapped model suggests a main effect of retrieval practice. The contrasts did not reach significance. No interaction was found (see Table 34). An illustration of the results is provided in Figure 14.

**Table 34**

*Results of The Bootstrap Regression Model on Total Recall at One-Month Delay*

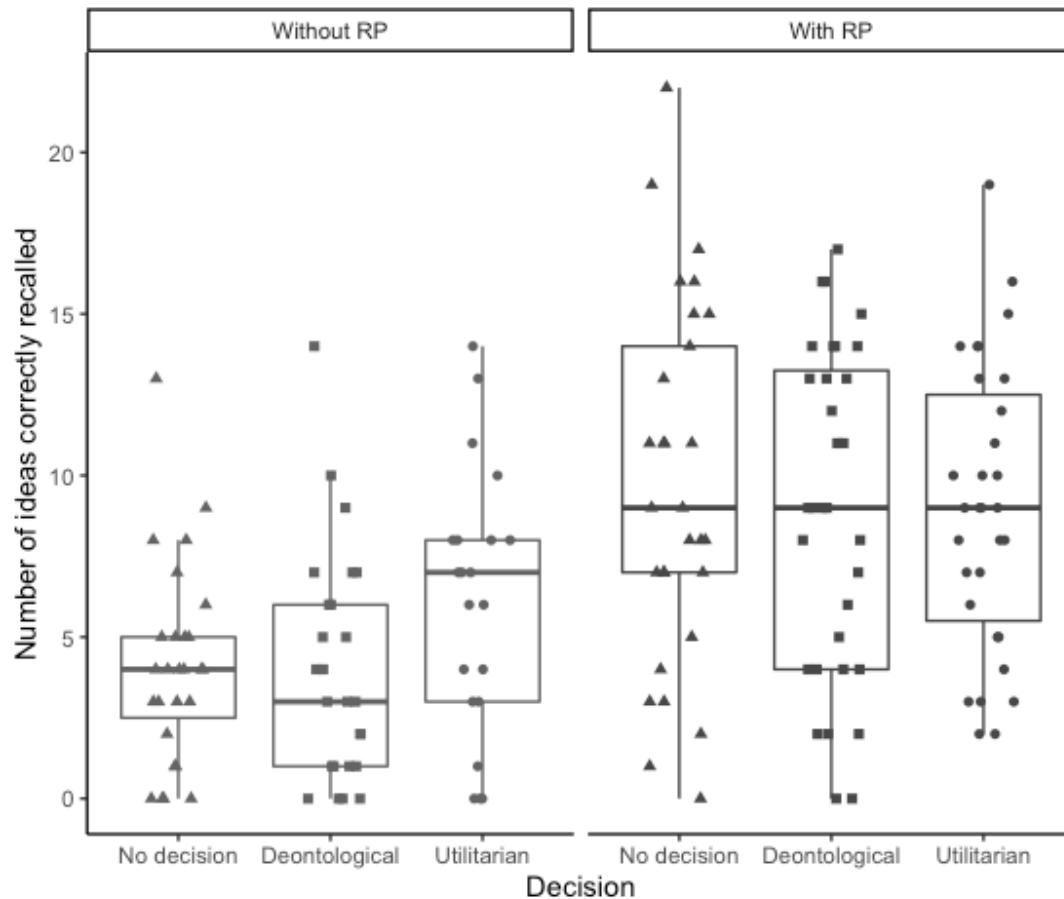
*Predicted by Decision and Retrieval Practice*

Parameter	Average bootstrap estimate	95% CI	
		LL	UL
Intercept: Without RP	4.75	3.93	5.62
With RP	4.42	3.06	5.78
Any Decision - No Decision	0.64	-0.44	1.72
Utilitarian - Deontological	1.03	-0.09	2.13
With RP: Any Decision - No Decision	-1.09	-2.99	0.79
With RP: Utilitarian - Deontological	-1.00	-2.60	0.61

*Note.* RP, CI, LL and UL, respectively stands for retrieval practice, confidence interval, lower limit and upper limit.

**Figure 14**

*Number of Ideas Correctly Recalled After a One-Month Delay By Retrieval Practice (RP) and Decision Group*



*Note:* the boxplots present for each experimental group: data points, the median, the upper (Q3) and lower (Q1) quartiles, the interquartile range (IQR), the highest ( $Q3 + 1.5 \cdot IQR$ ) and lowest ( $Q1 - 1.5 \cdot IQR$ ) when excluding potential outliers.

AIC and BIC were used to distinguished between the three models describing the relationship between retrieval practice, decision-making and the story recall. The decision model had lower AIC (539.41) and BIC (555.15) than the interaction model (AIC= 540.76, BIC= 562.79). Overall, the null model had the lowest AIC (537.87) and BIC (547.31).

***Exploratory analysis: effect of administration conditions on post-decision recall***

To verify whether having different administration conditions (online vs in-person) had an impact on post-decision recall, I ran a Mann-Whitney U test.

Descriptive statistics

Descriptive statistics for the number of ideas correctly recalled post-decision for each administration condition is available in Table 35.

**Table 35**

Number of Ideas Correctly Recalled per Administration Conditions

Administration	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
In-person	140.00	12.72	5.58	1.00	13.00	24.00
Online	38.00	13.63	4.70	4.00	14.00	22.00

*Note.* The Dilemma Story recall is scored over 27 ideas. *N*, *M*, *SD*, *Min*, and *Max* refer to sample size, mean, standard deviation, minimum, and maximum respectively.

Inferential statistics

Diagnostic plots for the means and distribution of residuals are available in Appendix 1. The Levene's test suggested that errors have constant variance,  $F(1, 175) = 2.39, p = .12$ . The Shapiro-Wilk's test was significant,  $W = 0.98, p = .01$ . The assumptions for running a parametric t-test were not met. The Mann-Whitney U test revealed no significant difference in post-decision recall between the in-person and

the online administration conditions,  $W= 2422$ ,  $p= .43$ . An illustration of the results is available in Appendix 1.

#### 4.1.4 Discussion

The current experiment investigates whether learning is a prerequisite to observing an impact of decision-making on memory. Learning was elicited using retrieval practice. A moral dilemma story was used to manipulate decision-making and measure long-term episodic memory retention. It was hypothesised that decision-making will interact with retrieval practice to impact long-term episodic memory retention. However, the results revealed no effect of decision-making on memory even when retrieval practice was induced.

One main limitation of the current experiment is the popularity of moral dilemmas. Considering that moral dilemmas are well-known to the public; participants may anticipate the decision-making task. In that case, the impact on the results would be two-fold. First, the control condition would not be valid. If participants engage in the process of making a decision even when not prompted to, this explains why no difference in retention can be observed between the no-decision and the decision groups. Moreover, any anticipation of the decision-making task would also explain that retrieval practice does not lead to a difference in retention between the utilitarian and the deontological groups. Indeed, following Hastie and Park (1986), participants must be “surprised” by a novel judgement to observe an impact on memory. Applying this assumption to the broader process of decision-making, the results are in line with Hastie and Park’s (1986) predictions.

Another limitation is that the groups' sizes are unbalanced. The difference in group size is due to participants' dropout. Although dropouts are inherent in longitudinal designs, the difference of 10 observations between the with-retrieval practice utilitarian group and the without-retrieval practice utilitarian group is rather substantial. Unbalanced groups' size can lead to heteroscedasticity and exacerbate statistical error related to the use of p-values as tests of significance (Johnson, in Little, 2013). However, this issue was addressed by using bootstrap methods and CIs interpretation for null hypothesis.

Considering the limitations discussed above, my results suggest Hastie and Park's (1986) distinction between online vs. memory-based judgement tasks can be generalised to the broader process of decision-making. According to the authors, judgement impacts memory in memory-based tasks but not in online tasks. Given that moral dilemmas are popular tools, and that the decision must be made right after listening to the dilemma story, my experimental paradigm falls within the online task category. This explains why, even when controlling for learning, no impact of decision-making on memory can be found.

Theories on relationship decision-making and memory highlight the role of learning in coordinating the two processes. When a decision is made, the association between the chosen option and its related outcomes will be learned and updated in memory (Fellows, 2018, Tremel et al., 2018). Thus, I hypothesised that a certain level of learning might be required to observe an impact of decision-making on memory. As the results revealed a main effect of retrieval practice on memory retention, the paradigm use was efficient in manipulating the learning of the dilemma story. However,

if learning was the underpinning process of the relationship between decision-making and memory, an interaction effect between learning should have been observed. The fact that, in the current experiment, learning does not interact with decision-making to impact memory should bring to reconsider the theory aforementioned.

#### 4.1.5 Chapter 4 Discussion

The current research aimed at investigating whether learning plays a role in the relationship between decision-making and memory. Two contradicting theoretical frameworks were considered. On the one hand, common characterisations of decision-making and memory describe the two processes as influencing one another through learning. On the other hand, models from related literature on judgment and memory suggests that an impact of the earlier process on the latter would be observed in memory-based tasks but not online tasks. My findings seem in line with the judgment literature as they show that, in moral dilemmas, decision-making does not influence memory even when promoting learning.

Such findings bring to reconsider current models of decision-making and memory. The literature provides an in-depth characterisation of the impact of memory on decision-making (Redish & Mizumori, 2015; Madan et al. 2014; Johnson & Busemeyer, 2010; Dougherty et al., 2003). Although this impact is assumed to be reciprocal, I am aware of only one experiment (Murty et al., 2019) attempting to provide empirical evidence of this assumption. After a careful examination of Murty et al.'s (2019), I suggested that the observed effect might be due to confounding effects of learning and substantial limitation in their elicitation of decision-making.

One reason for the scarcity of empirical evidence of the impact of decision-making on memory might be that previous experimental attempts have been, like the current one, unsuccessful. Previous investigations might have also led to null results, but researchers would have been reluctant to report their results given the strong theoretical assumptions in the field and difficulties to explain the absence of an effect. This phenomenon is also known as publication bias (Rosenthal, 1979). A complete description of the prevalence and impact of this issue in psychological sciences can be found in Heene and Fergusson (2018). Although this conclusion might seem extreme, it would explain why models describing the impact of decision-making on memory (Fellows, 2018 and Tremel et al., 2018) do not provide citations to support this specific assumption.

The literature on judgment and memory provides another reason to believe that the impact of decision-making on memory might not be as systematic as expected. Let us remind the reader that judgment and decision-making are related in the sense that judgment is a step in the broader process of decision-making. Research on judgment and memory provide in-depth descriptions of inconsistencies in providing supportive evidence of the impact on one another. These descriptions are based on seminal work by Hastie and Park (1986). The authors suggested a dichotomic model of judgment task as a function of judgment elicitation. According to their model, an impact of judgment on memory can be observed in tasks where no previous judgments are available, and information related to the judgement to be made is stored in long-term episodic memory. They specifically refer to this type of judgment as memory-based. Tasks, where the judgment is made while the information is still being

processed, are referred to as online judgment tasks. In online judgment tasks, a judgment is made simultaneously to the acquisition of information which is thus stored in memory independently from the specific episodic facts that are learned (Lichtenstein & Srull, 1987). Therefore, in online tasks, judgment should not impact memory recall.

This model is very influential in research on judgment but reference to this model in the context of decision-making are rarer (Redlaw & Lau, 2013; Newell & Shanks, 2014). The present research suggests that the online vs memory-based dichotomy might apply to certain decision-making tasks, such as moral dilemmas. Future research should further investigate whether Hastie and Park's (1986) characterisation of the relationship between judgment and memory could be generalised to the broader process of decision-making.

## Chapter 5: General Discussion

### 5.1 Summary of the findings

The current thesis presents six experiments reporting novel findings on the relationship between decision-making and memory. The investigation arose from the accrued assumption that this relationship is biunivocal. However, empirical evidence focuses solely on the impact of memory on decision-making. Therefore, I aimed at exploring the reciprocal hypothesis, that is, the impact that decision-making might have on memory. In Chapter 2, I developed an original experimental paradigm based on memory of a moral dilemma. The following chapters relied on this new paradigm to investigate the effect of decision-making and moral decision type on episodic memory.

Chapter 3 reports two experiments - Experiments 4 and 5. Experiment 4 found no effect of decision-making on memory retention after one day, regardless of the type of moral decisions taken - utilitarian or deontological. Experiment 5 replicated these findings with longer intervals. Making a utilitarian or a deontological decision led to equivalent retention after one day, one week, and one month. A descriptive inspection of the results suggested that decision-making does not even affect the dilemma story's retention pattern. Indeed, participants seemed to remember the same elements of the dilemma story regardless of their decision.

Chapter 4 questioned the role of learning in the relationship between decision-making and memory. Experiment 6 analysed the impact of decision-making on episodic memory when promoting learning over a one-month interval. Learning was promoted with retrieval practice. Although retrieval practice enhanced long-term episodic memory retention, this effect did not interact with decision-making nor with

the type of moral decision. This replicated the finding from Experiment 5, showing no impact of decision-making and moral decision type on forgetting rate.

In sum, taken together, the findings reported in my thesis contradict the long-running assumption of an equipoise, reciprocal relationship between decision-making and memory. The outcome from my thesis does not provide support to the hypothesis of an effect of decision-making on episodic memory. The results suggest that decision-making do not influence episodic memory independently from moral decision types, even when promoting learning.

## **5.2 Limitations**

Several limitations must be considered when interpreting the research findings reported in this thesis. First, the COVID-19 outbreak has impacted two experiments (Experiment 1, and Experiment 6). Following the lockdowns, the data collection was interrupted. The tasks had to be adapted before resuming data collection.

In Experiment 1, the paper-and-pencil, in-person tasks were computerised to enable online administration. The statistical analysis revealed a significant effect of these alterations on the experiment's dependent variable. The computerised version of the task led to lower story recall. However, the effect of tasks' administration did not interact with the effect of the experiment's independent variable. In other words, the decrease in recall due to online administration was constant across the one-day, one-week and one-month delayed recall.

In Experiment 6, the computerised, in-person tasks had to be programmed in a format allowing the online administration of the task. Although the differences between

the in-person and the online version of the task had no significant did not impact story recall, I acknowledge that such alterations of the procedure and materials are not ideal.

Another limitation to consider is the display format of the moral dilemma story (Experiments 4, 5, 6). Studies using moral dilemmas traditionally present a written version of the dilemma stories (Christensen & Gomila, 2012). In Experiments 4, 5 and 6, participants listened to an audio version of the dilemma story. This format was imposed by the research needs. Indeed, the three studies involved measuring the memory retention of the story. Using a written story would have allowed minimal control on the participant's exposure to the materials. For instance, a participant could have read a piece of the story several times or taken more time to read the story. The audio version of the dilemma guaranteed that exposure to the study materials remained constant across participants and experimental conditions. However, moral dilemmas involve complex pieces of prose (Experiment 2). Using an audio instead of a written version of the story might have impacted its intelligibility, which in turn could account for the absence of decision-making and moral decision type effects. Yet, it seems impossible to control for participants' level of understanding without impacting the initial exposure to the study materials.

Finally, it must be acknowledged that moral dilemmas are sensitive stimuli. The slightest variations in the story's framing, presentation, words count, or perspective can impact moral decisions (for a review, see Christensen & Gomila, 2012). As presented in Experiment 3, the dilemma story used in Experiments 4, 5 and 6 is a modified version of the Nobel Prize dilemma (Greene et al., 2008). Although the design of the experiment was carefully piloted (see Experiments 1, 2, 3), the changes made

to the story content might have weakened participants' utilitarian or deontological inclination. Indeed, Conway and Gawronski (2013) suggested that inclinations toward moral decision types are not two dimensions of a bipolar continuum between utilitarianism and deontology. According to the authors, the traditional dilemma methodology (e.i., endorsing either the deontological or the utilitarian option) does not distinguish between selecting one option and rejecting the other. If the utilitarian and deontological judgement are two distinct processes, then both inclinations might be activated simultaneously. The traditional approach, used in the present thesis, cannot capture the different level of activation between the two processes. To delineate the independent activation of utilitarian and deontological inclinations, Conway and Gawronski (2013) suggested applying a process dissociation procedure (Jacoby, 1991). To do so, the authors distinguished two types of dilemmas: incongruent and congruent. In incongruent dilemmas, utilitarian and deontological inclinations are opposed. This is the case my in modified version of the Nobel Prize dilemma, the colleagues should be killed so their discovery will not be used to harm other people. In congruent dilemmas, the outcome of the decision varies so that the suggested action is unacceptable by either utilitarian or deontological reasons. A congruent version of the dilemma used in this thesis could be that the colleague wants to patent the discovery. Then, killing him would not morally support any of the two decision types. By comparing answers between congruent and incongruent dilemmas, Conway and Gawronski (2013) have showed that moral judgements are concomitantly elicit utilitarian and deontological concerns. Overall, Conway and Gawronski's (2013) findings provide a potential explanation for the lack of effect observe in my thesis.

Indeed, if my modified version of the Nobel Prize dilemma elicits both moral inclinations to a great extent, then it is expected that participants would highly process the story content, independently of their decision, leading to equivalent memory recall.

### **5.3 Implications**

#### **5.3.1 Theoretical implications**

The most compelling finding of the current thesis is the remarkable replications of the lack of an effect of decision-making on episodic memory. The literature on decision-making considers the two processes as inherently related and mutually influenced. Albeit this assumption is widely consensual, there is little empirical evidence supporting the impact of decision-making on memory. By contrasting the only published findings on this hypothesis (Murty et al., 2019), the current thesis paves the way to new research prospects.

The absence of difference in episodic memory following different moral decisions was also unexpected. Such an observation brings us to reconsider the dual-process model (Greene et al., 2008) and current heuristics and bias in decision-making research (Brust-Renck, 2021). Overall, this finding suggests that different moral decisions may arise from comparable levels of processing of the dilemma story content.

Finally, this thesis draws attention to the need for research in the field to define decision-making carefully. At this stage, it would be premature to consider that the relationship between memory and decision-making is unidirectional. Indeed, it seems more reasonable to assume that specific decision-making contexts, unexplored in this

thesis, may impact memory. A clear distinction between the conditions under which decision-making can affect memory would only be achieved with cautious definitions of the concept of interest. This prescription is of utmost relevance given the discrepancy and inconsistencies reported in the field (Johnson & Busemeyer, 2010).

The suggestion above is in accordance with findings from the literature on judgment and memory. As discussed in Experiment 5 and Experiment 6, Hastie and Park (1986) highlight divergent conclusions in research on the relationship between judgment and memory. They suggest that the incoherence gleaned from the literature is due to conceptual differences in experimental paradigms. Hastie and Park (1986) suggest a distinction between "online" and "memory-based" tasks. In memory-based tasks, the judgment involves retrieving concrete evidence from long-term episodic memory. In online tasks, the judgment is concomitant to the presentation of information on which the judgment should be based. Their results indicate that judgment affects memory in memory-based tasks but not online tasks. Given that judgment is a step of decision-making, it is worth considering whether this finding could be generalised to the wider decision-making process.

### 5.3.2 Methodological implications

The current thesis offers an original experimental paradigm for decision-making and memory research. While I am not the first to measure moral dilemma stories' recall (Arbuckle & Harsany, 1985), such a design has never been used to study the relationship between decision-making and memory. The majority of research on the topic used correlational methods or focused on memory's impact on decision-making. Hence, most designs measure decision-making rather than manipulating it. My

experimental paradigm allows manipulating decision-making (e.i, with a treatment and control condition) and comparing the impact of making different decisions (e.i, with the utilitarian and deontological groups). Overall, this experimental paradigm provides a new approach to decision-making and memory.

#### **5.4 Future directions**

I must admit that my research raises more questions than answers. The thesis aimed to provide evidence of the influence of decision-making on memory. I did not find such evidence. Therefore, the main recommendations for future research are related to the limitations of the thesis. Prospective studies should attempt to replicate the findings with a set of different dilemmas stories. As stated in Section 5.2, moral dilemmas are sensitive stimuli. Using a set of moral dilemmas would allow verifying whether the absence of effects reported here could be due to the story's characteristics.

Future research may use my experimental paradigm to test the impact of memory on decision-making. It could be done by reversing the order of the decision-making and the memory tasks. Such a reversal would challenge the experimental paradigm. Indeed, the literature provides strong evidence of the impact of memory on decision-making. If no effect were to be found, it would suggest that the experimental design may not be suitable to study the relationship between decision-making and memory.

Moreover, I suggest that future studies consider whether Hastie and Park 's (1985) distinction between online and memory-based judgment tasks could be

generalised to the entire decision-making process. Such research could address the issues of inconsistent findings within the field.

Another source of ambiguity could be the use of the traditional dichotomy between utilitarian and deontological decisions (Gawronski & Beer, 2017). In line with Conway and Gawronski (2013) process dissociation procedure, future research should compare memory recall of the story content under congruent vs incongruent decision outcomes. Finally, given the lack of research, it seems essential to further investigate the effect of decision-making using alternative research paradigms.

## **5.5 Conclusion**

This thesis investigates the long-running assumption of a reciprocal relationship between decision-making and memory. In Chapter 2, three experiments were conducted to develop a unique experimental paradigm bringing together decision-making and memory tasks. In Chapter 3, this experimental paradigm was applied to test the impact of decision-making and moral decision types on episodic memory. Chapter 4 provide a partial replication of the findings reported in Chapter 3 and further considers the role of learning in the relationship between decision-making and memory. The results revealed an absence of effect of decision-making on memory, regardless of the moral decision type and induced learning. Given the originality of the experimental paradigm, precautions must be taken when concluding on the relationship between decision-making and episodic memory. To conclude, this thesis questions the extent of decision-making impacts on memory. Whether this relationship may only be observed under specific decision-making contexts should be considered in future research.

## References

- Adjeroud, N., Besnard, J., Verny, C., Prundean, A., Scherer, C., Gohier, B.,  
Bonneau, D., Massioui, N. E., & Allain, P. (2017). Dissociation between  
decision-making under risk and decision-making under ambiguity in  
premanifest and manifest Huntington's disease. *Neuropsychologia*, *103*, 87–  
95. <https://doi.org/10.1016/j.neuropsychologia.2017.07.011>
- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B.  
(2008). Examining the testing effect with open- and closed-book tests. *Applied  
Cognitive Psychology*, *22*(7), 861–876. <https://doi.org/10.1002/acp.1391>
- Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes.  
*Psychological bulletin*, *82*(2), 261.
- Alarcón, D., Amián, J. G., & Sánchez-Medina, J. A. (2015). Enhancing emotion-  
based learning in decision-making under uncertainty. *Psicothema*, *27*(4), 368-  
373.
- Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque:  
Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*,  
*21*(4), 503. <https://doi.org/10.2307/1907921>
- American Psychiatric Association. (1994). Diagnostic and statistical manual of  
mental disorders (4th ed.).
- Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a  
replication study: Beyond statistical significance. *Psychological Methods*,  
*21*(1), 1–12. <https://doi.org/10.1037/met0000051>

- Arbuckle, T. Y., & Harsany, M. (1985). Adult age differences in recall of a moral dilemma under intentional, incidental and dual task instructions. *Experimental Aging Research*, 11(3), 175–177.  
<https://doi.org/10.1080/03610738508259183>
- Baddeley, A. D. (1997). *Human memory: Theory and practice* (Rev. ed). Psychology Press.
- Baron, J., Gürçay, B., Moore, A. B., & Starcke, K. (2012). Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese*, 189(1), 107-117. 10.1007/s11229-012-0121-z
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1–3), 7–15. [https://doi.org/10.1016/0010-0277\(94\)90018-3](https://doi.org/10.1016/0010-0277(94)90018-3)
- Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, 269(5227), 1115-1118.
- Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex. *Cerebral Cortex*, 6(2), 215–225.  
<https://doi.org/10.1093/cercor/6.2.215>
- Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision-making within the human prefrontal cortex. *Journal of neuroscience*, 18(1), 428-437.

- Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22(1), 23. <https://doi.org/10.2307/1909829> (Original work published in 1735)
- Blanchette, I., & Dunbar, K. (2002). Representational change and analogy: How analogical inferences alter target representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 672. 10.1037/0278-7393.28.4.672
- Bogartz, R. S. (1990). Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 138–148. <https://doi.org/10.1037/0278-7393.16.1.138>
- Brauer, M., & McClelland, G. (2005). L'utilisation des contrastes dans l'analyse des données: Comment tester les hypothèses spécifiques dans la recherche en psychologie ? *L'année psychologique*, 105(2), 273–305. <https://doi.org/10.3406/psy.2005.29696>
- Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of personality and social psychology*, 92(5), 938. Brust-Renck, P. G., Weldon, R. B., & Reyna, V. F. (2021). Judgment and Decision-making. In P. G. Brust-Renck, R. B. Weldon, & V. F. Reyna, *Oxford Research Encyclopedia of Psychology*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190236557.013.536>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438–448. <https://doi.org/10.3758/MC.36.2.438>

- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>
- Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00607>
- Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience & Biobehavioral Reviews*, 36(4), 1249–1264. <https://doi.org/10.1016/j.neubiorev.2012.02.008>
- Clark, L., Cools, R., & Robbins, T. W. (2004). The neuropsychology of ventral prefrontal cortex: decision-making and reversal learning. *Brain and cognition*, 55(1), 41-53.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Coman, D., Coman, A., & Hirst, W. (2013). Memory Accessibility and Medical Decision-Making for Significant Others: The Role of Socially Shared Retrieval-Induced Forgetting. *Frontiers in Behavioral Neuroscience*, 7. <https://doi.org/10.3389/fnbeh.2013.00072>
- Conway, P. (2022). Moral judgment. In D. E. Carlston, K. Johnson, & K. Hugenberg (Eds.), *The Oxford Handbook of Social Cognition*, 2nd Edition. Oxford University Press.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of*

*Personality and Social Psychology*, 104(2), 216–235.

<https://doi.org/10.1037/a0031021>

Coverdale, M. E., & Nairne, J. S. (2019). The mnemonic effect of choice.

*Psychonomic Bulletin & Review*, 26(4), 1310–1316.

<https://doi.org/10.3758/s13423-019-01575-z>

Craik, F. I. (2002). Levels of processing: Past, present... and future?. *Memory*, 10(5-6), 305-318. 10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)

Cubelli, R. (2010). A new taxonomy of memory and forgetting. *Forgetting*, 35-48.

Cui, J. F., Wang, Y., Shi, H. S., Liu, L. L., Chen, X. J., & Chen, Y. H. (2015). Effects of working memory load on uncertain decision-making: Evidence from the Iowa Gambling Task. *Frontiers in Psychology*, 6, 162.

Cunningham, S. J., Brady-Van den Bos, M., & Turk, D. J. (2011). Exploring the effects of ownership and choice on self-memory biases. *Memory*, 19(5), 449–461. <https://doi.org/10.1080/09658211.2011.584388>

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

Del Missier, F., Mäntylä, T., Hansson, P., Bruine de Bruin, W., Parker, A. M., & Nilsson, L.-G. (2013). The multifold relationship between memory and decision-making: An individual-differences study. *Journal of Experimental*

*Psychology: Learning, Memory, and Cognition*, 39(5), 1344–1364.

<https://doi.org/10.1037/a0032379>

Della Sala, S. (Ed.). (2010). *Forgetting*. Psychology Press.

Dong, L., Lee, J. Y., & Harvey, A. G. (2017). Do improved patient recall and the provision of memory support enhance treatment adherence? *Journal of Behavior Therapy and Experimental Psychiatry*, 54, 219–228.

<https://doi.org/10.1016/j.jbtep.2016.08.017>

Dougherty, M. R. P., Gronlund, S. D., & Gettys, C. F. (2003). Memory as a Fundamental Heuristic for Decision-making. In S. L. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (1st ed., pp. 125–164). Cambridge University Press.

<https://doi.org/10.1017/CBO9780511609978.007>

Dubrow, S., Eberts, E. A., & Murty, V. P. (2019). A common mechanism underlying choice's influence on preference and memory. *Psychonomic Bulletin & Review*, 26(6), 1958–1966. <https://doi.org/10.3758/s13423-019-01650-5>

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius Trans.). New York: Dover. (Original work published 1885)

Edwards, W. (1954). The theory of decision-making. *Psychological Bulletin*, 51(4), 380–417. <https://doi.org/10.1037/h0053870>

Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. J. (2013). *Robust Statistical Estimation*. Oxford University Press.

<https://doi.org/10.1093/oxfordhb/9780199934874.013.0019>

- Ernst, M., & Paulus, M. P. (2005). Neurobiology of Decision-making: A Selective Review from a Neurocognitive and Clinical Perspective. *Biological Psychiatry*, 58(8), 597–604. <https://doi.org/10.1016/j.biopsych.2005.06.004>
- Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The Role of Medial Prefrontal Cortex in Memory and Decision-making. *Neuron*, 76(6), 1057–1070. <https://doi.org/10.1016/j.neuron.2012.12.002>
- Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59(1), 255-278. [10.1146/annurev.psych.59.103006.093629](https://doi.org/10.1146/annurev.psych.59.103006.093629)
- Fellows, L. K. (2018). The Neuroscience of Human Decision-Making Through the Lens of Learning and Memory. In R. E. Clark & S. J. Martin (Eds.), *Behavioral Neuroscience of Learning and Memory* (Vol. 37, pp. 231–251). Springer International Publishing. [https://doi.org/10.1007/7854\\_2016\\_468](https://doi.org/10.1007/7854_2016_468)
- Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral cortex*, 15(1), 58-63.
- Fischhoff, B., & Broomell, S. B. (2020). Judgment and Decision-making. *Annual Review of Psychology*, 71(1), 331–355. <https://doi.org/10.1146/annurev-psych-010419-050747>
- Foot, P. (2002). *Virtues and Vices*. Oxford University Press.  
<https://doi.org/10.1093/0199252866.001.0001> (Original work published 1967)

- Gansler, D. A., Jerram, M. W., Vannorsdall, T. D., & Schretlen, D. J. (2011). Does the Iowa Gambling Task Measure Executive Function? *Archives of Clinical Neuropsychology*, 26(8), 706–717. <https://doi.org/10.1093/arclin/acr082>
- Gawronski, B., & Beer, J. S. (2016). What makes moral dilemma judgments “utilitarian” or “deontological”? *Social Neuroscience*, 1–7. <https://doi.org/10.1080/17470919.2016.1248787>
- Giguère, G., & Love, B. C. (2013). Limits in decision-making arise from limits in memory retrieval. *Proceedings of the National Academy of Sciences*, 110(19), 7613–7618. <https://doi.org/10.1073/pnas.1219674110>
- Gilovich, T., & Griffin, D. (2012). Introduction – Heuristics and Biases: Then and Now. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (1st ed., pp. 1–18). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808098.002>
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107(3), 1144–1154. <https://doi.org/10.1016/j.cognition.2007.11.004>
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron*, 44(2), 389–400. <https://doi.org/10.1016/j.neuron.2004.09.027>
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>

- Groman, S. M., Ikemoto, S., Rushworth, M., Taylor, J. R., & Whelan, R. (2020). Introducing the PLOS ONE Collection on the neuroscience of reward and decision-making. *PLOS ONE*, *15*(10), e0240505.  
<https://doi.org/10.1371/journal.pone.0240505>
- Gupta, R., Duff, M. C., Denburg, N. L., Cohen, N. J., Bechara, A., & Tranel, D. (2009). Declarative memory is critical for sustained advantageous complex decision-making. *Neuropsychologia*, *47*(7), 1686–1693.  
<https://doi.org/10.1016/j.neuropsychologia.2009.02.007>
- Gutbrod K, Krouzel C, Hofer H, Müri R, Perrig W, Ptak R. (2006) Decision-making in amnesia: do advantageous decisions require conscious knowledge of previous behavioural choices? *Neuropsychologia*.;44(8):1315-24. doi: 10.1016/j.neuropsychologia.2006.01.014.
- Hanawalt, N.G., 1937. Memory trace for figures in recall and recognition. *Arch. Psychol.* 31 (216), 5–89.
- Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*(22), 1315–1318.  
[https://doi.org/10.1016/S0960-9822\(00\)80053-6](https://doi.org/10.1016/S0960-9822(00)80053-6)
- Harrison, B. J., Pujol, J., López-Solà, M., Hernández-Ribas, R., Deus, J., Ortiz, H., Soriano-Mas, C., Yücel, M., Pantelis, C., & Cardoner, N. (2008). Consistency and functional specialization in the default mode brain network. *Proceedings of the National Academy of Sciences*, *105*(28), 9781–9786.  
<https://doi.org/10.1073/pnas.0711791105>

- Hastie, R. (2001). Problems for Judgment and Decision-making. *Annual Review of Psychology*, 52(1), 653–683. <https://doi.org/10.1146/annurev.psych.52.1.653>
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93(3), 258–268. <https://doi.org/10.1037/0033-295X.93.3.258>
- Heene, M., & Ferguson, C. J. (2017). Psychological Science's Aversion to the Null, and Why Many of the Things You Think Are True, Aren't. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological Science Under Scrutiny* (pp. 34–52). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119095910.ch3>
- Hesterberg, T. C. (2015). What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum. *The American Statistician*, 69(4), 371–386. <https://doi.org/10.1080/00031305.2015.1089789>
- Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, 19(3), 290–304. <https://doi.org/10.1080/09658211.2011.560121>
- Horner, A. J. (2019). Retrieving to remember: In theory and practice. *York Scholarship of Teaching and Learning Journal*, 3, 23-32.
- Iglesias-Parro, S., & Gómez-Ariza, C. J. (2006). Biasing decision-making by means of retrieval practice. *European Journal of Cognitive Psychology*, 18(6), 899–908. <https://doi.org/10.1080/09541440500501189>
- Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision-

making. *NeuroImage*, 32(2), 978–987.

<https://doi.org/10.1016/j.neuroimage.2006.02.018>

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541. [https://doi.org/10.1016/0749-596X\(91\)90025-F](https://doi.org/10.1016/0749-596X(91)90025-F)

Johnson, J. G., & Busemeyer, J. R. (2010). Decision-making under risk and uncertainty. *WIREs Cognitive Science*, 1(5), 736–749.

<https://doi.org/10.1002/wcs.76>

Johnson, P. E. (2013). *Monte Carlo Analysis in Academic Research*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199934874.013.0022>

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430–454.

[https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3)

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263. <https://doi.org/10.2307/1914185>

Kant, I., (2001) *Critique of pure reason* (Adorno, T. W., Tans). Polity. (Original work published in 1785).

Karpicke, J. D. (2017). Retrieval-Based Learning: A Decade of Progress. In *Learning and Memory: A Comprehensive Reference* (pp. 487–514). Elsevier.

<https://doi.org/10.1016/B978-0-12-809324-5.21055-9>

Karpicke, J. D., & Aue, W. R. (2015). The Testing Effect Is Alive and Well with Complex Materials. *Educational Psychology Review*, 27(2), 317–326.

<https://doi.org/10.1007/s10648-015-9309-3>

- Katzman, P. L., & Hartley, C. A. (2020). The value of choice facilitates subsequent memory across development. *Cognition*, *199*, 104239.  
<https://doi.org/10.1016/j.cognition.2020.104239>
- Kent, P. (2013). The Evolution of the Wechsler Memory Scale: A Selective Review. *Applied Neuropsychology: Adult*, *20*(4), 277–291.  
<https://doi.org/10.1080/09084282.2012.689267>
- Knight Frank, H. (1921). Risk, uncertainty and profit. книга.
- Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and decision making*, *8*(5), 527.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 609–622. <https://doi.org/10.1037/0278-7393.32.3.609>
- Krueger, J. I. (Ed.). (2011). *Social judgment and decision-making*. Psychology Press.
- Levy, I. (2017). Neuroanatomical Substrates for Risk Behavior. *The Neuroscientist*, *23*(3), 275–286. <https://doi.org/10.1177/1073858416672414>
- Lichtenstein, M., & Srull, T. K. (1987). Processing objectives as a determinant of the relationship between recall and judgment. *Journal of Experimental Social Psychology*, *23*(2), 93–118. [https://doi.org/10.1016/0022-1031\(87\)90027-8](https://doi.org/10.1016/0022-1031(87)90027-8)
- Little, T. D. (Ed.). (2013). *The Oxford Handbook of Quantitative Methods in Psychology, Vol. 1* (1st ed.). Oxford University Press.  
<https://doi.org/10.1093/oxfordhb/9780199934874.001.0001>

- MacPherson, S. E., Della Sala, S., Cox, S. R., Girardi, A., & Iveson, M. H. (2015). *Handbook of Frontal Lobe Assessment (DRAFT)*. Oxford University Press.  
<https://doi.org/10.1093/med:psych/9780199669523.001.0001>
- Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, *21*(3), 629–636.  
<https://doi.org/10.3758/s13423-013-0542-9>
- Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2017). The Role of Memory in Distinguishing Risky Decisions from Experience and Description. *Quarterly Journal of Experimental Psychology*, *70*(10), 2048–2059.  
<https://doi.org/10.1080/17470218.2016.1220608>
- Markant, D., DuBrow, S., Davachi, L., & Gureckis, T. M. (2014). Deconstructing the effect of self-directed study on episodic memory. *Memory & Cognition*, *42*(8), 1211–1224. <https://doi.org/10.3758/s13421-014-0435-9>
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The Read-Recite-Review Study Strategy: Effective and Portable. *Psychological Science*, *20*(4), 516–522. <https://doi.org/10.1111/j.1467-9280.2009.02325.x>
- McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment*, *9*(3), 199–213.  
[https://doi.org/10.1016/0740-5472\(92\)90062-S](https://doi.org/10.1016/0740-5472(92)90062-S)
- Memon, M. A., Ting, H., Cheah, J.-H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample Size for Survey Research: Review and Recommendations.

*Journal of Applied Structural Equation Modeling*, 4(2), i–xx.

[https://doi.org/10.47263/JASEM.4\(2\)01](https://doi.org/10.47263/JASEM.4(2)01)

Mill, J. S. (1998). *Utilitarianism* (Crisp, R., Trans.). Oxford University Press. (Original work published in 1861)

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who Shalt Not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment. *Psychological Science*, 19(6), 549–557.

<https://doi.org/10.1111/j.1467-9280.2008.02122.x>

Morelli, M., Casagrande, M., & Forte, G. (2021). Decision-making: A Theoretical Review. *Integrative Psychological and Behavioral Science*.

<https://doi.org/10.1007/s12124-021-09669-x>

Morris, J., Swier-Vosnos, A., Woodworth, C., Umfleet, L. G., Czipri, S., & Kopald, B. (2014). Development of Alternate Paragraphs for the Logical Memory Subtest of the Wechsler Memory Scale-IV. *Applied Neuropsychology: Adult*, 21(2), 143–147. <https://doi.org/10.1080/09084282.2013.780172>

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The Simple Act of Choosing Influences Declarative Memory. *Journal of Neuroscience*, 35(16), 6255–6264.

<https://doi.org/10.1523/JNEUROSCI.4181-14.2015>

Murty, V. P., DuBrow, S., & Davachi, L. (2019). Decision-making Increases Episodic Memory via Postencoding Consolidation. *Journal of Cognitive Neuroscience*, 31(9), 1308–1317. [https://doi.org/10.1162/jocn\\_a\\_01321](https://doi.org/10.1162/jocn_a_01321)

Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of*

*Experimental Psychology: General*, 145(5), 548–558.

<https://doi.org/10.1037/xge0000158>

Navarro, D., (n.d.). Learning Statistics with R, <https://learningstatisticswithr.com/>

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory:

Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 263–273.

<https://doi.org/10.1037/0278-7393.33.2.263>

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision-making:

A critical review. *Behavioral and Brain Sciences*, 37(1), 1–19.

<https://doi.org/10.1017/S0140525X12003214>

Ni, Y., & Li, J. (2021). Neural mechanisms of social learning and decision-making.

*Science China Life Sciences*, 64(6), 897–910. <https://doi.org/10.1007/s11427-020-1833-8>

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing

learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, 14(2), 187–193. <https://doi.org/10.3758/BF03194050>

Racsmány, M., Szöllősi, Á., & Bencze, D. (2018). Retrieval practice makes

procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(1), 157–166. <https://doi.org/10.1037/xlm0000423>

Redish, A. D. (2013). *The mind within the brain: How we make decisions and how those decisions go wrong*. Oxford university press.

- Redish, A. D., & Mizumori, S. J. Y. (2015). Memory and decision-making. *Neurobiology of Learning and Memory*, 117, 1–3. <https://doi.org/10.1016/j.nlm.2014.08.014>
- Redlawsk, D. P., & Lau, R. R. (2013). *Behavioral Decision-Making*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199760107.013.0005>
- Reisberg, D., & Heuer, F. (2004). Memory for Emotional Events. In D. Reisberg & P. Hertel (Eds.), *Memory and Emotion* (pp. 1–41). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195158564.003.0001>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger, H. L., & Karpicke, J. D. (2006a). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roediger, H. L., & Karpicke, J. D. (2006b). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten Benefits of Testing and Their Applications to Educational Practice. In *Psychology of Learning and Motivation* (Vol. 55, pp. 1–36). Elsevier. <https://doi.org/10.1016/B978-0-12-387691-1.00001-6>

- Rogers, R. (1999). Dissociable Deficits in the Decision-Making Cognition of Chronic Amphetamine Abusers, Opiate Abusers, Patients with Focal Damage to Prefrontal Cortex, and Tryptophan-Depleted Normal Volunteers Evidence for Monoaminergic Mechanisms. *Neuropsychopharmacology*, *20*(4), 322–339. [https://doi.org/10.1016/S0893-133X\(98\)00091-8](https://doi.org/10.1016/S0893-133X(98)00091-8)
- Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences* (2d ed). Holt, Rinehart and Winston.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge University Press.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (1999). *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511804403>
- Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734–760. <https://doi.org/10.1037/0033-295X.103.4.734>
- Schacter, D. L., Benoit, R. G., De Brigard, F., & Szpunar, K. K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory*, *117*, 14–21. <https://doi.org/10.1016/j.nlm.2013.12.008>

- Schaper, M. L., Mieth, L., & Bell, R. (2019). Adaptive memory: Source memory is positively associated with adaptive social decision-making. *Cognition*, *186*, 7–14. <https://doi.org/10.1016/j.cognition.2019.01.014>
- Schlichting, M. L., & Preston, A. R. (2017). The Hippocampus and Memory Integration: Building Knowledge to Navigate Future Decisions. In D. E. Hannula & M. C. Duff (Eds.), *The Hippocampus from Cells to Systems* (pp. 405–437). Springer International Publishing. [https://doi.org/10.1007/978-3-319-50406-3\\_13](https://doi.org/10.1007/978-3-319-50406-3_13)
- Scholl, J., & Klein-Flügge, M. (2018). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. *Behavioural Brain Research*, *355*, 56–75. <https://doi.org/10.1016/j.bbr.2017.09.050>
- Schwarz, N., & Vaughn, L. (2002). The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 103-119). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511808098.007
- Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1712–1725. <https://doi.org/10.1037/a0033569>

- Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The Critical Importance of Retrieval—And Spacing—For Learning. *Psychological Science, 27*(2), 223–230. <https://doi.org/10.1177/0956797615617778>
- Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive psychology, 31*(3), 307-346. 10.1006/cogp.1996.0019
- Takemura, K. (2021). Decision-Making Phenomenon and Behavioral Decision Theory. In K. Takemura, *Behavioral Decision Theory* (pp. 3–13). Springer Singapore. [https://doi.org/10.1007/978-981-16-5453-4\\_1](https://doi.org/10.1007/978-981-16-5453-4_1)
- Taylor K, Rodriguez J. Visual Discrimination. [Updated 2021 Sep 21]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK563242/>
- The Scottish Voice (2011), Cerevoice Stuart version 4.0.1 [software], <https://www.thescottishvoice.org.uk/>
- Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal, 94*(6), 1395. <https://doi.org/10.2307/796133>
- Tremel, J. J., Ortiz, D. M., & Fiez, J. A. (2018). Manipulating memory efficacy affects the behavioral and neural profiles of deterministic learning and decision-making. *Neuropsychologia, 114*, 214–230. <https://doi.org/10.1016/j.neuropsychologia.2018.04.023>
- Tulving, E. (1972). 12. Episodic and semantic memory. *Organization of memory*/Eds E. Tulving, W. Donaldson, NY: Academic Press, 381-403.

- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232.  
[https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157), 1124-1131.
- van der Pligt, J. (1996). Judgement and Decision-making. In *Applied Social Psychology* (pp. 30–64). SAGE Publications Ltd.  
<https://doi.org/10.4135/9781446250556.n2>
- van Gog, T., & Sweller, J. (2015). Not New, but Nearly Forgotten: The Testing Effect Decreases or even Disappears as the Complexity of Learning Materials Increases. *Educational Psychology Review*, 27(2), 247–264.  
<https://doi.org/10.1007/s10648-015-9310-x>
- Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior* (60th anniversary ed). Princeton University Press. (Original worked published in 1944).
- Voss, J. L., Warren, D. E., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Spontaneous revisitation during visual exploration as a link among strategic behavior, learning, and the hippocampus. *Proceedings of the National Academy of Sciences*, 108(31), E402-E409.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804.  
<https://doi.org/10.3758/BF03194105>

- Watanabe, T., & Soraci, S. A. (2004). The self-choice effect from a multiple-cue perspective. *Psychonomic Bulletin & Review*, 11(1), 168–172.  
<https://doi.org/10.3758/BF03206478>
- Wechsler, D. (2008), *WAIS-IV Administration and Scoring Manual*, San Antonio, TX: The Psychological Corporation
- Wechsler, D. (2009), *WMS-IV Administration and Scoring Manual*, San Antonio, TX, The Psychological Corporation
- Weiss, J. W., & Weiss, D. J. (2008). *A Science of Decision-making*. Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780195322989.001.0001>
- Wilson, M. J., & Vassileva, J. (2018). Decision-Making Under Risk, but Not Under Ambiguity, Predicts Pathological Gambling in Discrete Types of Abstinent Substance Users. *Frontiers in Psychiatry*, 9, 239.  
<https://doi.org/10.3389/fpsy.2018.00239>
- Wimmer, G. E., Shohamy D., (2012) Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*.338(6104):270-3.  
doi: 10.1126/science.1223252
- Zhou, Y. & Skidmore, S. (2018). A Reassessment of ANOVA Reporting Practices: A Review of Three APA Journals. *Journal of Methods and Measurement in the Social Sciences*, 8(1). <https://doi.org/10.2458/v8i1.22019>

## Appendixes

### Appendix 1

*Statistical analysis R markdown*<sup>7</sup>

PhD\_thesis\_statistical\_analyses

Carolane Helin, PhD Student in Psychology, The University of Edinburgh

```
## Loading required package: carData
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
```

```
##
```

```
##      between, first, last
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##      recode
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

---

<sup>7</sup> This document is available under a Rmarkdown format along with a package containing the dataset is available at [https://osf.io/uj3at/?view\\_only=a47b2a6def8849bdf7ec8b03ea1ad66](https://osf.io/uj3at/?view_only=a47b2a6def8849bdf7ec8b03ea1ad66).

```
##  
## Attaching package: 'gridExtra'  
## The following object is masked from 'package:dplyr':  
##  
##      combine  
## Loading required package: lpSolve  
##  
## Attaching package: 'janitor'  
## The following objects are masked from 'package:stats':  
##  
##      chisq.test, fisher.test  
##  
## Attaching package: 'kableExtra'  
## The following object is masked from 'package:dplyr':  
##  
##      group_rows  
##  
## Attaching package: 'psych'  
## The following objects are masked from 'package:ggplot2':  
##  
##      %+%, alpha
```

```
## The following object is masked from 'package:effectsize':  
##  
##   phi  
## The following object is masked from 'package:car':  
##  
##   logit  
##  
## Attaching package: 'purrr'  
## The following object is masked from 'package:data.table':  
##  
##   transpose  
## The following object is masked from 'package:car':  
##  
##   some  
##  
## Attaching package: 'rstatix'  
## The following object is masked from 'package:janitor':  
##  
##   make_clean_names  
## The following objects are masked from 'package:effectsize':  
##  
##   cohens_d, eta_squared
```

```
## The following object is masked from 'package:stats':  
##  
##   filter  
##  
## Attaching package: 'reshape2'  
## The following objects are masked from 'package:data.table':  
##  
##   dcast, melt  
## Learn more about sjPlot with 'browseVignettes("sjPlot")'.  
## Loading required package: MASS  
##  
## Attaching package: 'MASS'  
## The following object is masked from 'package:rstatix':  
##  
##   select  
## The following object is masked from 'package:dplyr':  
##  
##   select  
## Loading required package: lme4  
## Loading required package: Matrix  
## Loading required package: lavaan  
## This is lavaan 0.6-9  
## lavaan is FREE software! Please report any bugs.
```

```
##
```

```
## Attaching package: 'lavaan'
```

```
## The following object is masked from 'package:psych':
```

```
##
```

```
## cor2cov
```

```
## Loading required package: parallel
```

```
## Loading required package: PearsonDS
```

Experiment 1: Investigating the Retention of a Piece of Prose Under Long-Term

Forgetting and Retrieval Practice

```
E1_data_illustration <- data.frame ('Int' = c(1, 1, 2, 2, 3, 3, 4,
4, 5, 5, 6, 6, 7, 7),
```

```

'RT' = c('Repreated Recall', 'Single Recall',
'Repreated Recall', 'Single
Recall',
```

```

'Repreated Recall', 'Single
Recall',
```

```

'Repreated Recall', 'Single
Recall',
```

```

'Repreated Recall', 'Single
Recall',
```

```

'Repreated Recall', 'Single
Recall',
```

```

'Repreated Recall', 'Single
```

```
Recall'),
```

```
      'R' = c(0.88, 0.79,
```

```
          0.85, 0.71,
```

```
          0.84, 0.65,
```

```
          0.84, 0.60,
```

```
          0.836, 0.58,
```

```
          0.834, 0.57,
```

```
          0.83, 0.56))
```

```
E1_illustration_plot<-ggplot(E1_data_illustration, aes(x = Int,
```

```
          y = R,
```

```
          ylim(0.5, 0.9),
```

```
          group= RT,
```

```
          shape= RT)) +
```

```
  geom_line(aes(linetype=RT)) +
```

```
  xlab('Retention interval') +
```

```
  ylab('Proportion correctly recalled') +
```

```
  scale_x_discrete(limit = c(1, 2, 4, 7),
```

```
                  labels = c('1 Day', '1 Week', '1 Month', '2
```

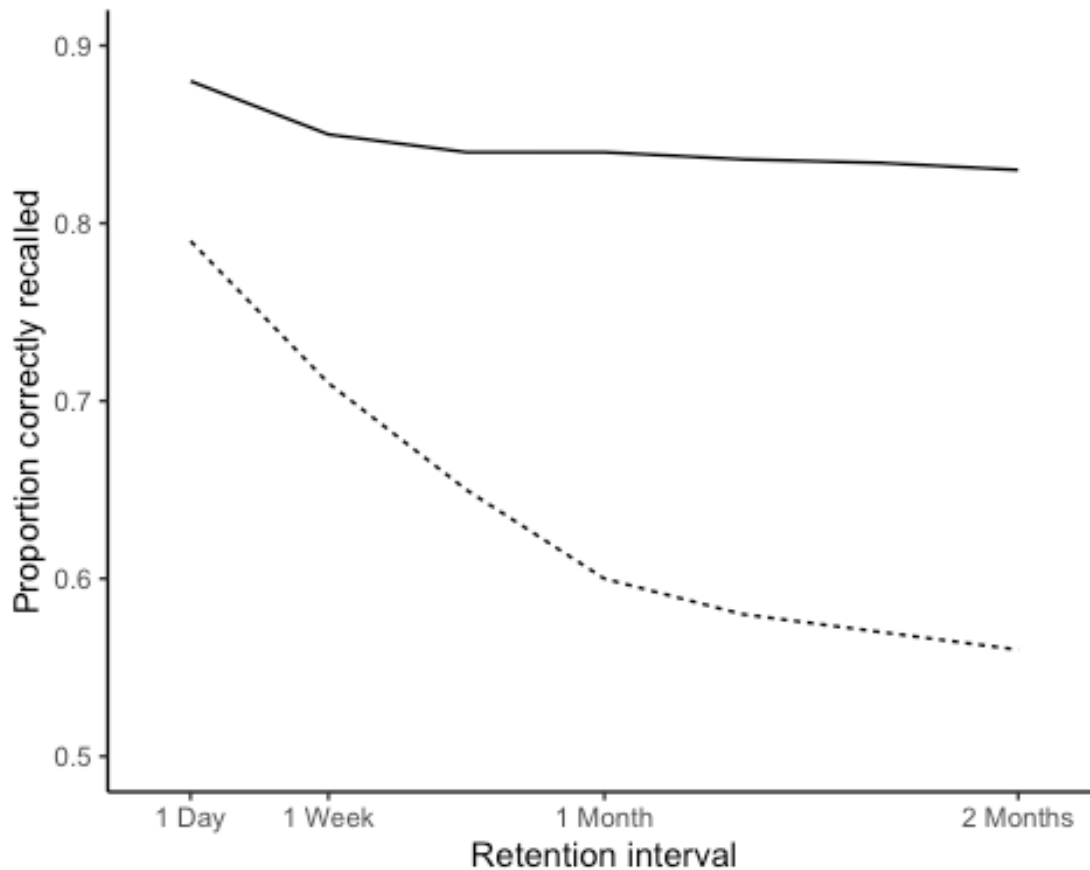
```
Months')) +
```

```
  expand_limits(y=c(0.5,0.9))+
```

```
  theme_classic()+
```

```
  theme(legend.position = 'none')
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale*_continuous()`?
E1_illustration_plot
```



Experiment 1a: The Forgetting of a Piece of Prose

```
data(Experiment1, package = 'PhDCarolane')
Experiment1[, c('Subject_ID', 'Retrieval_Practice', 'Delay',
               'Experimenter', 'Administration')] <- lapply(Experiment1[,
               c('Subject_ID', 'Retrieval_Practice', 'Delay', 'Experimenter',
               'Administration')], factor)
Experiment1$Delay <- factor(Experiment1$Delay, levels
```

```

=c('Immediate', 'Day', 'Week', 'Month'))
Experiment1[, c('DSF', 'DSB', 'Total_Recall')] <-
sapply(Experiment1[, c('DSF', 'DSB', 'Total_Recall')], as.numeric)
Experiment1a<- subset(Experiment1, Retrieval_Practice ==
'Without_RP',
                      select = c('Subject_ID', 'Delay',
'Administration', 'Experimenter', 'Total_Recall', 'DSF', 'DSB'))
Estimation of the required sample size
E1a_pwr<-pwr.anova.test(k= 3, f= 0.40, sig.level = 0.05, power= 0.8)
E1a_data_ctrl<-subset(Experiment1a, Delay == 'Immediate', select =
c('Subject_ID', 'Total_Recall', 'DSF', 'DSB'))
E1a_data_ctrl<- rename(E1a_data_ctrl,
'Immediate_Recall'='Total_Recall')
E1a_data_ctrl_delay<- subset(Experiment1a, Delay != 'Immediate')
E1a_data_ctrl_delay$Delay<-factor(E1a_data_ctrl_delay$Delay, levels
=c('Day', 'Week', 'Month'))
E1a_data_ctrl$Condition<- E1a_data_ctrl_delay$Delay
head(E1a_data_ctrl)
## # A tibble: 6 x 5
##   Subject_ID Immediate_Recall   DSF   DSB Condition
##   <fct>                <dbl> <dbl> <dbl> <fct>
## 1 20                    11    11    11 Day
## 2 23                    2     4     9 Day

```

```
## 3 30          12   10    9 Week
## 4 33          4   10    7 Week
## 5 34          12   12   12 Week
## 6 35          5   10    7 Week
```

Effect of delay groups on immediate recall

### Descriptive statistics

```
E1a_imm_descr <-E1a_data_ctrl %>%
  group_by(Condition) %>%
  summarise( N = n(), M = mean(Immediate_Recall), SD =
sd(Immediate_Recall),
            Min = min(Immediate_Recall),
            Median = median(Immediate_Recall),
            Max = max(Immediate_Recall))
E1a_imm_descr<-APA.descr.table(E1a_imm_descr)
E1a_imm_descr%>%
  tab_header(title = 'Table: Story C immediate recall for each delay
group')%>%
  tab_options(table.border.top.color = "White",
            source_notes.font.size = 12,
            heading.align = "left",
            heading.padding = px(30))
```

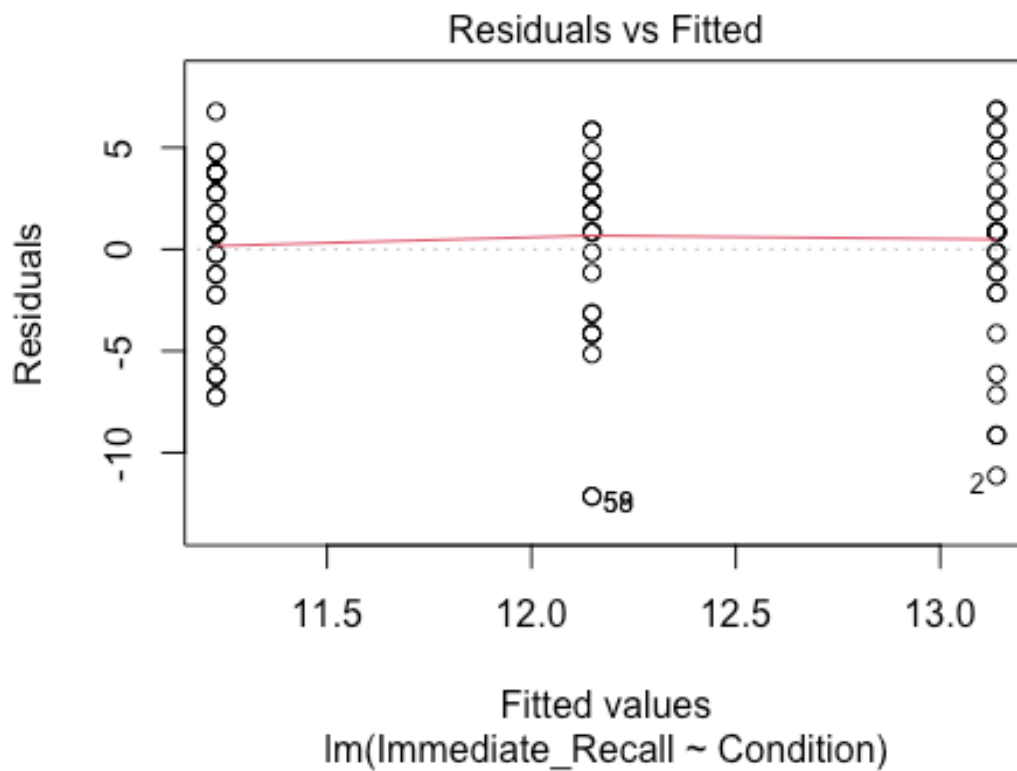
### Inferential statistics

*Assumption checking*

```

contrasts(E1a_data_ctrl$Condition)<- contr.sum
E1a_model_imm<- lm(Immediate_Recall ~ Condition, data =
E1a_data_ctrl)
plot(E1a_model_imm, which=1)

```

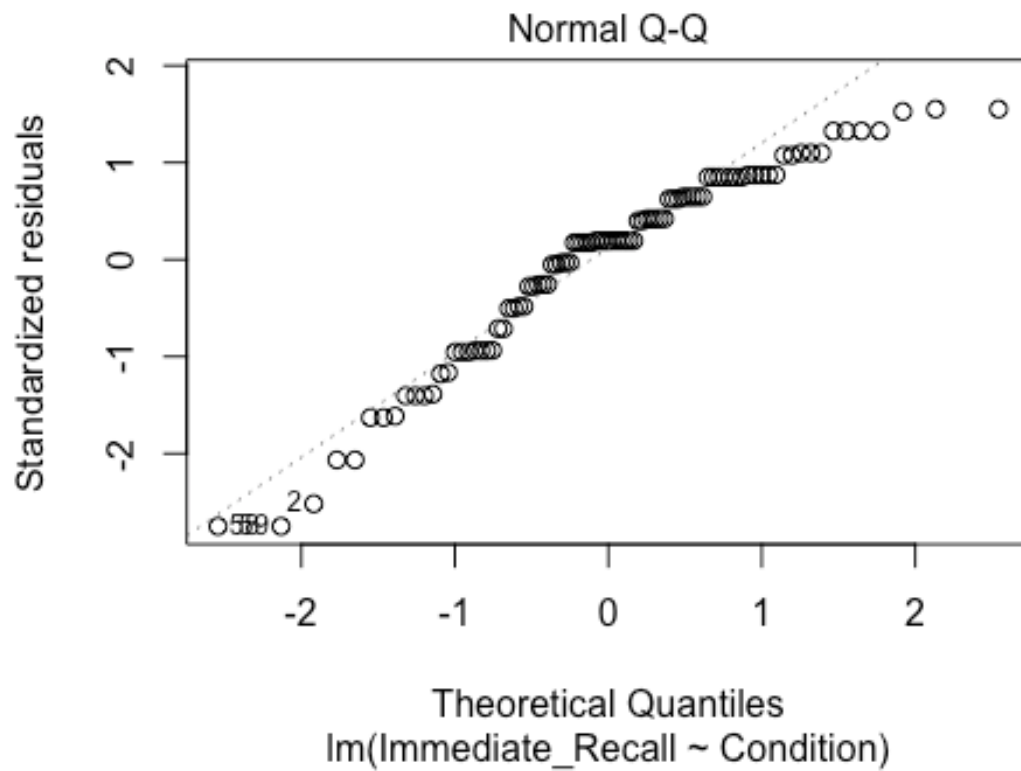


```

E1a_imm_levene<-leveneTest(Immediate_Recall ~ Condition, data =
E1a_data_ctrl)
E1a_imm_levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)

```

```
## group 2 0.0969 0.9078
## 88
plot(E1a_model_imm, which= 2)
```



```
E1a_imm_SW<-shapiro.test(resid(E1a_model_imm))
```

```
E1a_imm_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: resid(E1a_model_imm)
```

```
## W = 0.93862, p-value = 0.0003266
```

### Kruskal-Wallis rank-sum test

```
E1a_imm_KW<-kruskal.test(Immediate_Recall ~ Condition, data=
E1a_data_ctrl)
E1a_imm_KW
##
## Kruskal-Wallis rank sum test
##
## data: Immediate_Recall by Condition
## Kruskal-Wallis chi-squared = 3.5718, df = 2, p-value = 0.1676
```

Effect of delay groups on DSF

### Descriptive statistics

```
E1a_DSF_descr <-E1a_data_ctrl %>%
  group_by(Condition) %>%
  summarise( N = n(), M = mean(DSF), SD = sd(DSF),
            Min = min(DSF),
            Median = median(DSF),
            Max = max(DSF))
E1a_DSF_descr<-APA.descr.table(E1a_DSF_descr)
E1a_DSF_descr%>%
  tab_header(title = 'Table: DSF scores for each delay group')%>%
  tab_options(table.border.top.color = "White",
            source_notes.font.size = 12,
```

```
heading.align = "left",
```

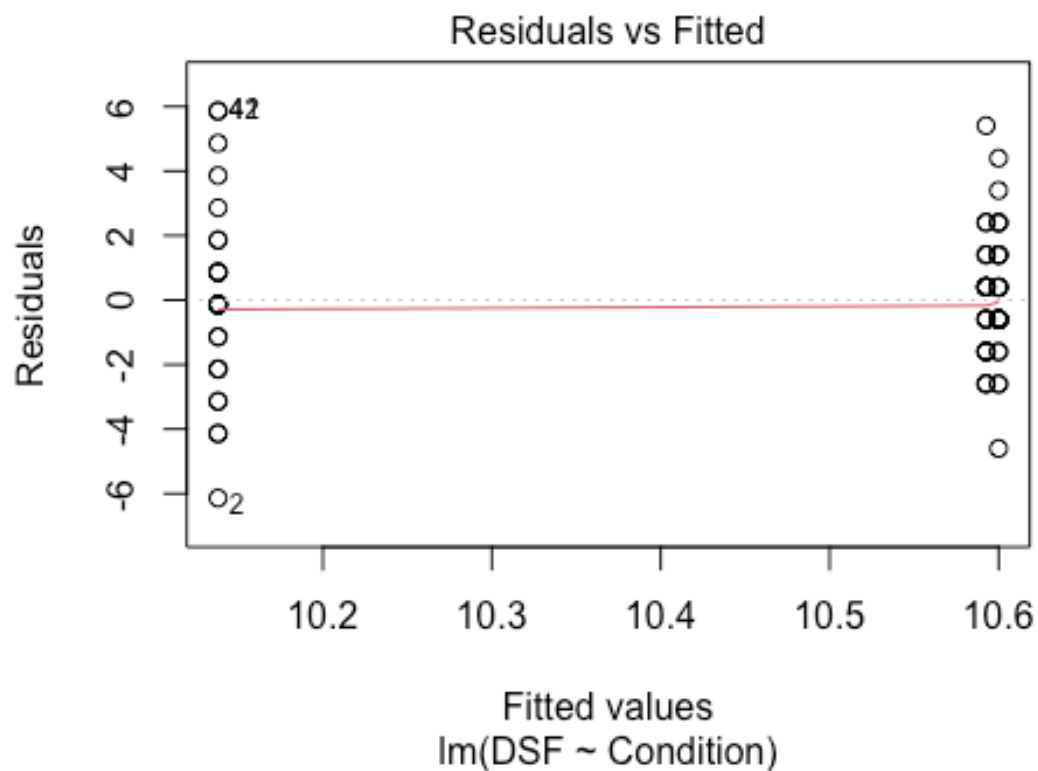
```
heading.padding = px(30))
```

### Inferential statistics

```
E1a_model_DSF<- lm(DSF ~ Condition, data = E1a_data_ctrl)
```

### Assumption checking

```
plot(E1a_model_DSF, which=1)
```



```
E1a_DSF_Levene<-leveneTest(DSF ~ Condition, data= E1a_data_ctrl)
```

```
E1a_DSF_Levene
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

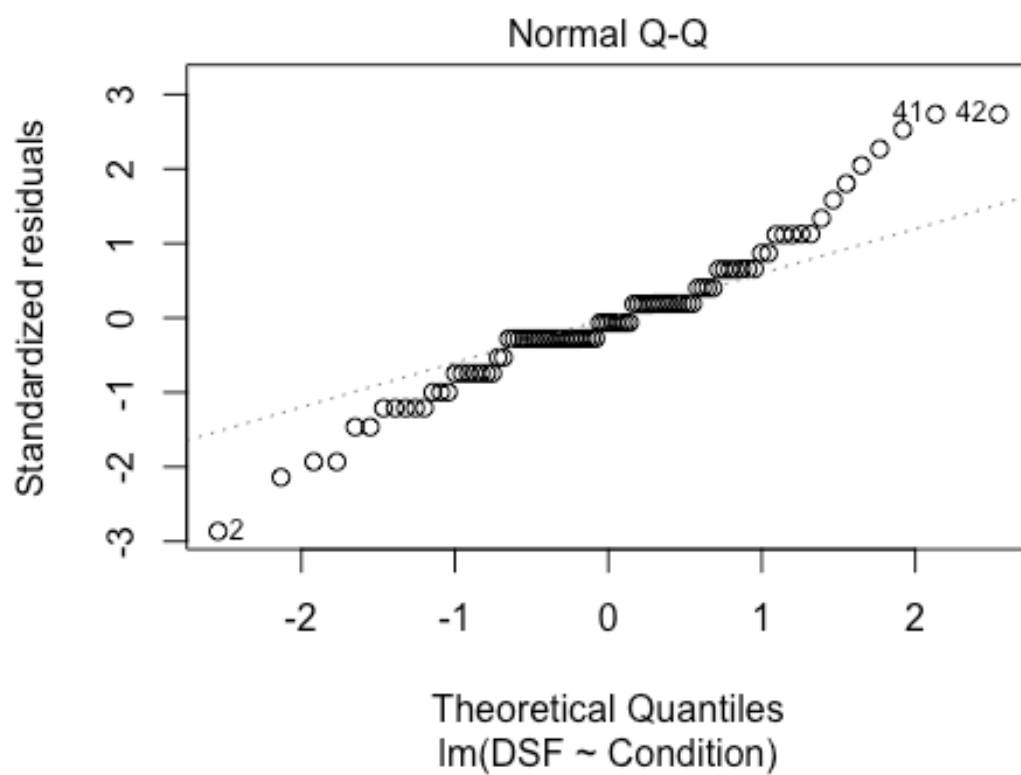
```
## group 2 2.8769 0.06162 .
```

```
## 88
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(E1a_model_DSF, which= 2)
```



```
E1a_DSF_SW<-shapiro.test(resid(E1a_model_DSF))
```

```
E1a_DSF_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```

## data: resid(E1a_model_DSF)

## W = 0.95995, p-value = 0.006831

Kruskal-Wallis rank-sum test

E1a_DSF_KW<-kruskal.test(DSF ~ Condition, data= E1a_data_ctrl)

E1a_DSF_KW

##

## Kruskal-Wallis rank sum test

##

## data: DSF by Condition

## Kruskal-Wallis chi-squared = 1.1245, df = 2, p-value = 0.5699

Effect of delay groups on DSB

Descriptive statistics

E1a_DSB_descr <-E1a_data_ctrl %>%

  group_by(Condition) %>%

  summarise( N = n(), M = mean(DSB), SD = sd(DSB),

             Min = min(DSB),

             Median = median(DSB),

             Max = max(DSB))

E1a_DSB_descr<-APA.descr.table(E1a_DSB_descr)

E1a_DSB_descr%>%

  tab_header(title = 'Table: DSB scores for each delay group')%>%

  tab_options(table.border.top.color = "White",

              source_notes.font.size = 12,

```

```
heading.align = "left",
```

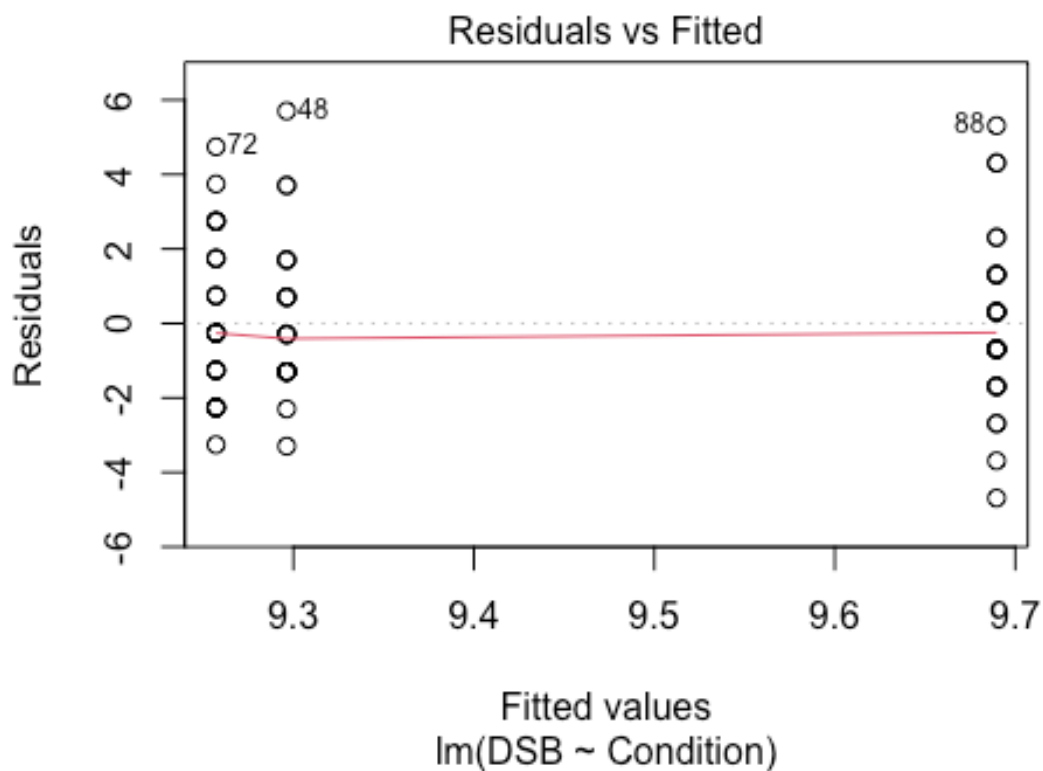
```
heading.padding = px(30))
```

### Inferential statistics

```
E1a_model_DSB<- lm(DSB ~ Condition, data = E1a_data_ctrl)
```

*Assumption checking*

```
plot(E1a_model_DSB, which=1)
```



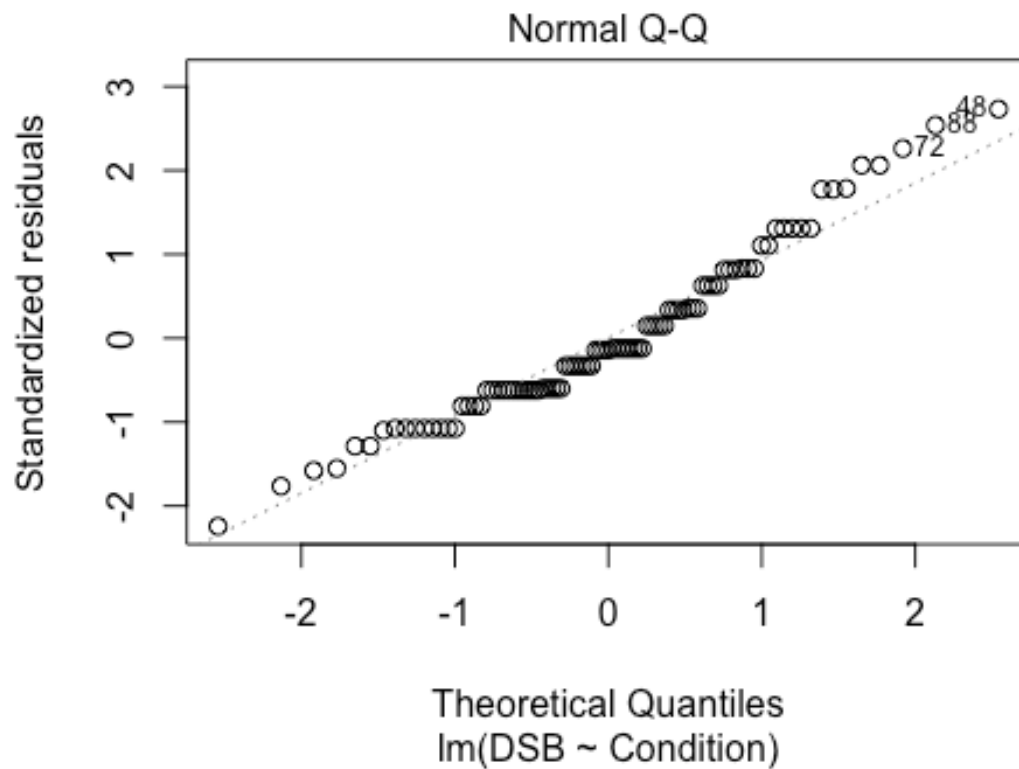
```
E1a_DSB_Levene<-leveneTest(DSB ~ Condition, data = E1a_data_ctrl)
```

```
E1a_DSB_Levene
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group 2 0.2051 0.8149
##      88
plot(E1a_model_DSB, which= 2)
```



```
E1a_DSB_SW<-shapiro.test(E1a_data_ctrl$DSB)
```

```
E1a_DSB_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: E1a_data_ctrl$DSB
```

```
## W = 0.94836, p-value = 0.001234
```

*Kruskal-Wallis rank-sum test*

```
E1a_DSB_KW<-kruskal.test(DSB ~ Condition, data= E1a_data_ctrl)
```

```
E1a_DSB_KW
```

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: DSB by Condition
```

```
## Kruskal-Wallis chi-squared = 1.1083, df = 2, p-value = 0.5746
```

Effect of delay groups on Story C retention

*N.B: As a matter of data protection, participants' demographics are not provided in the dataset*

```
E1a_data_delay<-subset(Experiment1a, Delay != 'Immediate', select =  
c('Subject_ID', 'Delay', 'Experimenter', 'Total_Recall'))
```

```
E1a_data_delay$Delay<- factor(E1a_data_delay$Delay, levels  
=c('Day', 'Week', 'Month'))
```

**Descriptive statistics**

```
E1a_descriptive_statistics <-E1a_data_delay %>%
```

```
  group_by(Delay) %>%
```

```
  summarise( N = n(), M = mean(Total_Recall), SD =
```

```
sd(Total_Recall),
```

```
    Min = min(Total_Recall),
```

```
    Median = median(Total_Recall),
```

```
    Max = max(Total_Recall))
```

```
E1a_descriptive_statistics <-
```

```
APA.descr.table(E1a_descriptive_statistics)
```

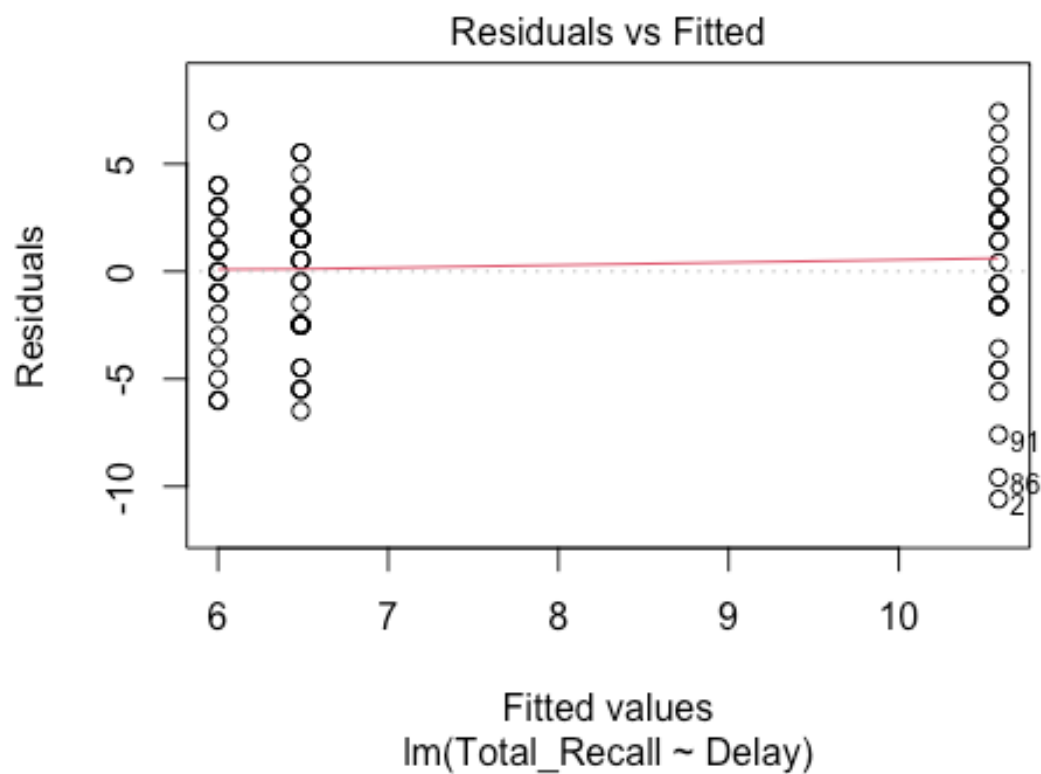
```
E1a_descriptive_statistics
```

**Inferential statistics**

```
E1a_model_delay <- lm(Total_Recall ~ Delay, data = E1a_data_delay)
```

*Assumption checking*

```
plot(E1a_model_delay, which=1)
```



```
E1a_delay_Levene <- leveneTest(Total_Recall ~ Delay, data =
```

```
E1a_data_delay)
```

```
E1a_delay_Levene
```

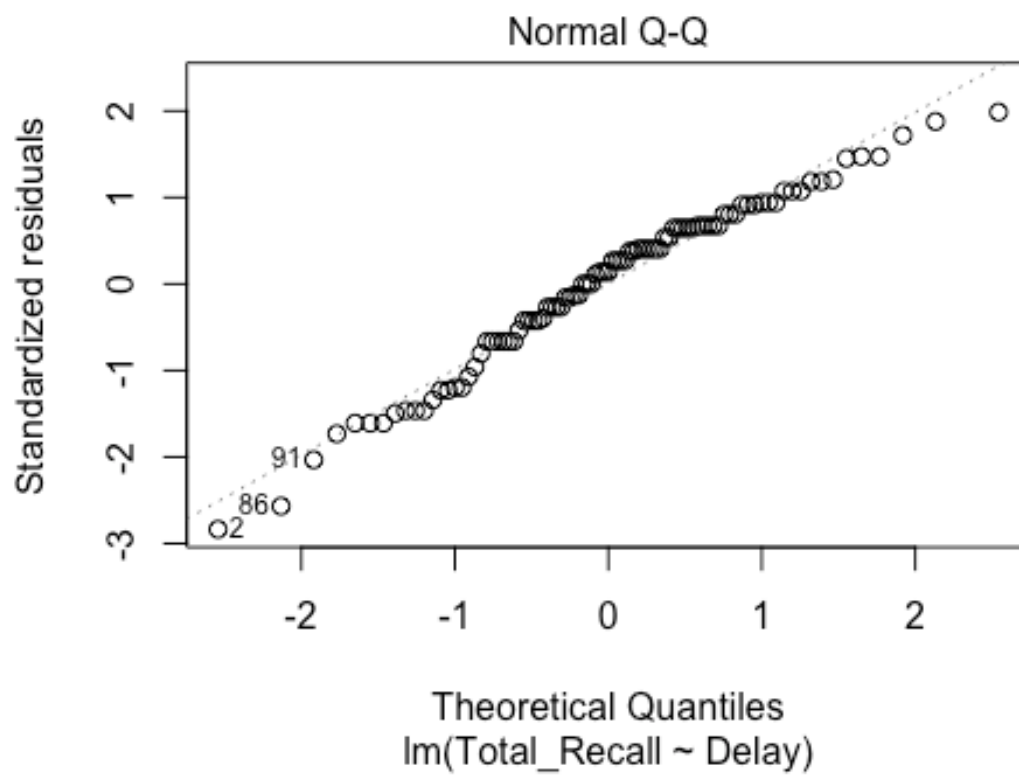
```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group 2  1.4465 0.2409
```

```
##      88
```

```
plot(E1a_model_delay, which= 2)
```



```
E1a_delay_SW<-shapiro.test(resid(E1a_model_delay))
```

```
E1a_delay_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```

## data: resid(E1a_model_delay)
## W = 0.97288, p-value = 0.05416
Analysis of Variance
anova(E1a_model_delay)
## Analysis of Variance Table
##
## Response: Total_Recall
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Delay      2  370.97  185.485   12.855 1.265e-05 ***
## Residuals 88 1269.78   14.429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
cohens_f(E1a_model_delay)
## For one-way between subjects designs, partial eta squared is
## equivalent to eta squared.
## Returning eta squared.
## # Effect Size for ANOVA
##
## Parameter | Cohen's f |          95% CI
## -----
## Delay     |      0.54 | [0.34,      Inf]
##
## - One-sided CIs: upper bound fixed at (Inf).

```

```
pairwise.t.test(E1a_data_delay$Total_Recall, E1a_data_delay$Delay,
p.adjust.method="bonferroni")
```

```
##
```

```
## Pairwise comparisons using t tests with pooled SD
```

```
##
```

```
## data: E1a_data_delay$Total_Recall and E1a_data_delay$Delay
```

```
##
```

```
##      Day      Week
```

```
## Week 0.00013 -
```

```
## Month 5.9e-05 1.00000
```

```
##
```

```
## P value adjustment method: bonferroni
```

Data illustration

```
E1a_plot_data <- E1a_data_delay %>%
```

```
  group_by(Delay) %>%
```

```
  summarise(
```

```
    sd= sd(Total_Recall),
```

```
    Total.Recall= mean(Total_Recall)
```

```
  )
```

```
E1a_Figure<- ggplot(E1a_plot_data, aes(x= Delay, y= Total.Recall,
```

```
ymin= Total.Recall-sd, ymax= Total.Recall+ sd))
```

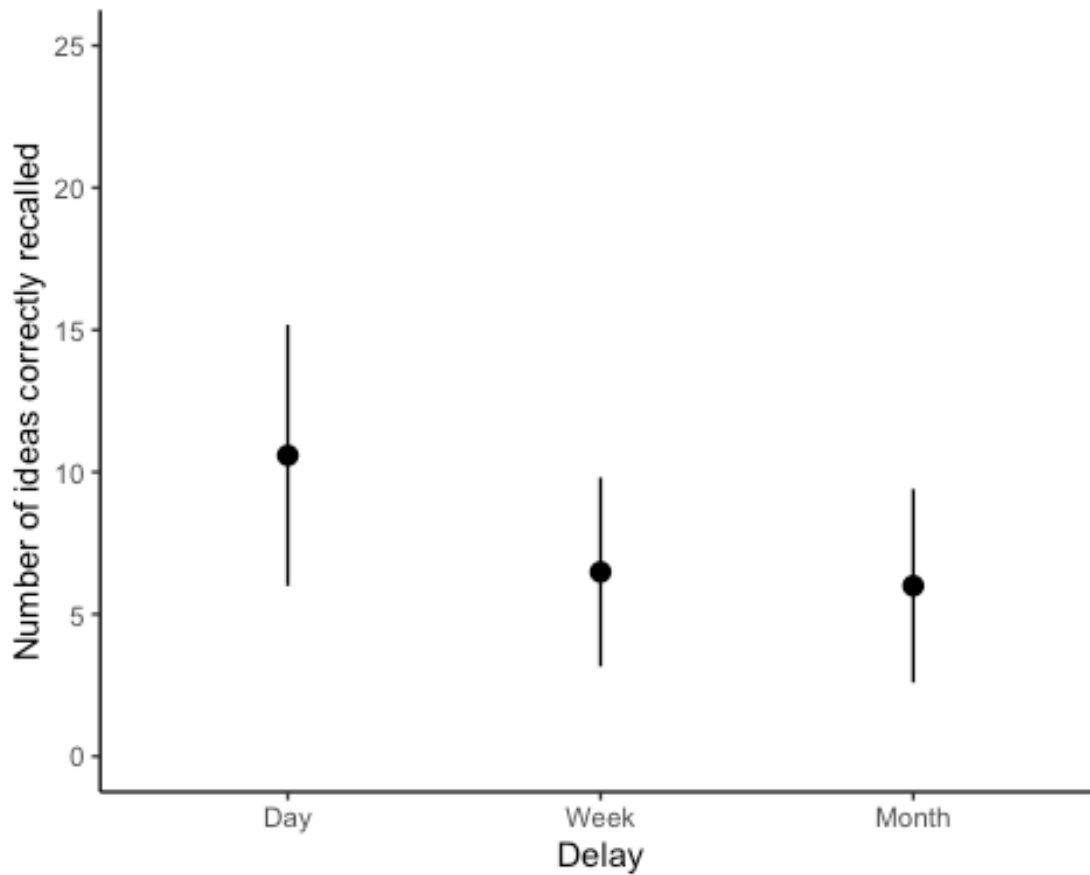
```
E1a_Figure <- E1a_Figure + geom_pointrange() +
```

```
  scale_y_continuous(name = 'Number of ideas correctly
```

```
recalled', limits = c(0,25)) +
```

```
theme_classic()
```

E1a\_Figure



Complementary analysis: Effect of experimenter on delay recall

### Descriptive statistics

```
E1a_expe <- E1a_data_delay %>%
```

```
  group_by(Experimenter, Delay) %>%
```

```
  summarise( N = n(), M = mean(Total_Recall), SD =
```

```
sd(Total_Recall),
```

```
  Min = min(Total_Recall),
```

```

      Median = median(Total_Recall),
      Max = max(Total_Recall))

## `summarise()` has grouped output by 'Experimenter'. You can
## override using the `.groups` argument.

E1a_expe<-APA.descr.table(E1a_expe)

E1a_expe

E1a_expe_GM <-E1a_data_delay %>%
  group_by(Experimenter) %>%
  summarise( N = n(), M = mean(Total_Recall), SD =
sd(Total_Recall),
      Min = min(Total_Recall),
      Median = median(Total_Recall),
      Max = max(Total_Recall))

E1a_expe_GM

## # A tibble: 2 x 7
##   Experimenter      N      M      SD      Min Median      Max
##   <fct>          <int> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 PhD              33  5.18  3.91     0     4    13
## 2 UG               58  9.05  3.83     0     9    18

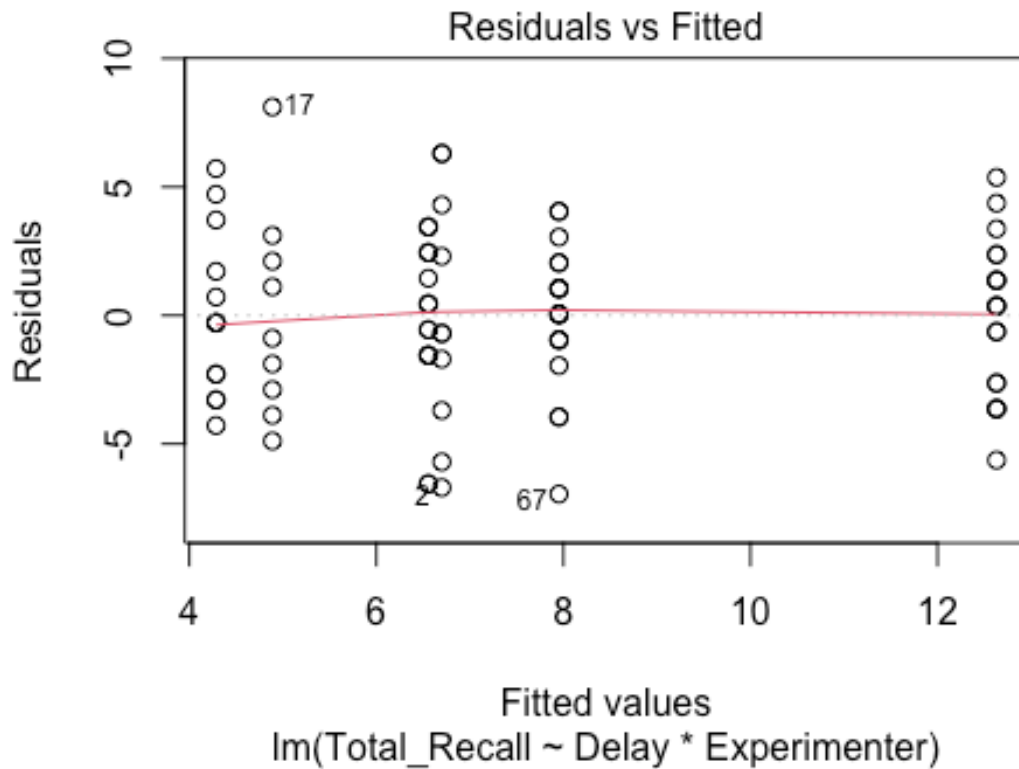
Inferential statistics

E1a_model_expe<- lm(Total_Recall ~ Delay * Experimenter, data =
E1a_data_delay)

Assumption checking

```

```
plot(E1a_model_expe, which= 1)
```



```
E1a_expe_Levene<-leveneTest(Total_Recall ~ Delay * Experimenter,  
data = E1a_data_delay)
```

```
E1a_expe_Levene
```

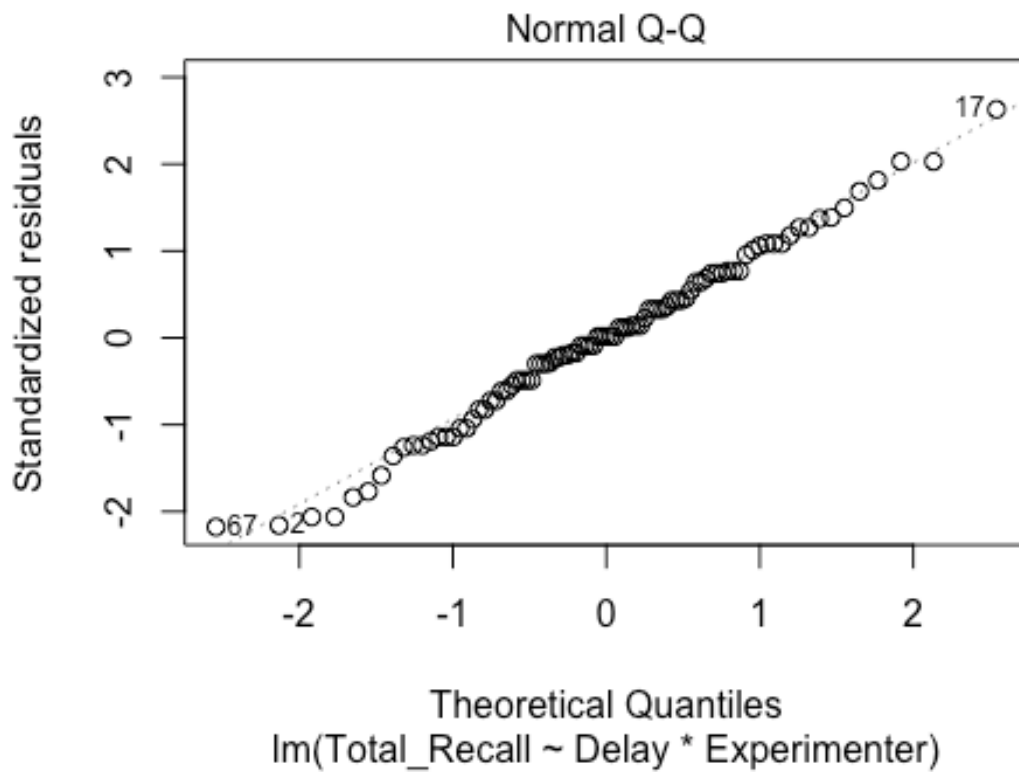
```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group 5  1.3236 0.2619
```

```
##      85
```

```
plot(E1a_model_expe, which= 2)
```



```
E1a_expe_SW <- shapiro.test(resid(E1a_model_expe))
```

```
E1a_expe_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: resid(E1a_model_expe)
```

```
## W = 0.99133, p-value = 0.8185
```

*Analysis of Variance*

```
anova(E1a_model_expe)
```

```

## Analysis of Variance Table
##
## Response: Total_Recall
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Delay           2  370.97  185.485  17.3319 4.863e-07 ***
## Experimenter    1  302.89  302.889  28.3023 8.288e-07 ***
## Delay:Experimenter 2   57.22   28.612   2.6736 0.07481 .
## Residuals      85  909.66   10.702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
cohens_f(E1a_model_expe, partial = FALSE)
## # Effect Size for ANOVA (Type I)
##
## Parameter          | Cohen's f |          95% CI
## -----
## Delay              |    0.54 | [0.33,      Inf]
## Experimenter       |    0.48 | [0.29,      Inf]
## Delay:Experimenter |    0.19 | [0.00,      Inf]
##
## - One-sided CIs: upper bound fixed at (Inf).
E1a_plot_data_2 <- E1a_data_delay %>%
  group_by(Delay, Experimenter) %>%
  summarise(

```

```

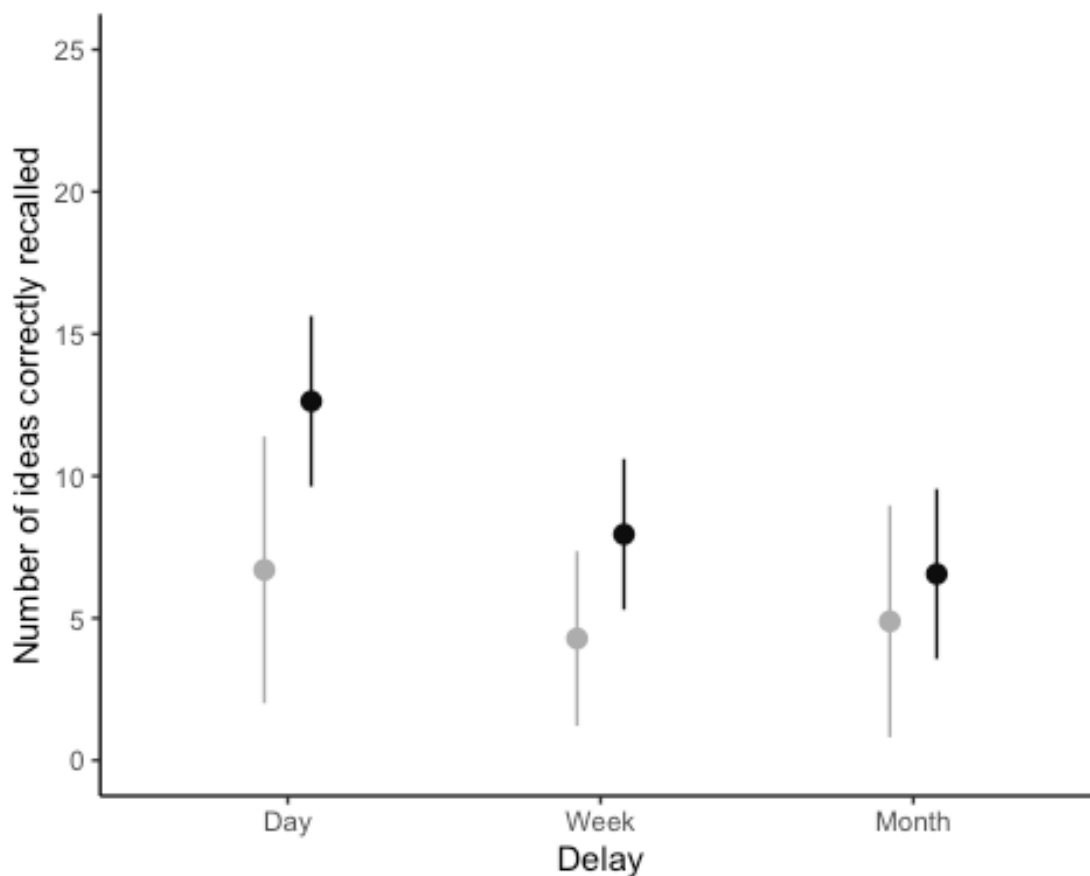
      sd= sd(Total_Recall),
      Total.Recall= mean(Total_Recall))
## `summarise()` has grouped output by 'Delay'. You can override
## using the `.groups` argument.
E1a_plot_data_2
## # A tibble: 6 x 4
## # Groups:   Delay [3]
##   Delay Experimenter   sd Total.Recall
##   <fct> <fct>         <dbl>     <dbl>
## 1 Day   PhD             4.69       6.7
## 2 Day   UG              3.00      12.6
## 3 Week  PhD             3.07       4.29
## 4 Week  UG              2.65       7.95
## 5 Month PhD             4.08       4.89
## 6 Month UG              2.99       6.56
E1a_Figure_2<- ggplot(E1a_plot_data_2, aes(x= Delay, y=
Total.Recall, ymin= Total.Recall-sd, ymax= Total.Recall+ sd, ,
color= Experimenter))
E1a_Figure_2<- E1a_Figure_2 + geom_pointrange(position =
position_dodge(0.3)) +
  scale_y_continuous(name = 'Number of ideas correctly recalled',
limits = c(0,25)) +
  scale_colour_manual(values = c('grey68', 'grey5')) +

```

```
theme_classic()+
```

```
theme(legend.position = "none")
```

E1a\_Figure\_2



Experiment 1b: The retention of a Piece of Prose Under Retrieval Practice with Incidental Free Recall

```
Experiment1b<- subset(Experiment1, Retrieval_Practice == 'With_RP',
                      select = c('Subject_ID', 'Delay',
                                'Administration', 'Experimenter', 'Total_Recall'))
```

```
E1b_data_delay<-subset(Experiment1b, Delay != 'Immediate', select =
c('Subject_ID', 'Delay', 'Administration', 'Total_Recall'))
```

```
E1b_data_delay$Delay<- factor(E1b_data_delay$Delay, levels
=c('Day', 'Week', 'Month'))
E1b_data_delay$Subject_ID<-as.numeric(E1b_data_delay$Subject_ID) #
to remove Levels from E1b
E1b_data_delay$Subject_ID<-as.factor(E1b_data_delay$Subject_ID)
E1b_data_delay$Administration<-
as.factor(E1b_data_delay$Administration)
```

### Descriptive statistics

```
E1b_descriptive_statistics <-E1b_data_delay %>%
  group_by(Delay) %>%
  summarise( N = n(), M = mean(Total_Recall), SD =
sd(Total_Recall),
  Min = min(Total_Recall),
  Median = median(Total_Recall),
  Max = max(Total_Recall))
E1b_descriptive_statistics <-
```

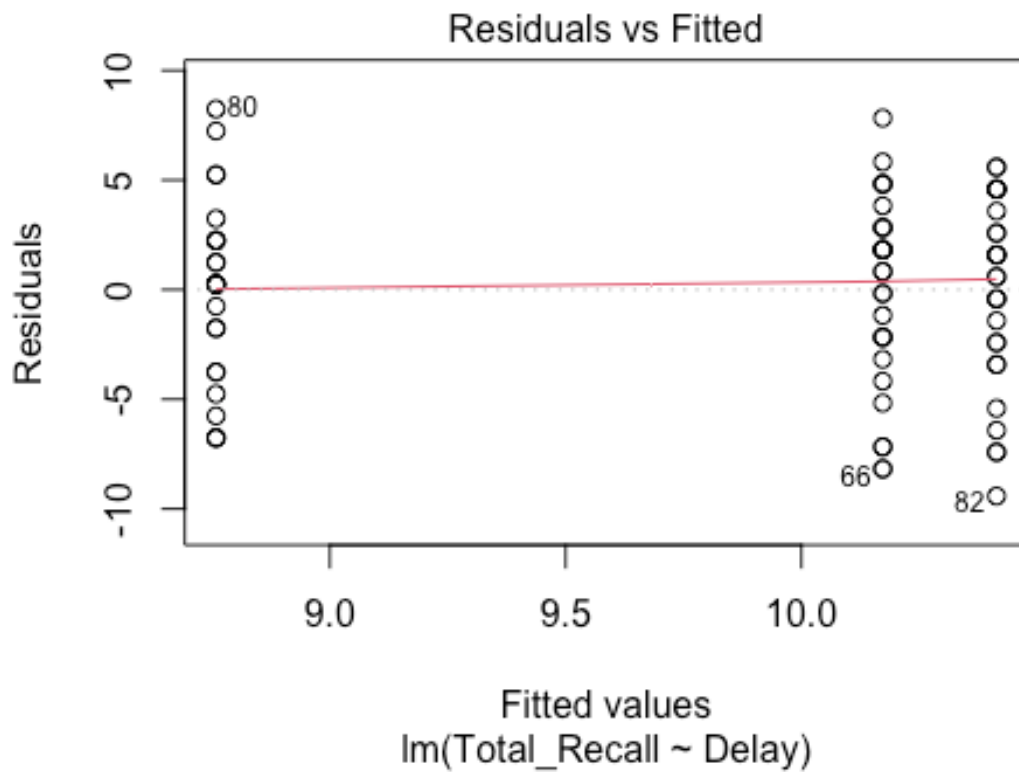
```
APA.descr.table(E1b_descriptive_statistics)
```

```
E1b_descriptive_statistics
```

### Inferential statistics

#### *Assumptions checking*

```
E1b_assum_model<- lm(Total_Recall ~ Delay, data= E1b_data_delay)
plot(E1b_assum_model, which = 1)
```



```
E1b_model<-anova_test(data = E1b_data_delay, dv = Total_Recall, wid
= Subject_ID, within = Delay)
```

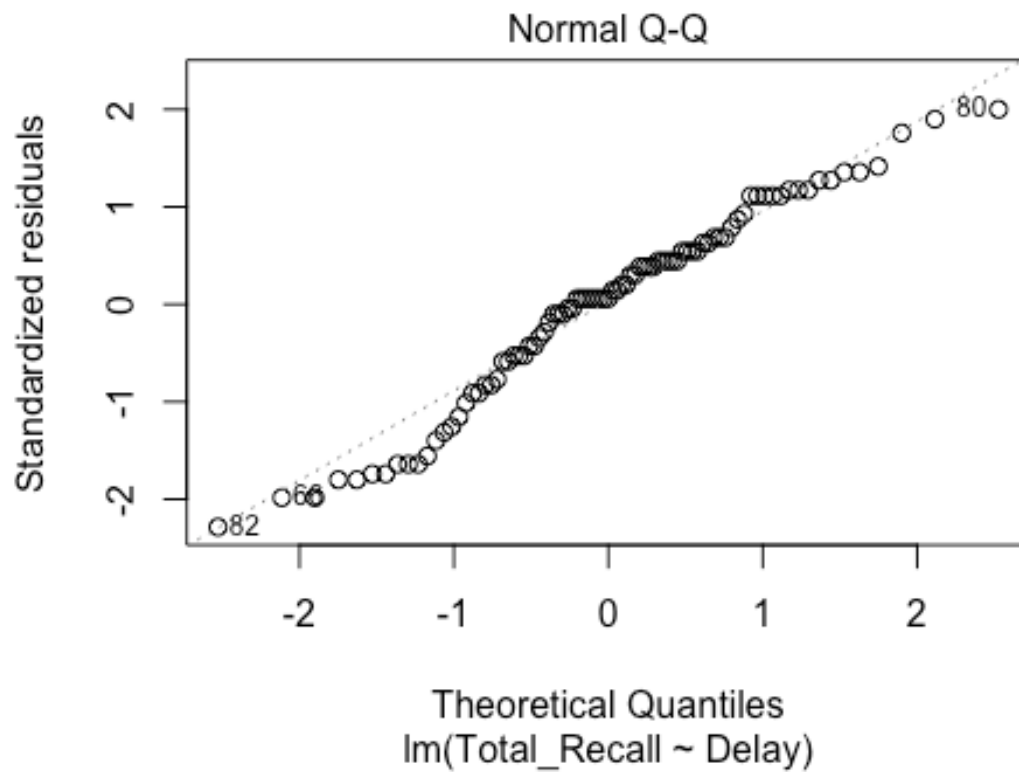
```
E1b_model$`Mauchly's Test for Sphericity`
```

```
## Effect W p p<.05
```

```
## 1 Delay 0.912 0.29
```

```
E1b_assum_model<- lm(Total_Recall ~ Delay, data= E1b_data_delay)
```

```
plot(E1b_assum_model, which= 2)
```



```
E1b_SW<-shapiro.test(resid(E1b_assum_model))
```

```
E1b_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: resid(E1b_assum_model)
```

```
## W = 0.96839, p-value = 0.03164
```

*Friedman's test*

```
friedman.test(y= E1b_data_delay$Total_Recall, groups =
```

```
E1b_data_delay$Delay, blocks = E1b_data_delay$Subject_ID)
```

```

##
## Friedman rank sum test
##
## data: E1b_data_delay$Total_Recall, E1b_data_delay$Delay and
E1b_data_delay$Subject_ID
## Friedman chi-squared = 13.134, df = 2, p-value = 0.001406
friedman_effsize(E1b_data_delay, Total_Recall ~ Delay | Subject_ID)
## # A tibble: 1 x 5
##   .y.          n effsize method  magnitude
## * <chr>      <int> <dbl> <chr>    <ord>
## 1 Total_Recall  29  0.226 Kendall W small
pairwise.wilcox.test(E1b_data_delay$Total_Recall,
E1b_data_delay$Delay, p.adj="bonferroni")
## Warning in wilcox.test.default(xi, xj, paired = paired, ...):
cannot compute
## exact p-value with ties

## Warning in wilcox.test.default(xi, xj, paired = paired, ...):
cannot compute
## exact p-value with ties

## Warning in wilcox.test.default(xi, xj, paired = paired, ...):

```

```

cannot compute

## exact p-value with ties

##

## Pairwise comparisons using Wilcoxon rank sum test with
continuity correction

##

## data: E1b_data_delay$Total_Recall and E1b_data_delay$Delay

##

##      Day Week
## Week 1.00 -
## Month 0.25 0.41

##

## P value adjustment method: bonferroni

E1b_plot_data <- E1b_data_delay %>%
  group_by(Delay) %>%
  summarise(
    sd= sd(Total_Recall),
    Total.Recall= mean(Total_Recall)
  )

E1b_Figure<- ggplot(E1b_plot_data, aes(x= Delay, y= Total.Recall,
group= 1, ymin= Total.Recall-sd, ymax= Total.Recall+ sd)) +
  geom_line() +
  geom_pointrange() +

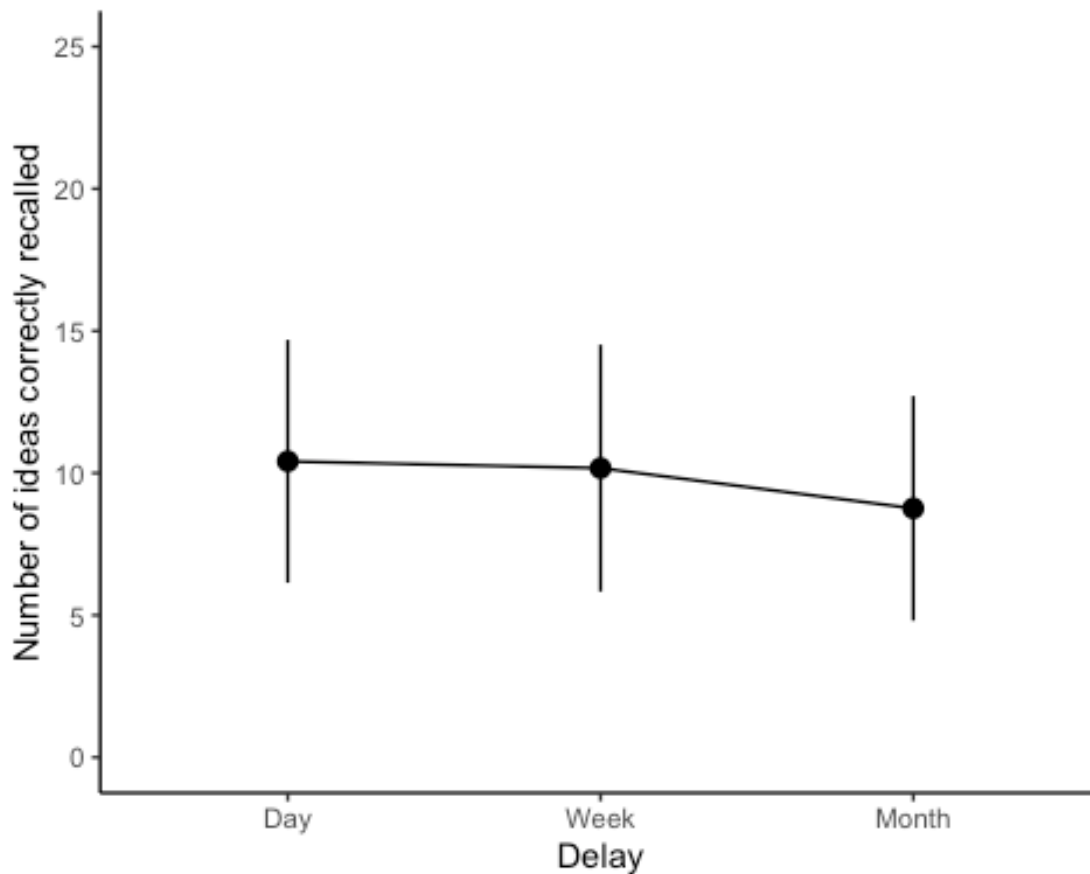
```

```

scale_y_continuous(name = 'Number of ideas correctly
recalled', limits = c(0,25))+
theme_classic()

```

E1b\_Figure



Analysis of variance also provided for information.

```
E1b_model$ANOVA
```

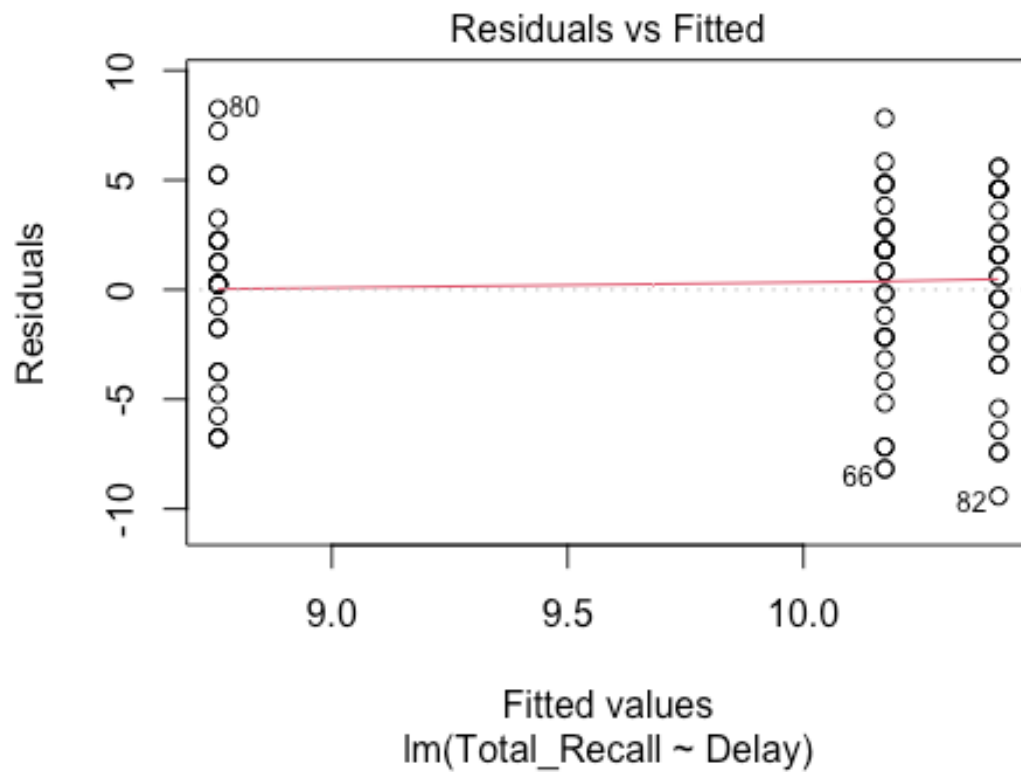
```

##   Effect DFn DFd    F      p p<.05 ges
## 1  Delay    2   56 11.7 5.68e-05 * 0.03

```

Complementary analysis: effect of administration conditions on the retention of the story

```
E1b_descriptive_statistics <- E1b_data_delay %>%  
  group_by(Administration, Delay) %>%  
  summarise( N = n(), M = mean(Total_Recall), SD =  
sd(Total_Recall),  
            Min = min(Total_Recall),  
            Median = median(Total_Recall),  
            Max = max(Total_Recall))  
## `summarise()` has grouped output by 'Administration'. You can  
## override using the `.groups` argument.  
E1b_descriptive_statistics <-  
APA.descr.table(E1b_descriptive_statistics)  
E1b_descriptive_statistics  
E1b_assum_model_2 <- lm(Total_Recall ~ Delay*Administration, data=  
E1b_data_delay)  
plot(E1b_assum_model, which = 1)
```



```
E1b_model_2<-anova_test(data = E1b_data_delay, dv = Total_Recall,
wid = Subject_ID, within = Delay, between= Administration)
```

```
E1b_model_2$`Mauchly's Test for Sphericity`
```

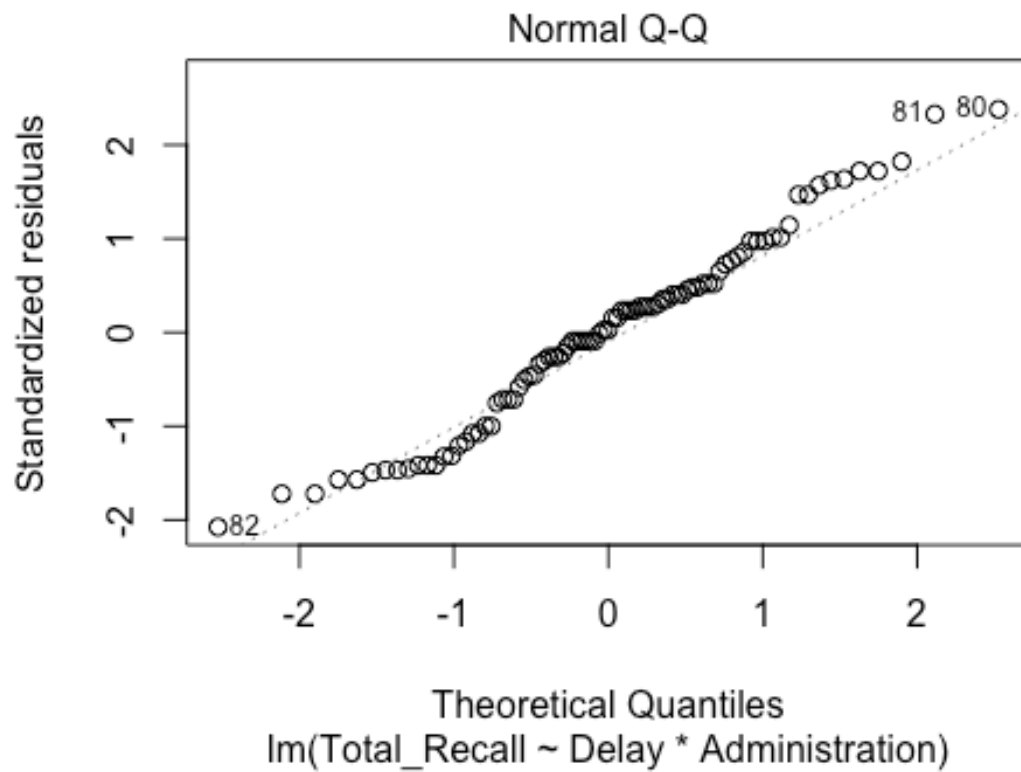
```
##           Effect      W      p p<.05
```

```
## 1                Delay 0.91 0.294
```

```
## 2 Administration:Delay 0.91 0.294
```

```
E1b_assum_model_2<- lm(Total_Recall ~ Delay*Administration, data=
E1b_data_delay)
```

```
plot(E1b_assum_model_2, which= 2)
```



```
E1b_SW_2<-shapiro.test(resid(E1b_assum_model_2))
```

```
E1b_SW_2
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: resid(E1b_assum_model_2)
```

```
## W = 0.97757, p-value = 0.1353
```

```
E1b_model_2$ANOVA
```

```
##           Effect DFn  DFd    F      p p<.05      ges
```

```
## 1 Administration    1   27  1.489 0.233000  0.048000
```

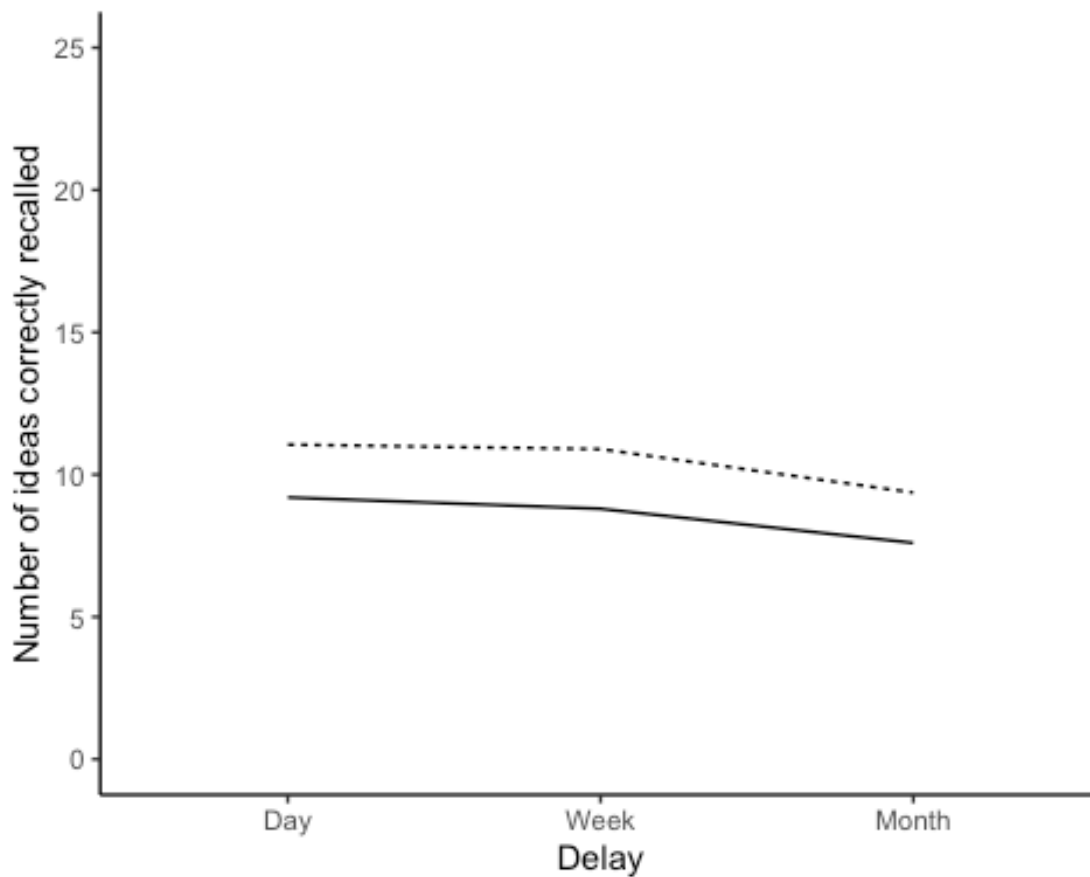
```
## 2          Delay    2  54  9.880 0.000221      * 0.028000
## 3 Administration:Delay    2  54  0.092 0.912000      0.000268

E1b_plot_data_2 <- E1b_data_delay %>%
  group_by(Delay, Administration) %>%
  summarise(
    sd= sd(Total_Recall),
    Total.Recall= mean(Total_Recall)
  )

## `summarise()` has grouped output by 'Delay'. You can override
using the `.groups` argument.

E1b_Figure_2<- ggplot(E1b_plot_data_2, aes(x= Delay, y=
Total.Recall, group= Administration, ymin= Total.Recall-sd, ymax=
Total.Recall+ sd, linetype= Administration)) +
  geom_line() +
  scale_y_continuous(name = 'Number of ideas correctly
recalled', limits = c(0,25)) +
  scale_colour_manual(values = c('grey40', 'grey5')) +
  theme_classic()+
  theme(legend.position = "none")

E1b_Figure_2
```

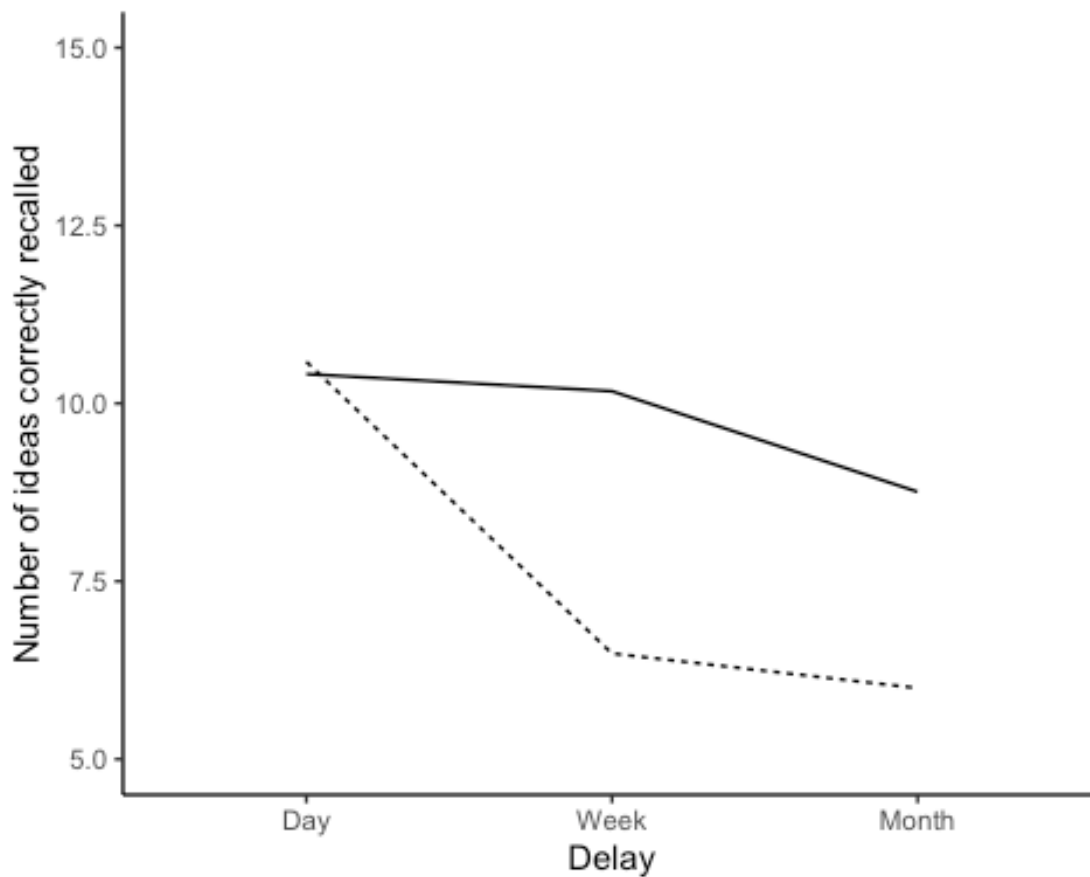


Experiment 1: Discussion illustration

```
E1_data_delay<- subset(Experiment1, Delay != 'Immediate',
                        select = c('Subject_ID', 'Delay',
                                  'Retrieval_Practice', 'Total_Recall'))
E1_data_delay$Delay<- factor(E1_data_delay$Delay, levels
                             =c('Day', 'Week', 'Month'))
E1_plot_data <- E1_data_delay %>%
  group_by(Retrieval_Practice, Delay) %>%
  summarise(Total_Recall= mean(Total_Recall)
            )
```

## `summarise()` has grouped output by 'Retrieval\_Practice'. You can override using the `.groups` argument.

```
E1_Figure_2<- ggplot(E1_plot_data,  
                    aes(x= Delay, y= Total_Recall, group=  
Retrieval_Practice,  
                    linetype= Retrieval_Practice)) +  
  geom_line() +  
  theme_classic() +  
  scale_y_continuous(name = 'Number of ideas correctly  
recalled', limits = c(5,15)) +  
  theme(legend.position = "none")  
E1_Figure_2
```




---

## Experiment 2: The Long-Term Retention of a Dilemma Story Under Retrieval

### Practice

*N.B: As a matter of data protection, participants' demographics are not provided in the dataset*

```
data(Experiment2, package = 'PhDCarolane')
```

```
Experiment2$Subject_ID<-as.factor(Experiment2$Subject_ID)
```

```
Experiment2$Story_type<-as.factor(Experiment2$Story_type)
```

```
Experiment2$Delay<- factor(Experiment2$Delay, levels
```

```
=c('Immediate', 'Day', 'Week', 'Month'))
```

```
# this set 'Immediate' as the first Delay, 'Day' as the second ...
```

```
otherwise automatically set in alphabetical order
```

```
Experiment2$Total_Recall<-as.numeric(Experiment2$Total_Recall)
```

```
Descriptive statistics
```

```
E2_table_story_type <- Experiment2 %>%
```

```
  group_by(Story_type) %>%
```

```
  summarise( N = n(), M = mean(Total_Recall), SD =
```

```
sd(Total_Recall),
```

```
    Min = min(Total_Recall),
```

```
    Median = median(Total_Recall),
```

```
    Max = max(Total_Recall)) # create a new object with
```

```
the descriptive statistics
```

```
E2_table_story_type<-APA.descr.table(E2_table_story_type) # From
```

```
PhDCarolane package, to format the table in APA style
```

```
E2_table_story_type<-E2_table_story_type %>%
```

```
  cols_label(Story_type = 'Story type') # relabel the column head
```

```
#gtsave(E2_table_Story_type, 'E2_Descriptives_per_story.pdf') #to
```

```
save as PDF
```

```
E2_table_story_type
```

```
E2_table_delay <- Experiment2 %>%
```

```
  group_by(Delay) %>%
```

```

summarise( N = n(), M = mean(Total_Recall), SD
= sd(Total_Recall),
          Min = min(Total_Recall),
          Median = median(Total_Recall),
          Max = max(Total_Recall)) # create a new
object with the descriptive statistics

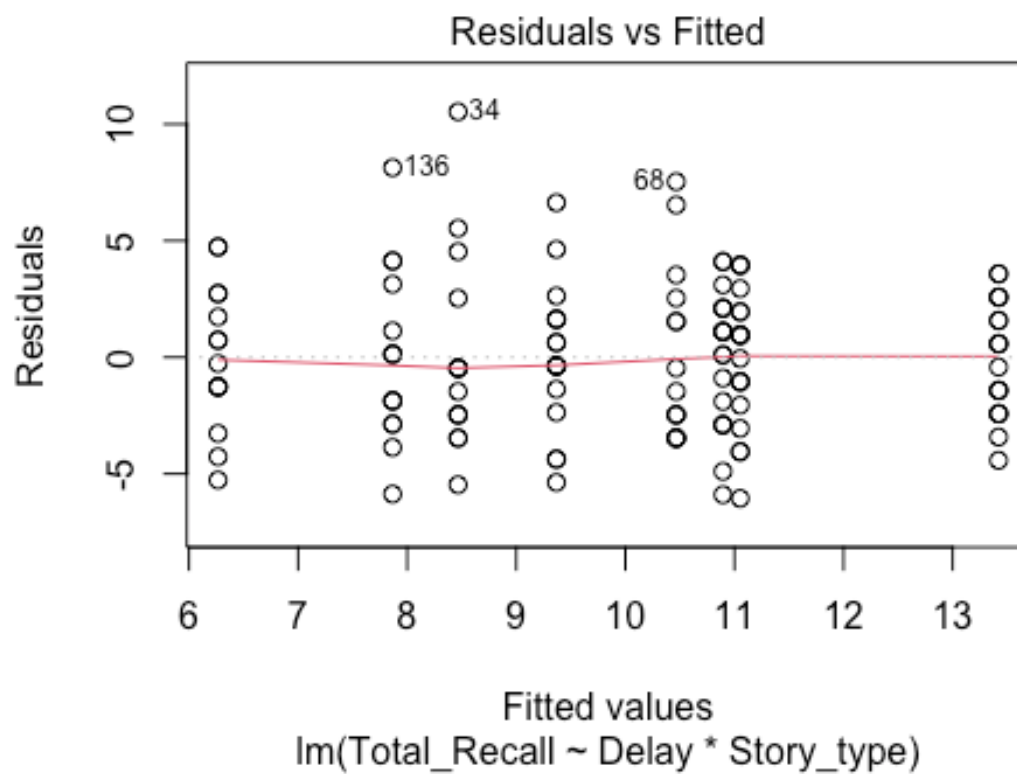
# Turn the dataframe into a APA style table
E2_table_delay<-APA.descr.table(E2_table_delay)
E2_table_delay
E2_Table_descriptives<-Experiment2 %>%
  group_by(Story_type, Delay) %>%
  summarise( N = n(), M = mean(Total_Recall), SD =
sd(Total_Recall),
          Min = min(Total_Recall),
          Median = median(Total_Recall),
          Max = max(Total_Recall))
## `summarise()` has grouped output by 'Story_type'. You can
override using the `.groups` argument.

# Turn the dataframe into a APA style table
E2_Table_descriptives<-APA.descr.table(E2_Table_descriptives)%>%
  cols_label(Delay='Condition')
E2_Table_descriptives
Inferential statistics

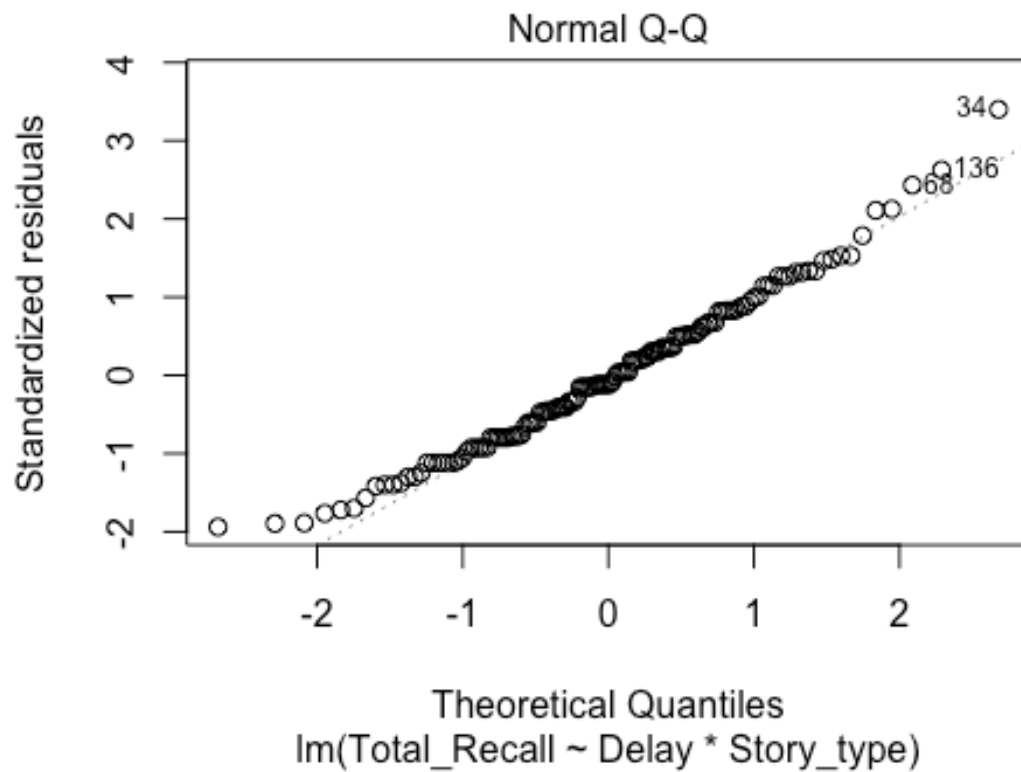
```

### Assumptions checking

```
E2_assum_model<- lm(Total_Recall ~ Delay*Story_type, data=
Experiment2)
plot(E2_assum_model, which = 1)
```



```
plot(E2_assum_model, which= 2)
```



```
E2_SW<-shapiro.test(resid(E2_assum_model))
```

```
E2_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: resid(E2_assum_model)
```

```
## W = 0.98376, p-value = 0.1064
```

```
E2_model<-anova_test(data = Experiment2, dv = Total_Recall, wid =  
Subject_ID, within = Delay, between = Story_type)
```

```
E2_model$`Mauchly's Test for Sphericity`
```

```
##           Effect      W      p p<.05
```

```
## 1           Delay 0.813 0.272
```

```
## 2 Story_type:Delay 0.813 0.272
```

### Mixed ANOVA

```
E2_anova<-ezANOVA(data= Experiment2, dv= Total_Recall, wid=
```

```
Subject_ID, within= Delay, between= Story_type)
```

```
## Warning: Data is unbalanced (unequal N per group). Make sure you
specified a
```

```
## well-considered value for the type argument to ezANOVA().
```

```
E2_anova
```

```
## $ANOVA
```

```
##           Effect DFn DFd           F           p p<.05
```

```
ges
```

```
## 2           Story_type      1      32  8.2230279 7.260958e-03      *
```

```
0.178229927
```

```
## 3           Delay          3      96 45.8189157 1.818020e-18      *
```

```
0.182572651
```

```
## 4 Story_type:Delay      3      96 0.2057504 8.921927e-01
```

```
0.001001952
```

```
##
```

```
## $`Mauchly's Test for Sphericity`
```

```
##           Effect      W      p p<.05
```

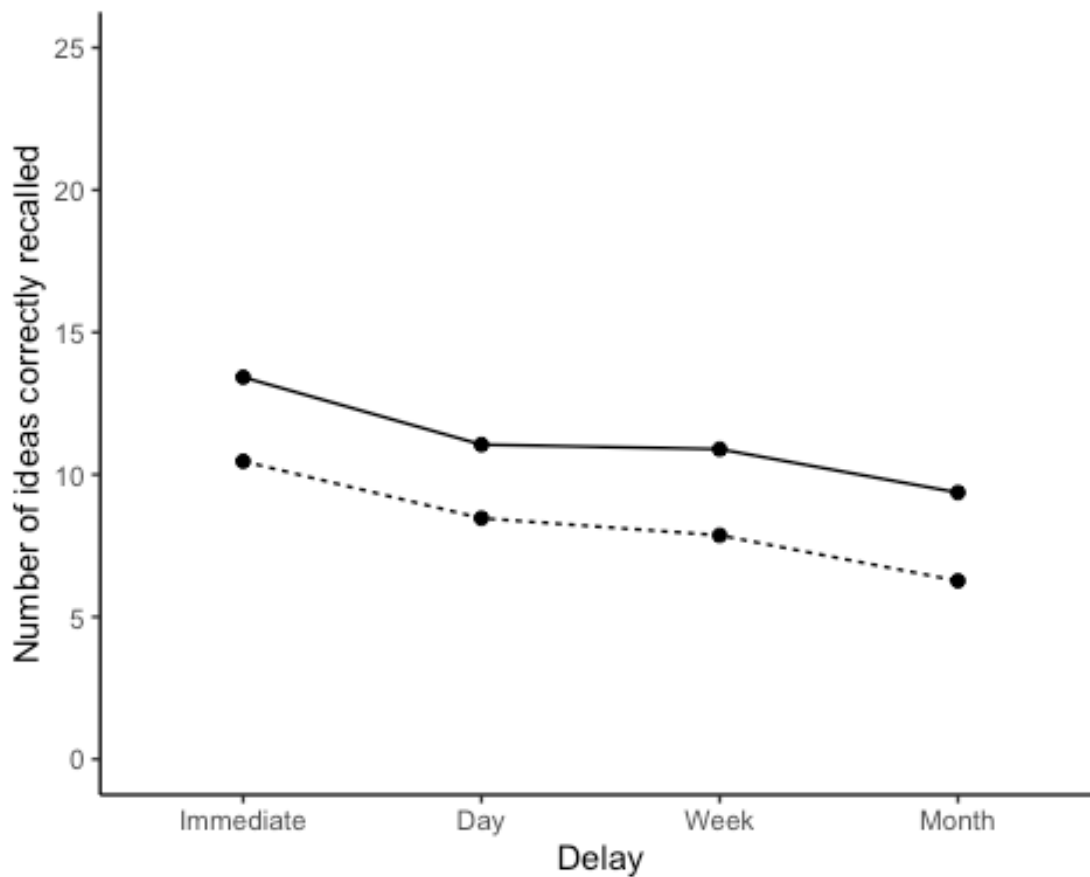
```
## 3           Delay 0.8128341 0.2723855
```

```

## 4 Story_type:Delay 0.8128341 0.2723855
##
## `$Sphericity Corrections`
##           Effect          GGe          p[GG] p[GG]<.05          HFe
p[HF]
## 3           Delay 0.875848 1.838814e-16          * 0.9613196
7.655262e-18
## 4 Story_type:Delay 0.875848 8.688414e-01          0.9613196
8.854723e-01
## p[HF]<.05
## 3          *
## 4
E2_multi.comp<-pairwise.t.test(Experiment2$Total_Recall,
Experiment2$Delay,
p.adj='bonf')
E2_multi.comp
##
## Pairwise comparisons using t tests with pooled SD
##
## data: Experiment2$Total_Recall and Experiment2$Delay
##
##           Immediate Day   Week
## Day 0.061 - -

```

```
## Week 0.018 1.000 -  
## Month 1.9e-05 0.152 0.404  
##  
## P value adjustment method: bonferroni  
E2_bonf.p.week<-E2_multi.comp$p.value[2,1]  
E2_bonf.p.month<-E2_multi.comp$p.value[3,1]  
E2_Figure<- ggline(Experiment2, x = 'Delay', y = 'Total_Recall',  
group = 'Story_type',  
linetype = 'Story_type',  
add = c('mean'))  
E2_Figure + scale_y_continuous(name = 'Number of ideas correctly  
recalled', limits = c(0,25))+  
theme_classic()+  
theme(legend.position= "none")
```




---

EXPERIMENT 3: A Comparison of Moral Decisions Types in a Moral Dilemmas Set.

```
data(Experiment3, package = 'PhDCarolane')
```

```
Experiment3[Experiment3=='']<-NA
```

Reporting moral decisions ratio

```
# The Bike Week
```

```
table_Bike_week_detailed_written<-
```

```
tabyl(Experiment3$Bike_week_detailed_written, show_na=FALSE,
```

```
show_missing_levels = FALSE)
```

```
table_Bike_week_detailed_audio<-
```

```
tabyl(Experiment3$Bike_week_detailed_audio, show_na=FALSE,
show_missing_levels = FALSE)

table_Bike_week_original_written<-
tabyl(Experiment3$Bike_week_original_written, show_na=FALSE,
show_missing_levels = FALSE)

table_Bike_week_original_audio<-
tabyl(Experiment3$Bike_week_original_audio, show_na=FALSE,
show_missing_levels = FALSE)

# The Nobel Price

table_Nobel_price_detailed_written<-
tabyl(Experiment3$Nobel_price_detailed_written, show_na=FALSE,
show_missing_levels = FALSE)

table_Nobel_price_detailed_audio<-
tabyl(Experiment3$Nobel_price_detailed_audio, show_na=FALSE,
show_missing_levels = FALSE)

table_Nobel_price_original_written<-
tabyl(Experiment3$Nobel_price_original_written, show_na=FALSE,
show_missing_levels = FALSE)

table_Nobel_price_original_audio<-
tabyl(Experiment3$Nobel_price_original_audio, show_na=FALSE,
show_missing_levels = FALSE)

# The Vaccine

table_Vaccine_detailed_written<-
```

```
tabyl(Experiment3$Vaccine_detailed_written, show_na=FALSE,
show_missing_levels = FALSE)
table_Vaccine_detailed_audio<-
tabyl(Experiment3$Vaccine_detailed_audio, show_na=FALSE,
show_missing_levels = FALSE)
table_Vaccine_original_written<-
tabyl(Experiment3$Vaccine_original_written, show_na=FALSE,
show_missing_levels = FALSE)
table_Vaccine_original_audio<-
tabyl(Experiment3$Vaccine_original_audio, show_na=FALSE,
show_missing_levels = FALSE)
# The Best Friend
table_Best_friends_detailed_written<-
tabyl(Experiment3$Best_friends_detailed_written, show_na=FALSE,
show_missing_levels = FALSE)
table_Best_friends_detailed_audio<-
tabyl(Experiment3$Best_friends_detailed_audio, show_na=FALSE,
show_missing_levels = FALSE)
table_Best_friends_original_written<-
tabyl(Experiment3$Best_friends_original_written, show_na=FALSE,
show_missing_levels = FALSE)
table_Best_friends_original_audio<-
```

```
tabyl(Experiment3$Best_friends_original_audio, show_na=FALSE,  
show_missing_levels = FALSE)  
  
Bike_week_detailed_written<-table_Bike_week_detailed_written$percent  
Bike_week_detailed_audio<-table_Bike_week_detailed_audio$percent  
Bike_week_original_written<-table_Bike_week_original_written$percent  
Bike_week_original_audio<-table_Bike_week_original_audio$percent  
  
Nobel_price_detailed_written<-  
table_Nobel_price_detailed_written$percent  
Nobel_price_detailed_audio<-table_Nobel_price_detailed_audio$percent  
Nobel_price_original_written<-  
table_Nobel_price_original_written$percent  
Nobel_price_original_audio<-table_Nobel_price_original_audio$percent  
  
Vaccine_detailed_written<-table_Vaccine_detailed_written$percent  
Vaccine_detailed_audio<-table_Vaccine_detailed_audio$percent  
Vaccine_original_written<-table_Vaccine_original_written$percent  
Vaccine_original_audio<-table_Vaccine_original_audio$percent  
  
Best_friends_detailed_written<-  
table_Best_friends_detailed_written$percent  
Best_friends_detailed_audio<-  
table_Best_friends_detailed_audio$percent
```

```

Best_friends_original_written<-
table_Best_friends_original_written$percent
Best_friends_original_audio<-
table_Best_friends_original_audio$percent

# Create the matrix
E3_decisions_ratio<-cbind(Bike_week_detailed_written,
                           Bike_week_detailed_audio,
                           Bike_week_original_written,
                           Bike_week_original_audio,
                           Nobel_price_detailed_written,
                           Nobel_price_detailed_audio,
                           Nobel_price_original_written,
                           Nobel_price_original_audio,
                           Vaccine_detailed_written,
                           Vaccine_detailed_audio,
                           Vaccine_original_written,
                           Vaccine_original_audio,
                           Best_friends_detailed_written,
                           Best_friends_detailed_audio,
                           Best_friends_original_written,
                           Best_friends_original_audio)

#Make it a dataframe

```

```
E3_decisions_ratio<-as.data.frame(t(E3_decisions_ratio)) #t
transpose the matrix so the answers become variables
#Naming the variables
setnames( E3_decisions_ratio, 'V1', 'Deontological' )
setnames( E3_decisions_ratio, 'V2', 'Utilitarian' )
write.csv(E3_decisions_ratio, file = 'E3_data_endorsement.csv')
```

```
E3_decisions_ratio
```

##	Deontological	Utilitarian
## Bike_week_detailed_written	0.5806452	0.41935484
## Bike_week_detailed_audio	0.7777778	0.22222222
## Bike_week_original_written	0.6415094	0.35849057
## Bike_week_original_audio	0.7222222	0.27777778
## Nobel_price_detailed_written	0.6923077	0.30769231
## Nobel_price_detailed_audio	0.5333333	0.46666667
## Nobel_price_original_written	0.7111111	0.28888889
## Nobel_price_original_audio	0.5000000	0.50000000
## Vaccine_detailed_written	0.9750000	0.02500000
## Vaccine_detailed_audio	0.5625000	0.43750000
## Vaccine_original_written	0.8409091	0.15909091
## Vaccine_original_audio	0.5454545	0.45454545
## Best_friends_detailed_written	0.1142857	0.88571429
## Best_friends_detailed_audio	0.9285714	0.07142857

```
## Best_friends_original_written      0.1224490  0.87755102
## Best_friends_original_audio        0.5384615  0.46153846
```

---

Experiment 4: The Impact of Decision-Making on Memory Retention after A Day

Call the dataset stored in the package 'PhDcarolaneh' and place it in the

environment. *N.B: As a matter of data protection, participants' demographics are not provided in the dataset*

```
data(Experiment4, package = 'PhDCarolane')
```

```
Experiment4$Decision[Experiment4$Decision == 'No_decision']<- 'No
decision'
```

```
# The demographics file not included here shows that 4 participants
are over 50.
```

```
# They are identified under the Subject_ID 47, 66, 83, 98
```

```
Experiment4<-Experiment4[Experiment4$Subject_ID != 47 &
```

```
Experiment4$Subject_ID != 66 & Experiment4$Subject_ID != 83 &
```

```
Experiment4$Subject_ID != 98, ]
```

Inter-rater reliability

```
data(E4_interrater, package = 'PhDCarolane')
```

```
# using the cohen.kappa() function directly on E4_interrater lead to
a ''list' object cannot be coerced to type 'double'' error.
```

```
# setting the factor as numeirc does not fix the error so we need to
use the passby procedure below
```

```
a<-E4_interrater$CH
```

```
b<-E4_interrater$GD
k<-as.data.frame(cbind(a,b))
agree(k, tolerance=0) # for agreement %
## Percentage agreement (Tolerance=0)
##
## Subjects = 270
## Raters = 2
## %-agree = 91.1
kappa2(k, 'unweighted') # for k and p
## Cohen's Kappa for 2 Raters (Weights: unweighted)
##
## Subjects = 270
## Raters = 2
## Kappa = 0.821
##
## z = 13.5
## p-value = 0
cohen.kappa(k) # for 95%CI
## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha,
levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and
confidence boundaries
```

```
##           lower estimate upper
## unweighted kappa  0.75     0.82  0.89
## weighted kappa   0.75     0.82  0.89
##
## Number of subjects = 270
Effect of decision on post decision recall
E4_data_ctrl<-subset(Experiment4, Delay == 'Immediate', select =
c('Subject_ID', 'Decision', 'DSF', 'DSB', 'Total_Recall'))
E4_data_ctrl$DSF<-as.numeric(E4_data_ctrl$DSF)
## Warning: NAs introduced by coercion
E4_data_ctrl$DSB<-as.numeric(E4_data_ctrl$DSB)
## Warning: NAs introduced by coercion
E4_data_ctrl$Total_Recall<-as.numeric(E4_data_ctrl$Total_Recall)
## Warning: NAs introduced by coercion
E4_data_ctrl$Decision<-as.factor(E4_data_ctrl$Decision)
E4_data_ctrl$Decision<- factor(E4_data_ctrl$Decision, levels= c('No
decision', 'Deontological', 'Utilitarian'))
Descriptive statistics
E4_post_recall<-E4_data_ctrl %>%
  group_by(Decision) %>%
  summarise( N = n(), M = mean(Total_Recall, na.rm=TRUE), SD =
sd(Total_Recall, na.rm=TRUE),
           Min = min(Total_Recall, na.rm=TRUE),
```

```
Median = median(Total_Recall, na.rm=TRUE),
```

```
Max = max(Total_Recall, na.rm=TRUE))
```

```
E4_post_recall<-APA.descr.table(E4_post_recall)
```

```
E4_post_recall
```

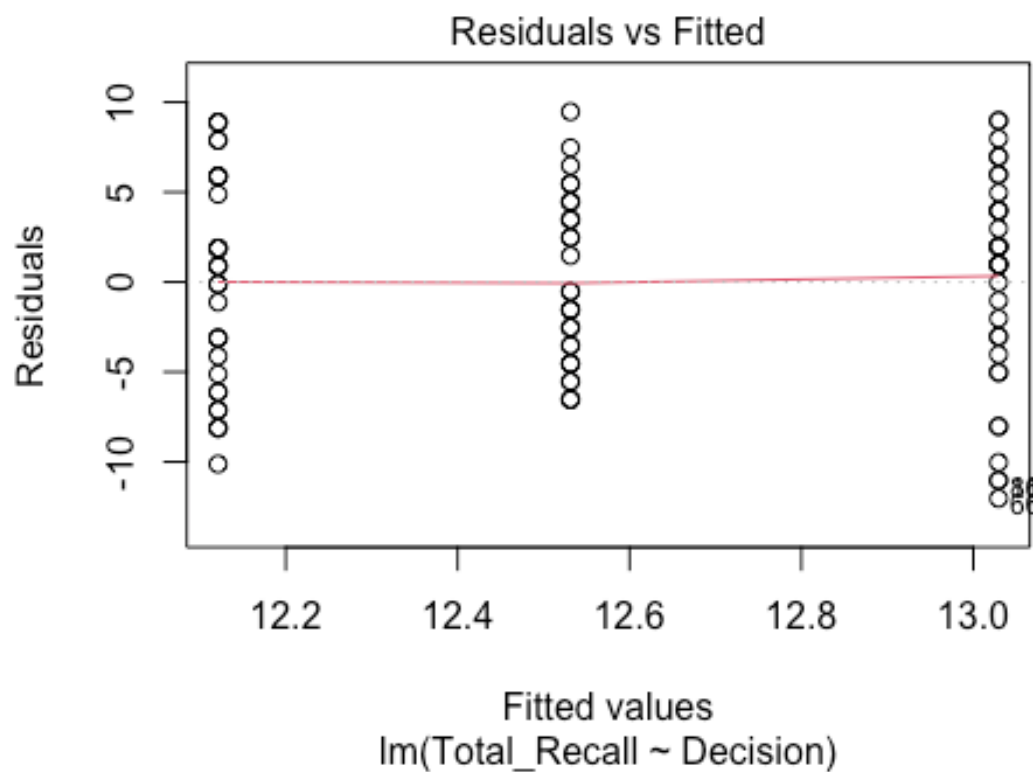
### Inferential statistics

#### *Assumptions checking*

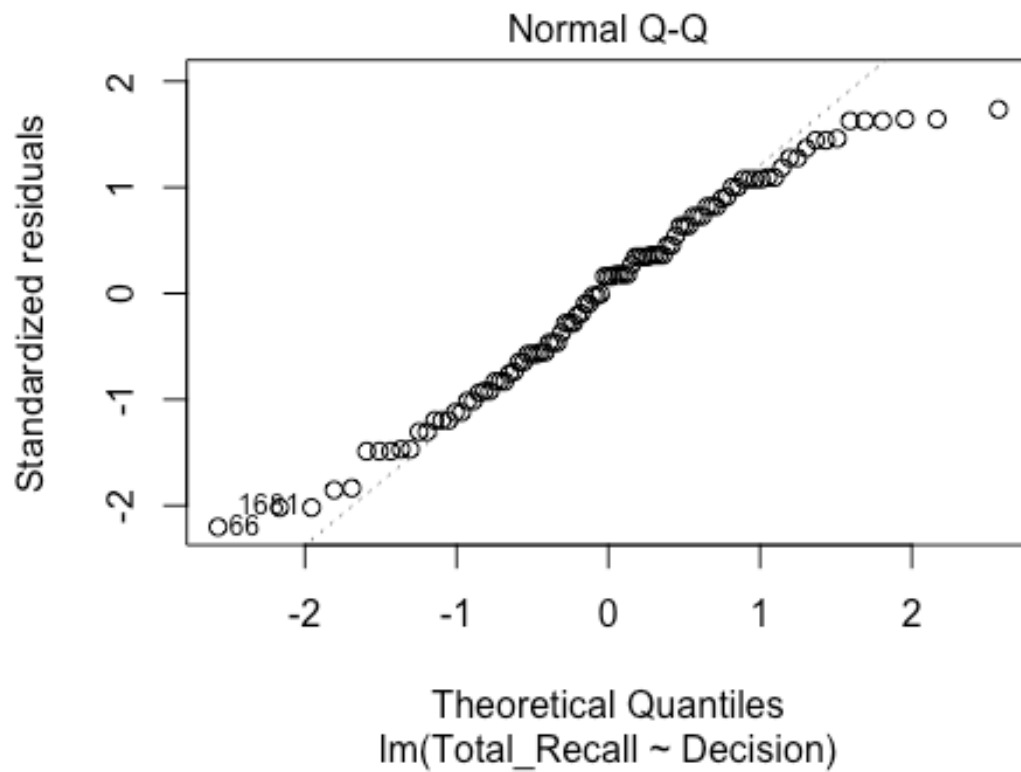
```
E4_model_post_recall<- lm(Total_Recall ~ Decision, data =
```

```
E4_data_ctrl)
```

```
plot(E4_model_post_recall, which=1)
```



```
E4_post_recall_Levene<-leveneTest(Total_Recall ~ Decision, data=
E4_data_ctrl)
E4_post_recall_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  2  0.7526 0.4739
##      96
E4_post_recall_residuals<-residuals(lm(Total_Recall ~ Decision, data
= E4_data_ctrl))
shapiro.test(E4_post_recall_residuals)
##
## Shapiro-Wilk normality test
##
## data:  E4_post_recall_residuals
## W = 0.97315, p-value = 0.04037
plot(E4_model_post_recall, which= 2)
```



*Kruskal-Wallis rank-sum test*

```
E4_post_recall_KW<-kruskal.test(Total_Recall ~ Decision, data=
E4_data_ctrl)
```

```
E4_post_recall_KW
```

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: Total_Recall by Decision
```

```
## Kruskal-Wallis chi-squared = 0.66329, df = 2, p-value = 0.7177
```

Effect of decision groups on DSF

## Descriptive statistics

```
E4_DSF<-E4_data_ctrl %>%  
  group_by(Decision) %>%  
  summarise( N = n(), M = mean(DSF, na.rm=TRUE), SD = sd(DSF,  
na.rm=TRUE),  
            Min = min(DSF, na.rm=TRUE),  
            Median = median(DSF, na.rm=TRUE),  
            Max = max(DSF, na.rm=TRUE))
```

```
E4_DSF<-APA.descr.table(E4_DSF)
```

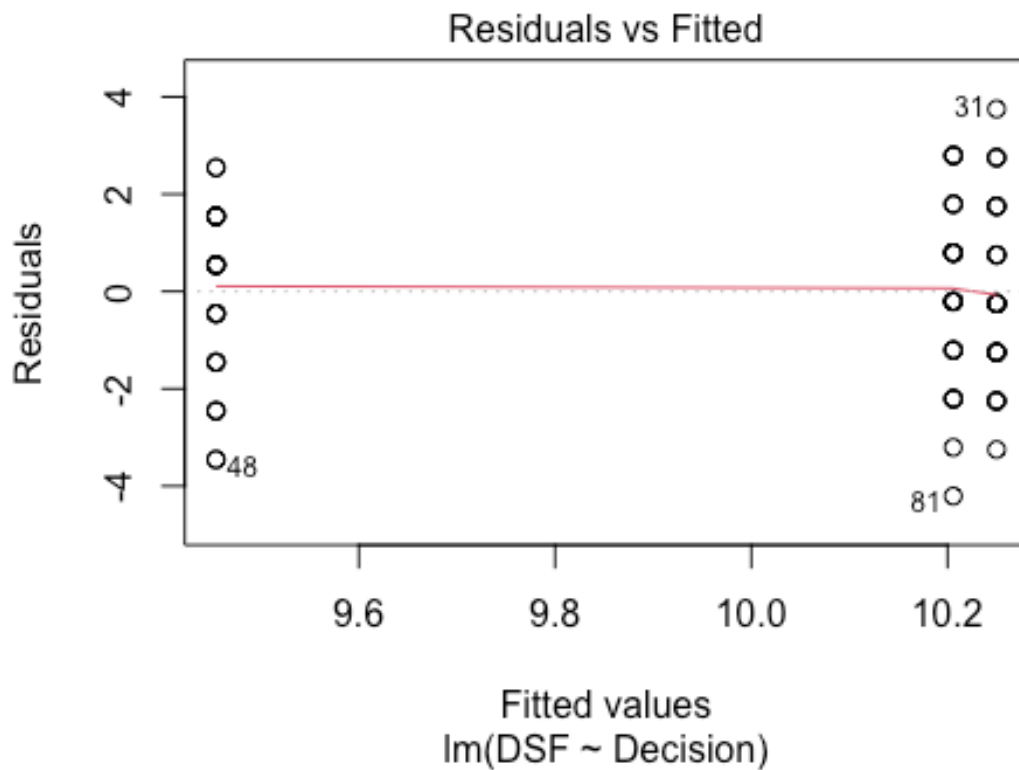
```
E4_DSF
```

## Inferential statistics

### *Assumptions checking*

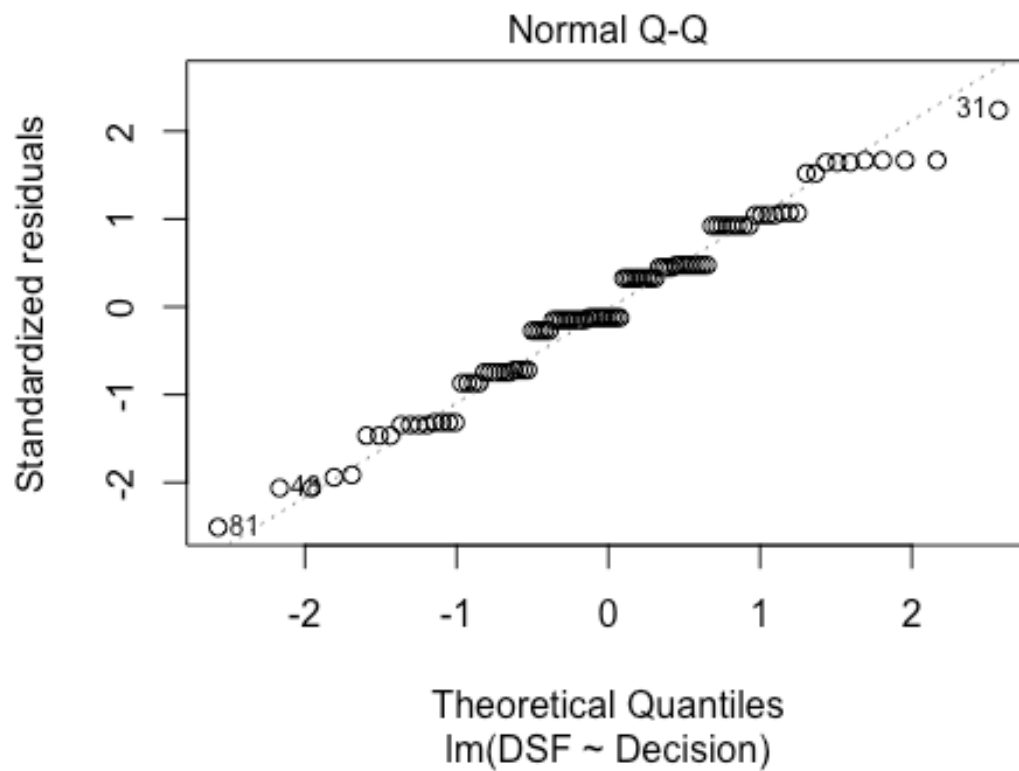
```
E4_model_DSF<- lm(DSF ~ Decision, data = E4_data_ctrl)
```

```
plot(E4_model_DSF, which=1)
```



```
E4_DSF_Levene<-leveneTest(DSF ~ Decision, data= E4_data_ctrl)
E4_DSF_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 2  0.0185 0.9816
##      96
E4_DSF_residuals<-residuals(lm(DSF ~ Decision, data = E4_data_ctrl))
shapiro.test(E4_DSF_residuals)
##
## Shapiro-Wilk normality test
```

```
##
## data: E4_DSF_residuals
## W = 0.97774, p-value = 0.0916
plot(E4_model_DSF, which= 2)
```



ANOVA

```
E4_DSF_anova<-aov(DSF ~ Decision, data= E4_data_ctrl)
E4_DSF_aov_table<-Anova(E4_DSF_anova, type = 'III')
E4_DSF_aov_table
## Anova Table (Type III tests)
##
```

```

## Response: DSF
##           Sum Sq Df    F value Pr(>F)
## (Intercept) 2949.82  1 1019.5935 <2e-16 ***
## Decision     13.17  2    2.2758 0.1082
## Residuals   277.74 96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Save an APA-style table.

apa.aov.table(E4_DSF_anova, filename = 'E4_DSF_ANOVA.doc',
table.number = 7)

##
##
## Table 7
##
## ANOVA results using DSF as the dependent variable
##
##
##   Predictor      SS df      MS      F      p partial_eta2
##   (Intercept) 2949.82  1 2949.82 1019.59 .000
##   Decision    13.17  2    6.58    2.28 .108          .05
##   Error       277.74 96    2.89
##   CI_90_partial_eta2
##   [.00, .12]

```

```
##
```

```
## Note: Values in square brackets indicate the bounds of the 90%
confidence interval for partial eta-squared
```

```
Effect of decision groups on DSB
```

### Descriptive statistics

```
E4_DSB<-E4_data_ctrl %>%
```

```
  group_by(Decision) %>%
```

```
  summarise( N = n(), M = mean(DSB, na.rm=TRUE), SD = sd(DSB,
na.rm=TRUE),
```

```
             Min = min(DSB, na.rm=TRUE),
```

```
             Median = median(DSB, na.rm=TRUE),
```

```
             Max = max(DSB, na.rm=TRUE))
```

```
E4_DSB<-APA.descr.table(E4_DSB)
```

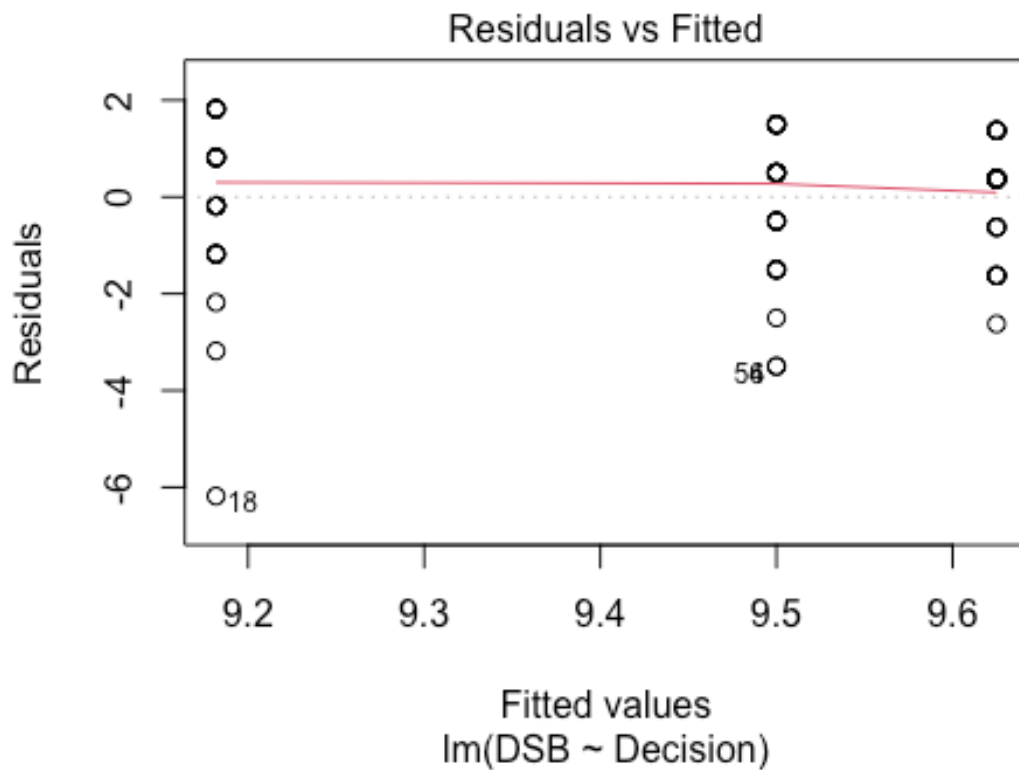
```
E4_DSB
```

### Inferential statistics

*Assumptions checking*

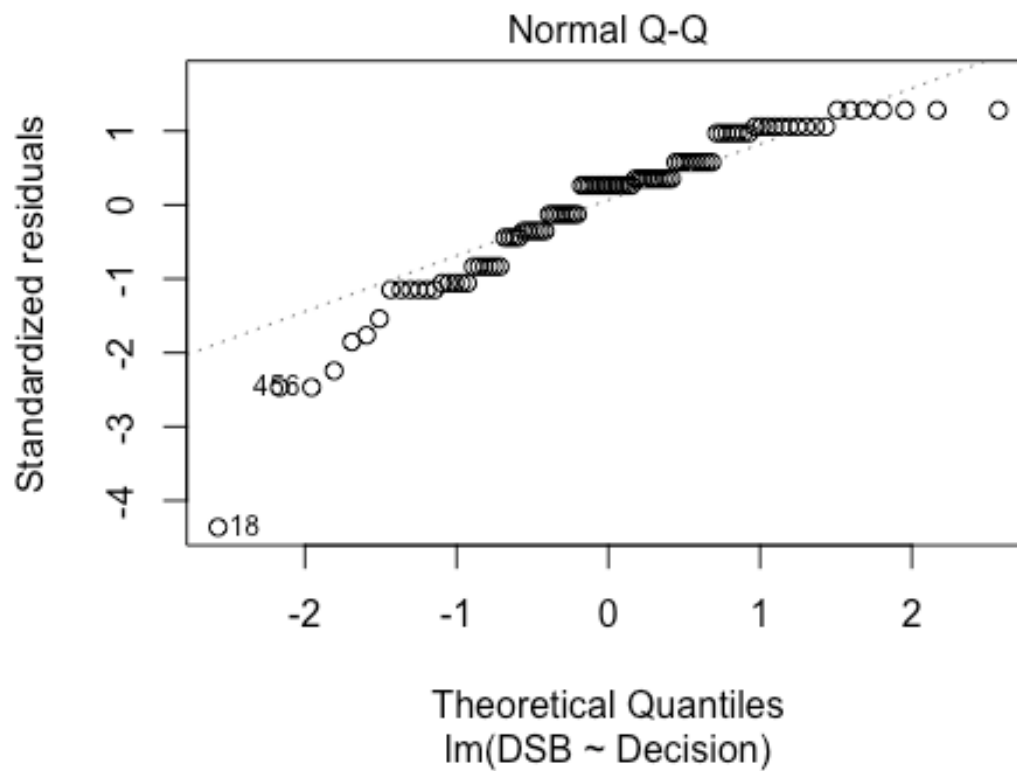
```
E4_model_DSB<- lm(DSB ~ Decision, data = E4_data_ctrl)
```

```
plot(E4_model_DSB, which=1)
```



```
E4_DSB_Levene<-leveneTest(DSB ~ Decision, data= E4_data_ctrl)
E4_DSB_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  2  1.2627 0.2875
##      96
E4_DSB_residuals<-residuals(lm(DSB ~ Decision, data = E4_data_ctrl))
shapiro.test(E4_DSB_residuals)
##
## Shapiro-Wilk normality test
```

```
##
## data: E4_DSB_residuals
## W = 0.90287, p-value = 2.141e-06
plot(E4_model_DSB, which= 2)
```



*Kruskal-Wallis rank-sum test*

```
E4_DSB_KW<-kruskal.test(DSB ~ Decision, data= E4_data_ctrl)
```

```
E4_DSB_KW
```

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: DSB by Decision
```

```
## Kruskal-Wallis chi-squared = 1.0398, df = 2, p-value = 0.5946
```

The effect of decision-making groups on story recall after a day.

```
E4_data_delay<-subset(Experiment4, Delay != 'Immediate', select =
c('Subject_ID', 'Delay', 'Decision', 'Total_Recall'))
```

```
E4_data_delay$Total_Recall<-as.numeric(E4_data_delay$Total_Recall)
```

```
## Warning: NAs introduced by coercion
```

```
E4_data_delay$Decision<-factor(E4_data_delay$Decision, levels= c('No
decision', 'Deontological', 'Utilitarian'))
```

```
E4_data_delay$Delay<-factor(E4_data_delay$Delay, levels= c('Day',
'Week', 'Month'))
```

### Descriptive statistics

```
E4_delay_recall<-E4_data_delay %>%
```

```
  group_by(Decision) %>%
```

```
  summarise( N = n(), M = mean(Total_Recall, na.rm=TRUE), SD =
```

```
sd(Total_Recall, na.rm=TRUE),
```

```
            Min = min(Total_Recall, na.rm=TRUE),
```

```
            Median = median(Total_Recall, na.rm=TRUE),
```

```
            Max = max(Total_Recall, na.rm=TRUE))
```

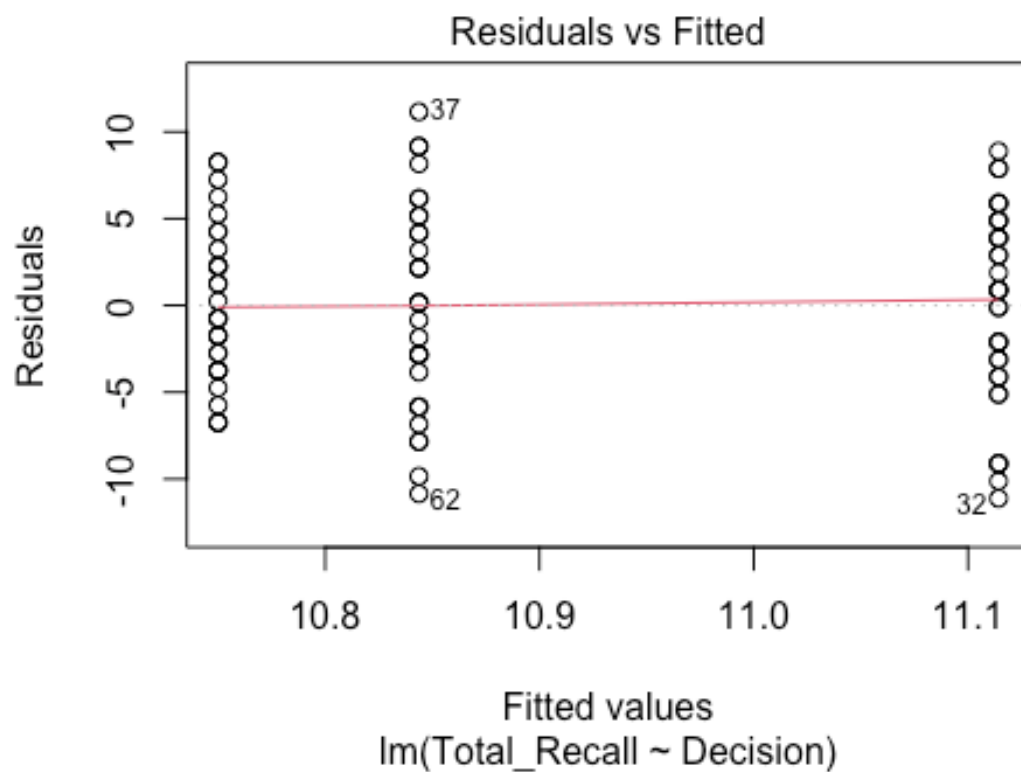
```
E4_delay_recall<-APA.descr.table(E4_delay_recall)
```

```
E4_delay_recall
```

### Inferential statistics

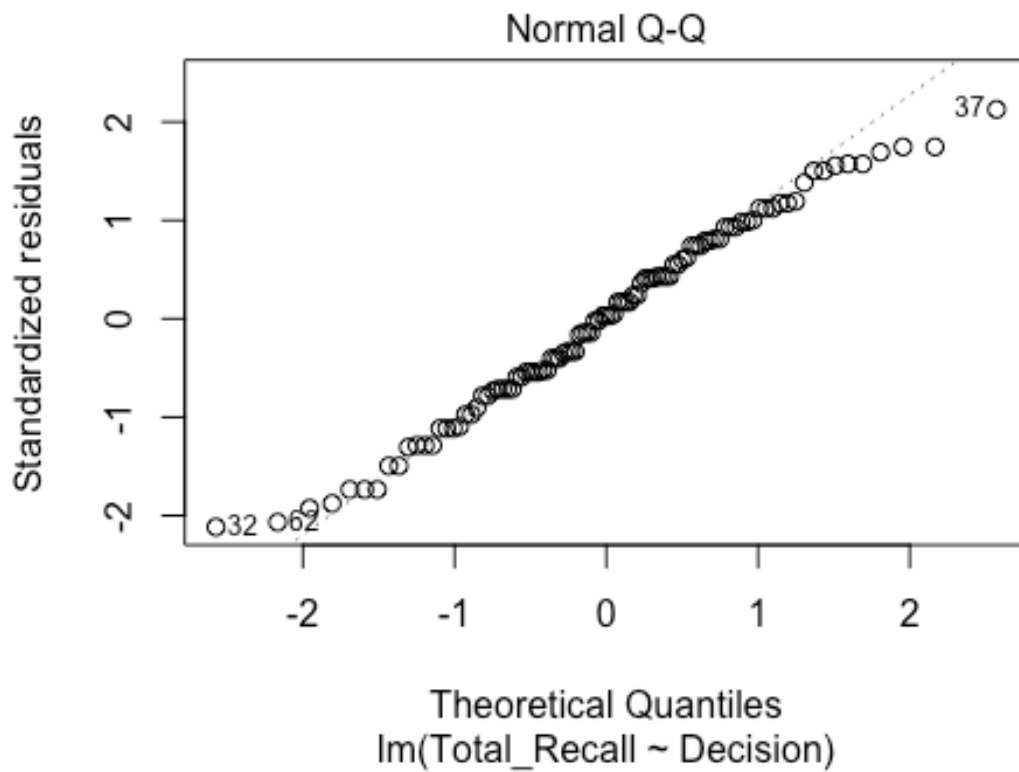
*Assumptions checking*

```
E4_model_delay_recall<- lm(Total_Recall ~ Decision, data =
E4_data_delay)
plot(E4_model_delay_recall, which=1)
```



```
E4_delay_recall_Levene<-leveneTest(Total_Recall ~ Decision, data=
E4_data_delay)
E4_delay_recall_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
```

```
## group 2 1.4976 0.2288
## 96
E4_delay_recall_residuals<-residuals(lm(Total_Recall ~ Decision,
data = E4_data_delay))
shapiro.test(E4_delay_recall_residuals)
##
## Shapiro-Wilk normality test
##
## data: E4_delay_recall_residuals
## W = 0.98338, p-value = 0.2472
plot(E4_model_delay_recall, which= 2)
```



### Inferential statistics

ANOVA

```
E4_anova<-aov(Total_Recall ~ Decision, data=E4_data_delay)
```

```
summary(E4_anova)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## Decision    2     2.4    1.21  0.043  0.958
```

```
## Residuals  96 2725.8   28.39
```

```
## 1 observation deleted due to missingness
```

```
contrast.matrix<-list('No decision vs Any Decision'=c(-1,0.5,0.5),
```

```
                      'Deontological vs Utilitarian'=c(0,-1,+1))
```

```

emm<-emmeans(E4_anova, ~ Decision)

E4_results<-contrast(emm, method=contrast.matrix, side = '>')

E4_contrasts<-summary(E4_results, infer=TRUE)

E4_contrasts

## contrast                estimate    SE df lower.CL upper.CL
t.ratio

## No decision vs Any Decision    0.0884 1.15 96    -1.81      Inf
0.077

## Deontological vs Utilitarian  -0.3643 1.30 96    -2.53      Inf
-0.280

## p.value

## 0.4693

## 0.6098

##

## Confidence level used: 0.95

## P values are right-tailed

E4_Figure<- ggboxplot(E4_data_delay, x = 'Decision', y =
'Total_Recall',
                    ylab = 'Total Recall',
                    add = 'jitter', shape = 'Decision',
                    color= 'Decision',
                    palette= c('grey40', 'grey28', 'grey16')),

```

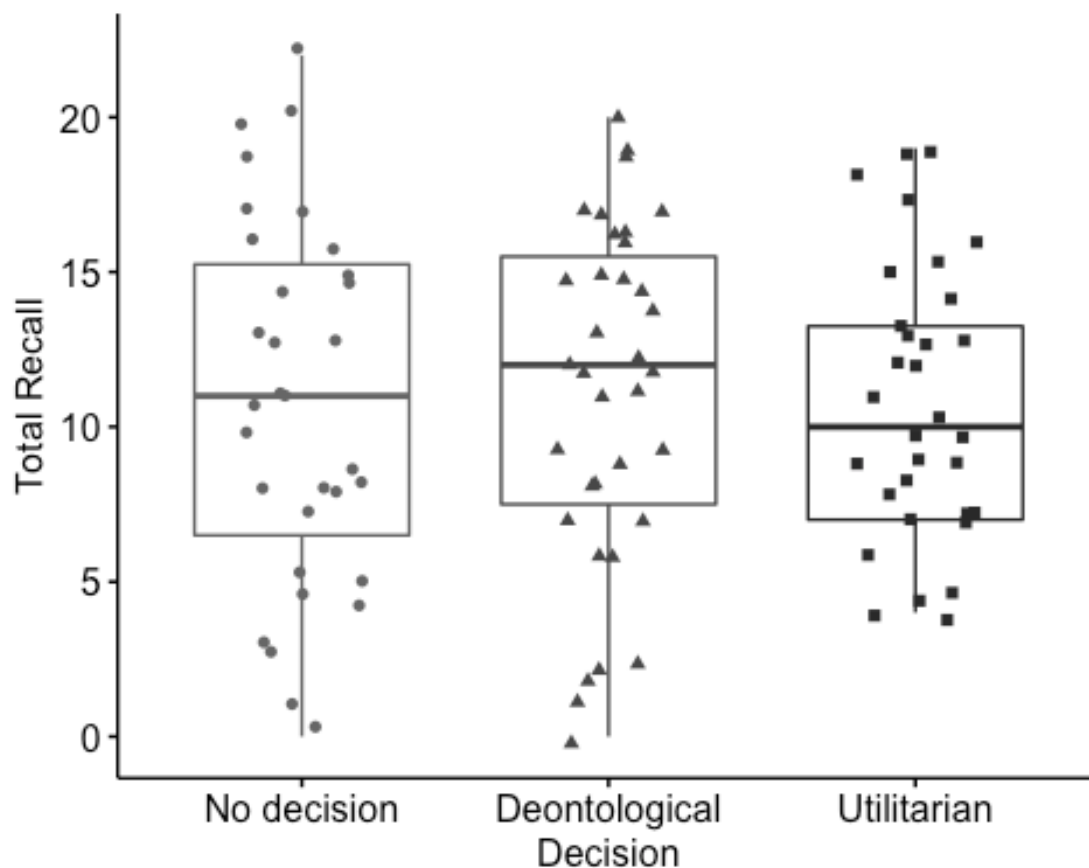
```
legend= "none")
```

```
E4_Figure
```

```
## Warning: Removed 1 rows containing non-finite values
```

```
(stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



Complementary analysis

Descriptive analysis of ideas retention.

```
# Rearranging from Item 1 ...
```

```
data_iI<-subset(Experiment4, Delay == 'Day', select =
```

```
c('Subject_ID', 'Decision', 'Item_1'))
```

```
data_iI$Item<- 1
data_iI<-rename(data_iI,Recall=Item_1)
data_iI<-rename(data_iI,Item_n=Item)
#...2...
data_iII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_2'))
data_iII$Item<- 2
data_iII<-rename(data_iII,Recall=Item_2)
data_iII<-rename(data_iII,Item_n=Item)
#...3...
data_iIII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision','Item_3'))
data_iIII$Item<- 3
data_iIII<-rename(data_iIII,Recall=Item_3)
data_iIII<-rename(data_iIII,Item_n=Item)
#...4...
data_iIV<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision','Item_4'))
data_iIV$Item<- 4
data_iIV<-rename(data_iIV,Recall=Item_4)
data_iIV<-rename(data_iIV,Item_n=Item)
#...5...
data_iV<-subset(Experiment4, Delay == 'Day', select =
```

```
c('Subject_ID', 'Decision', 'Item_5'))  
data_iV$Item<- 5  
data_iV<-rename(data_iV,Recall=Item_5)  
data_iV<-rename(data_iV,Item_n=Item)  
#...6...  
data_iVI<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_6'))  
data_iVI$Item<- 6  
data_iVI<-rename(data_iVI,Recall=Item_6)  
data_iVI<-rename(data_iVI,Item_n=Item)  
#...7...  
data_iVII<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_7'))  
data_iVII$Item<- 7  
data_iVII<-rename(data_iVII,Recall=Item_7)  
data_iVII<-rename(data_iVII,Item_n=Item)  
#...8...  
data_iVIII<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_8'))  
data_iVIII$Item<- 8  
data_iVIII<-rename(data_iVIII,Recall=Item_8)  
data_iVIII<-rename(data_iVIII,Item_n=Item)  
#...9...
```

```
data_iIX<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_9'))
data_iIX$Item<- 9
data_iIX<-rename(data_iIX,Recall=Item_9)
data_iIX<-rename(data_iIX,Item_n=Item)
#...10...
data_iX<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_10'))
data_iX$Item<- 10
data_iX<-rename(data_iX,Recall=Item_10)
data_iX<-rename(data_iX,Item_n=Item)
#...11...
data_iXI<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_11'))
data_iXI$Item<- 11
data_iXI<-rename(data_iXI,Recall=Item_11)
data_iXI<-rename(data_iXI,Item_n=Item)
#...12...
data_iXII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_12'))
data_iXII$Item<- 12
data_iXII<-rename(data_iXII,Recall=Item_12)
data_iXII<-rename(data_iXII,Item_n=Item)
```

```
#...13...
```

```
data_iXIII<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_13'))
```

```
data_iXIII$Item<- 13
```

```
data_iXIII<-rename(data_iXIII,Recall=Item_13)
```

```
data_iXIII<-rename(data_iXIII,Item_n=Item)
```

```
#...14...
```

```
data_iXIV<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_14'))
```

```
data_iXIV$Item<- 14
```

```
data_iXIV<-rename(data_iXIV,Recall=Item_14)
```

```
data_iXIV<-rename(data_iXIV,Item_n=Item)
```

```
#...15...
```

```
data_iXV<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_15'))
```

```
data_iXV$Item<- 15
```

```
data_iXV<-rename(data_iXV,Recall=Item_15)
```

```
data_iXV<-rename(data_iXV,Item_n=Item)
```

```
#...16...
```

```
data_iXVI<-subset(Experiment4, Delay == 'Day', select =  
c('Subject_ID', 'Decision', 'Item_16'))
```

```
data_iXVI$Item<- 16
```

```
data_iXVI<-rename(data_iXVI,Recall=Item_16)
```

```
data_iXVI<-rename(data_iXVI,Item_n=Item)
#...17...
data_iXVII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_17'))
data_iXVII$Item<- 17
data_iXVII<-rename(data_iXVII,Recall=Item_17)
data_iXVII<-rename(data_iXVII,Item_n=Item)
#...18...
data_iXVIII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_18'))
data_iXVIII$Item<- 18
data_iXVIII<-rename(data_iXVIII,Recall=Item_18)
data_iXVIII<-rename(data_iXVIII,Item_n=Item)
#...19...
data_iXIX<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_19'))
data_iXIX$Item<- 19
data_iXIX<-rename(data_iXIX,Recall=Item_19)
data_iXIX<-rename(data_iXIX,Item_n=Item)
#...20...
data_iXX<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_20'))
data_iXX$Item<- 20
```

```
data_iXX<-rename(data_iXX,Recall=Item_20)
data_iXX<-rename(data_iXX,Item_n=Item)
#...21...
data_iXXI<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_21'))
data_iXXI$Item<- 21
data_iXXI<-rename(data_iXXI,Recall=Item_21)
data_iXXI<-rename(data_iXXI,Item_n=Item)
#...22...
data_iXXII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_22'))
data_iXXII$Item<- 22
data_iXXII<-rename(data_iXXII,Recall=Item_22)
data_iXXII<-rename(data_iXXII,Item_n=Item)
#...23...
data_iXXIII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_23'))
data_iXXIII$Item<- 23
data_iXXIII<-rename(data_iXXIII,Recall=Item_23)
data_iXXIII<-rename(data_iXXIII,Item_n=Item)
#...24...
data_iXXIV<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_24'))
```

```

data_iXXIV$Item<- 24

data_iXXIV<-rename(data_iXXIV,Recall=Item_24)

data_iXXIV<-rename(data_iXXIV,Item_n=Item)

#...25...

data_iXXV<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_25'))

data_iXXV$Item<- 25

data_iXXV<-rename(data_iXXV,Recall=Item_25)

data_iXXV<-rename(data_iXXV,Item_n=Item)

#...26...

data_iXXVI<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_26'))

data_iXXVI$Item<- 26

data_iXXVI<-rename(data_iXXVI,Recall=Item_26)

data_iXXVI<-rename(data_iXXVI,Item_n=Item)

#...27...

data_iXXVII<-subset(Experiment4, Delay == 'Day', select =
c('Subject_ID', 'Decision', 'Item_27'))

data_iXXVII$Item<- 27

data_iXXVII<-rename(data_iXXVII,Recall=Item_27)

data_iXXVII<-rename(data_iXXVII,Item_n=Item)

#combining them altogether

E4_data_Item<-rbind(data_iI,data_iII, data_iIII,data_iIV,

```

```

data_iV,data_iVI, data_iVII, data_iVIII, data_iIX,data_iX,
    data_iXI,data_iXII, data_iXIII,data_iXIV,
data_iXV,data_iXVI, data_iXVII, data_iXVIII, data_iXIX,data_iXX,
    data_iXXI,data_iXXII, data_iXXIII,data_iXXIV,
data_iXXV,data_iXXVI, data_iXXVII)

#and setting variables

E4_data_Item$Recall<-as.numeric(E4_data_Item$Recall)

## Warning: NAs introduced by coercion

E4_data_Item$Item_n<-as.factor(E4_data_Item$Item_n)

E4_data_Item$Subject_ID<-as.factor(E4_data_Item$Subject_ID)

E4_data_Item$Decision<- factor(E4_data_Item$Decision, levels =c('No
decision','Deontological','Utilitarian'))

# Count the occurrence of each ideas being recall at each delay and
for each decision group

E4_data_Item <- aggregate( Recall ~ Decision + Item_n,

    FUN = sum,

    data = E4_data_Item )

# Plotting the data

E4_plot_idea_recall <- ggbarplot(E4_data_Item, x = 'Item_n', y =
'Recall',

    color = 'Decision', fill= 'Decision', palette =
c('grey40','grey28', 'grey16'),

    ylab = 'Recall occurrence', xlab = 'idea number_',

```

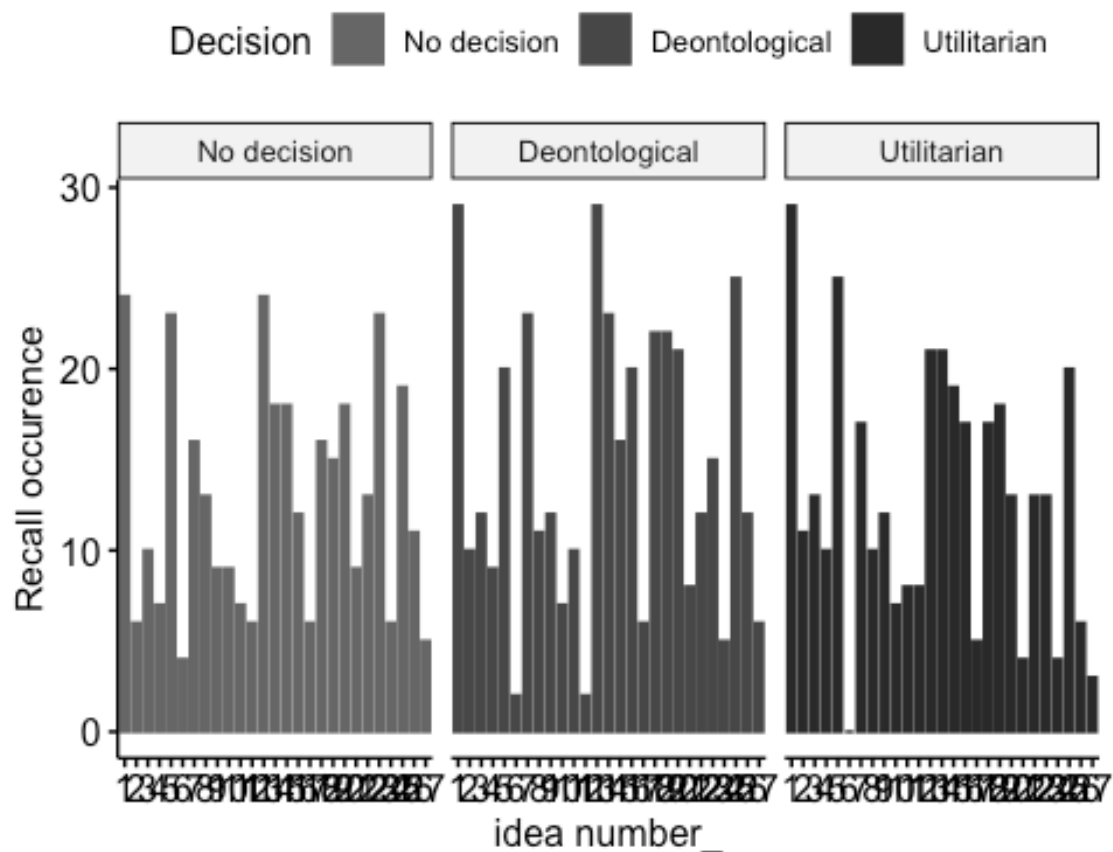
```
position = position_dodge(),
```

```
)
```

```
E4_Figure_idea_recall_occurrence<-facet(E4_plot_idea_recall,
```

```
facet.by = c('Decision'))
```

```
E4_Figure_idea_recall_occurrence
```



Forgetting score (like Murty et al., 2019)

```
# Setting up the data
```

```
E4_data_ctrl<-rename(E4_data_ctrl,ctrl_Recall=Total_Recall)
```

```
E4_data_delay<-rename(E4_data_delay,Delay_Recall=Total_Recall)
```

```
E4_data_forget<-cbind(E4_data_ctrl, E4_data_delay['Delay_Recall'])
```

```
E4_data_forget$ctrl_Recall<-as.numeric(E4_data_forget$ctrl_Recall)
```

```
E4_data_forget<-na.omit(E4_data_forget)
```

```
#Compute forgetting score
```

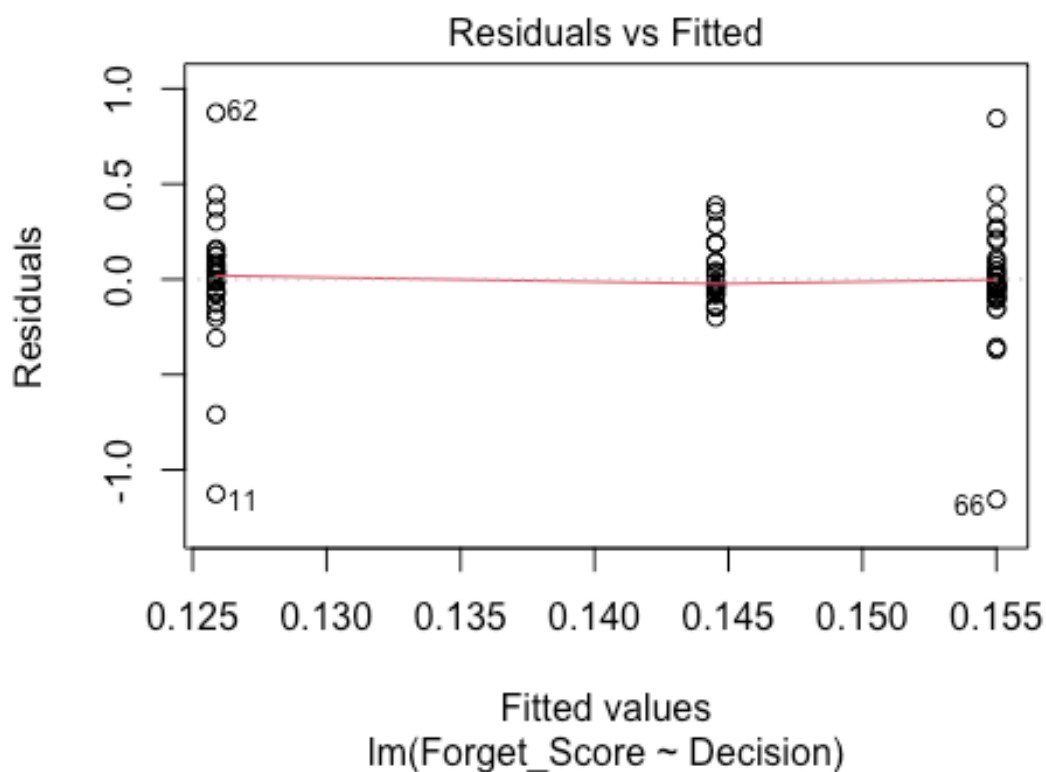
```
E4_data_forget$Forget_Score<-(E4_data_forget$ctrl_Recall-
```

```
E4_data_forget$Delay_Recall)/E4_data_forget$ctrl_Recall
```

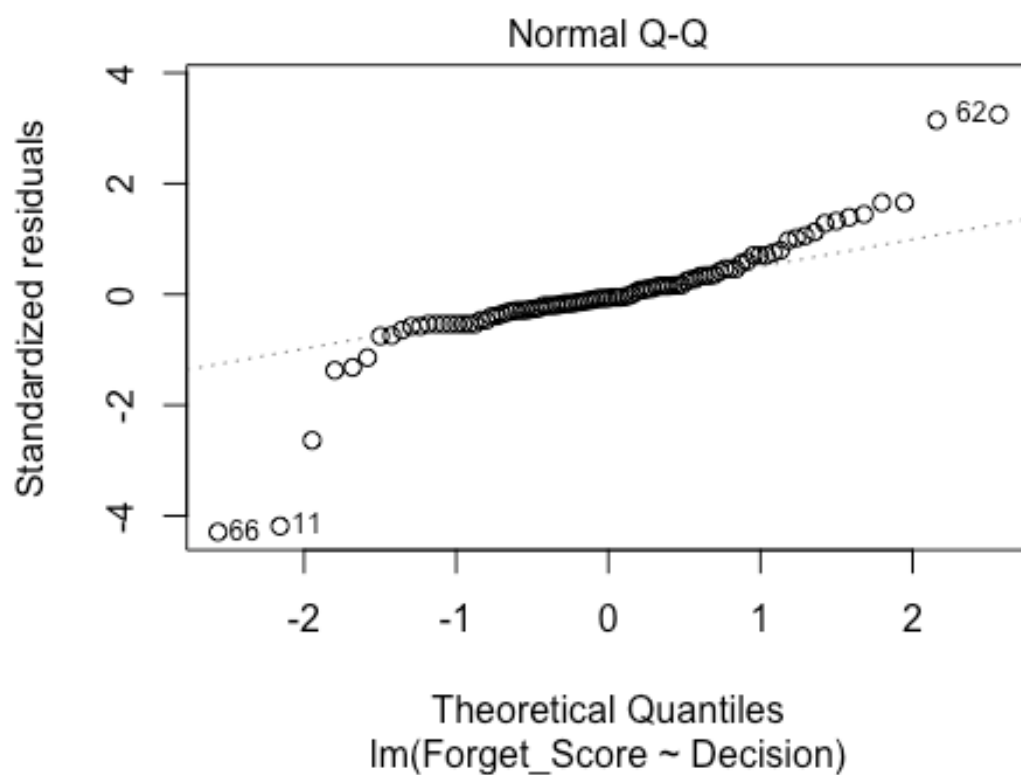
```
Assumptions checking
```

```
E4_model_forget<- lm(Forget_Score ~ Decision, data = E4_data_forget)
```

```
plot(E4_model_forget, which=1)
```



```
E4_forget_Levene<-leveneTest(Forget_Score ~ Decision, data=
E4_data_forget)
E4_forget_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  2  1.2582 0.2889
##      94
E4_forget_residuals<-residuals(lm(Forget_Score ~ Decision, data =
E4_data_forget))
shapiro.test(E4_forget_residuals)
##
## Shapiro-Wilk normality test
##
## data:  E4_forget_residuals
## W = 0.82801, p-value = 2.946e-09
plot(E4_model_forget, which= 2)
```



```
E4_forget_KW<-kruskal.test(Forget_Score ~ Decision, data=
E4_data_forget)
E4_forget_KW
##
## Kruskal-Wallis rank sum test
##
## data: Forget_Score by Decision
## Kruskal-Wallis chi-squared = 0.7259, df = 2, p-value = 0.6956
E4_anova_comp<-aov(Forget_Score ~ Decision, data=E4_data_forget)
summary(E4_anova_comp)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Decision     2  0.014  0.00705   0.094   0.91
## Residuals   94  7.018  0.07466
```

---

Experiment 5: The Impact of Moral Decision Types on Long-Term Forgetting Up to One Month Delay

```
data(Experiment5, package = 'PhDCarolane')
```

```
E5_data_ctrl<-subset(Experiment5, TI == 'Immediate_Recall', select =
c('Subject_ID', 'Decision', 'Total_Recall', 'DSF', 'DSB'))
```

```
E5_data_ctrl$Decision<-as.factor(E5_data_ctrl$Decision)
```

```
E5_data_ctrl$Total_Recall<-as.numeric(E5_data_ctrl$Total_Recall)
```

*N.B: As a matter of data protection, participants' demographics are not provided in the dataset*

Effect of moral decision type on post decision recall

### Descriptive statistics

```
E5_post_recall<-E5_data_ctrl %>%
```

```
  group_by(Decision) %>%
```

```
  summarise( N = n(), M = mean(Total_Recall, na.rm=TRUE), SD =
sd(Total_Recall, na.rm=TRUE),
```

```
            Min = min(Total_Recall, na.rm=TRUE),
```

```
            Median = median(Total_Recall, na.rm=TRUE),
```

```
            Max = max(Total_Recall, na.rm=TRUE))
```

```
E5_post_recall<-APA.descr.table(E5_post_recall)
```

```
E5_post_recall
```

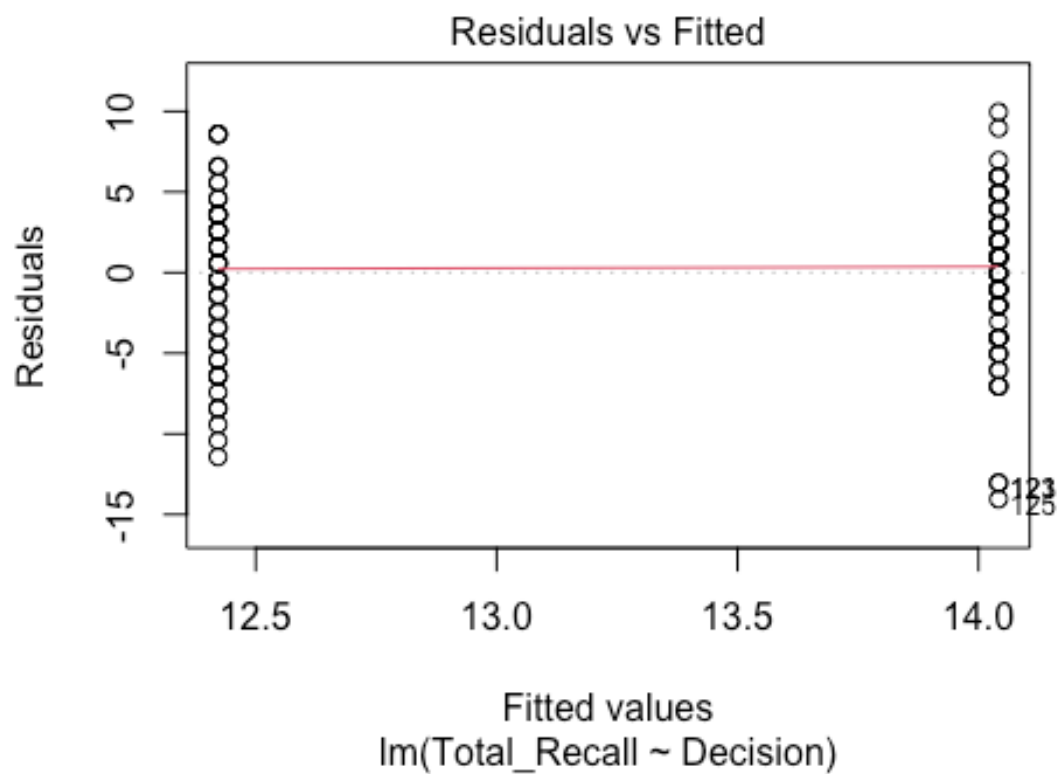
### Inferential statistics

*Assumptions checking*

```
E5_model_post_recall<- lm(Total_Recall ~ Decision, data =
```

```
E5_data_ctrl)
```

```
plot(E5_model_post_recall, which=1)
```



```
E5_post_recall_Levene<-leveneTest(Total_Recall ~ Decision, data=
```

```
E5_data_ctrl)
```

```
E5_post_recall_Levene
```

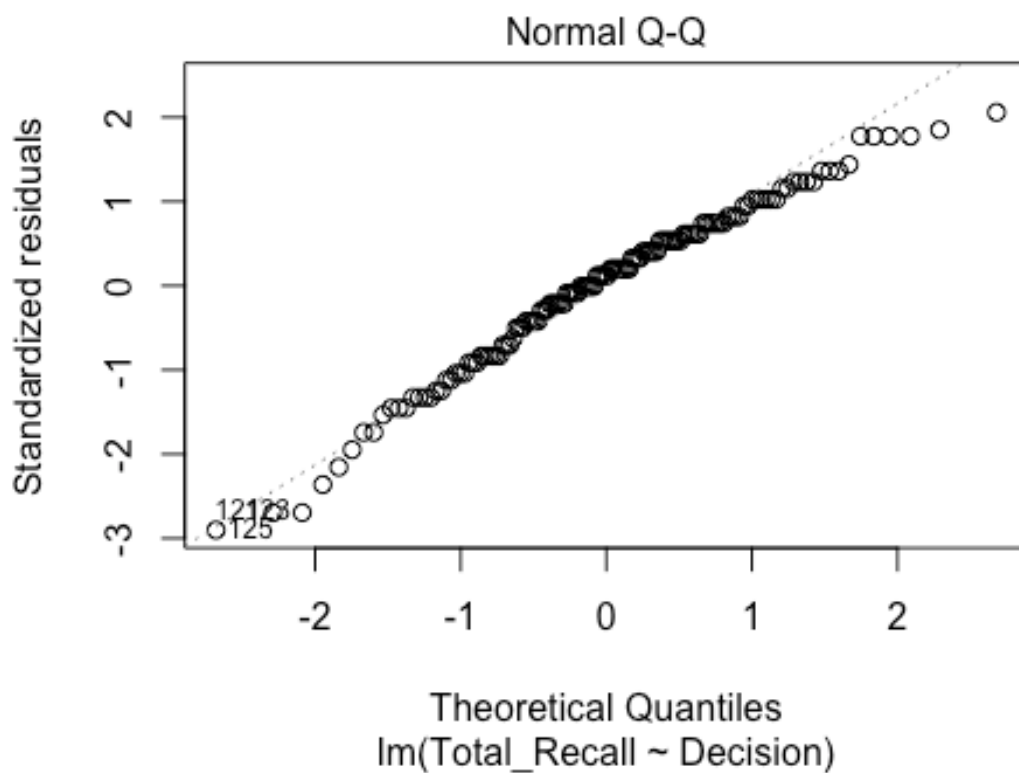
```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  1   0.386 0.5355
##      134

E5_post_recall_residuals<-residuals(lm(Total_Recall ~ Decision, data
= E5_data_ctrl))

shapiro.test(E5_post_recall_residuals)

##
## Shapiro-Wilk normality test
##
## data:  E5_post_recall_residuals
## W = 0.9774, p-value = 0.02321

plot(E5_model_post_recall, which= 2)
```



```
E5_post_recall_Utest<-
wilcox.test(E5_data_ctrl$Total_Recall~E5_data_ctrl$Decision,
var.equal=TRUE)
E5_post_recall_Utest
##
## Wilcoxon rank sum test with continuity correction
##
## data:  E5_data_ctrl$Total_Recall by E5_data_ctrl$Decision
## W = 1858.5, p-value = 0.05178
## alternative hypothesis: true location shift is not equal to 0
```

Effect of moral decision type on DSF

### Descriptive statistics

```
E5_DSf<-E5_data_ctrl %>%  
  group_by(Decision) %>%  
  summarise( N = n(), M = mean(DSF, na.rm=TRUE), SD = sd(DSF,  
na.rm=TRUE),  
            Min = min(DSF, na.rm=TRUE),  
            Median = median(DSF, na.rm=TRUE),  
            Max = max(DSF, na.rm=TRUE))
```

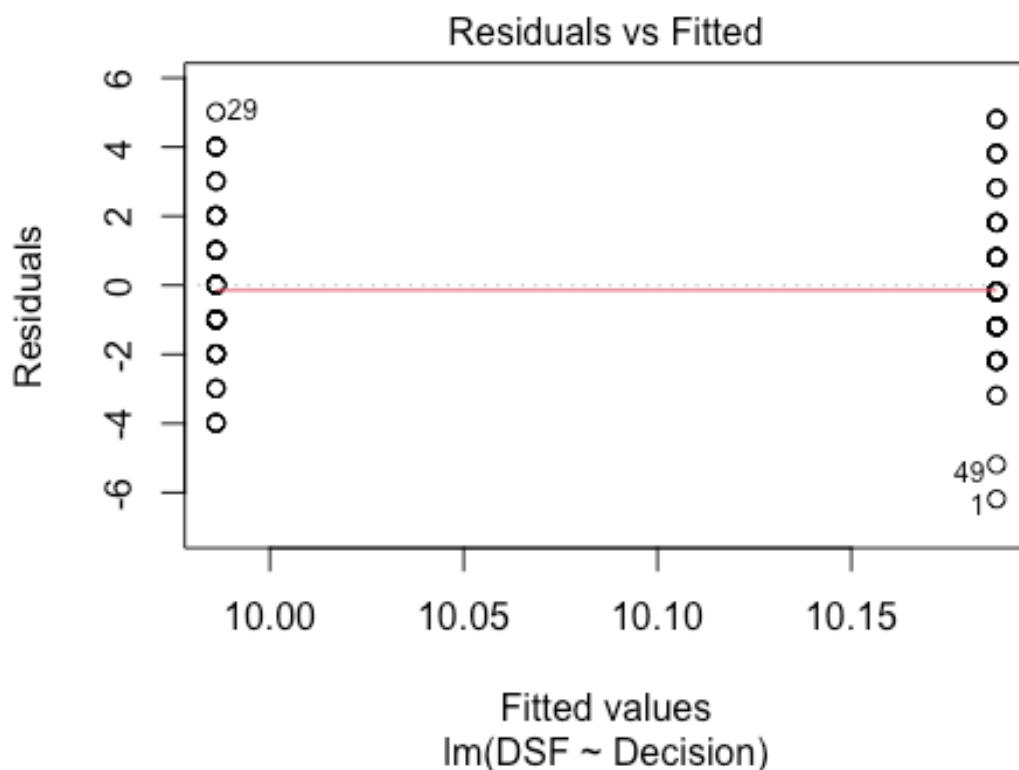
```
E5_DSf<-APA.descr.table(E5_DSf)
```

```
E5_DSf
```

### Inferential statistics

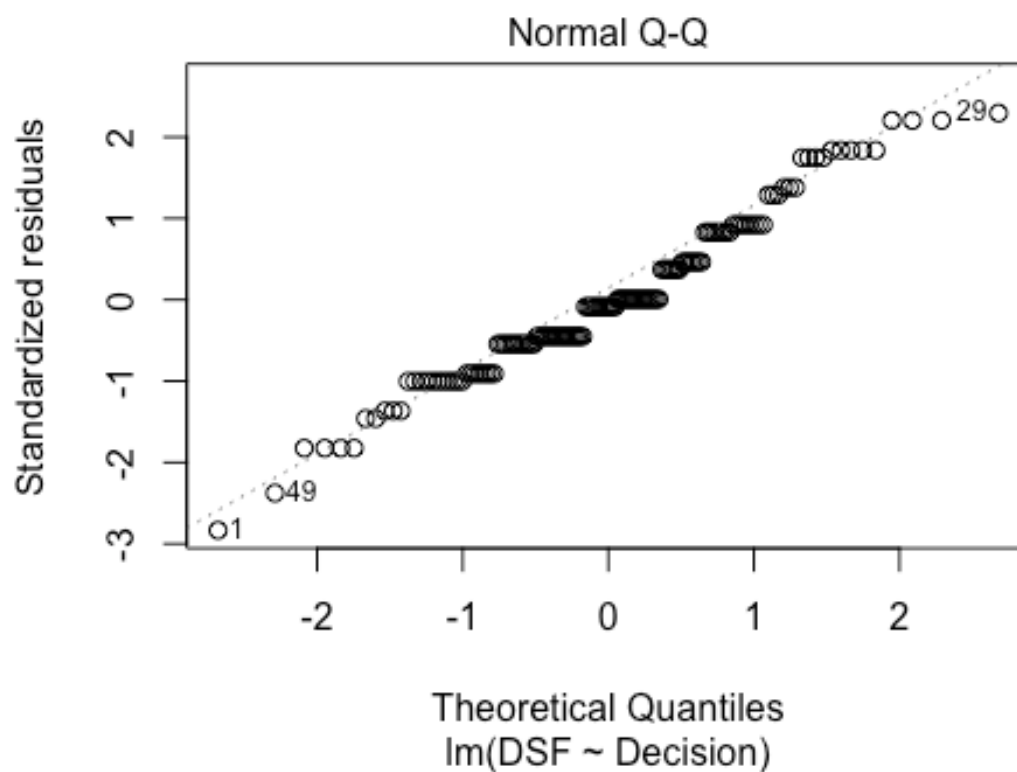
*Assumptions checking*

```
E5_model_DSf<- lm(DSF ~ Decision, data = E5_data_ctrl)  
plot(E5_model_DSf, which=1)
```



```
E5_DSF_Levene<-leveneTest(DSF ~ Decision, data= E5_data_ctrl)
E5_DSF_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  1  0.5819 0.4469
##      134
E5_DSF_residuals<-residuals(lm(DSF ~ Decision, data = E5_data_ctrl))
shapiro.test(E5_DSF_residuals)
##
## Shapiro-Wilk normality test
```

```
##
## data: E5_DSF_residuals
## W = 0.97401, p-value = 0.01054
plot(E5_model_DSF, which= 2)
```



```
E5_DSF_Utest<-wilcox.test(E5_data_ctrl$DSF~E5_data_ctrl$Decision,
var.equal=TRUE)
```

```
E5_DSF_Utest
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: E5_data_ctrl$DSB by E5_data_ctrl$Decision
## W = 2408.5, p-value = 0.6462
## alternative hypothesis: true location shift is not equal to 0

Effect of decision groups on DSB

Descriptive statistics

E5_DSB<-E5_data_ctrl %>%
  group_by(Decision) %>%
  summarise( N = n(), M = mean(DSB, na.rm=TRUE), SD = sd(DSB,
na.rm=TRUE),
            Min = min(DSB, na.rm=TRUE),
            Median = median(DSB, na.rm=TRUE),
            Max = max(DSB, na.rm=TRUE))

E5_DSB<-APA.descr.table(E5_DSB)

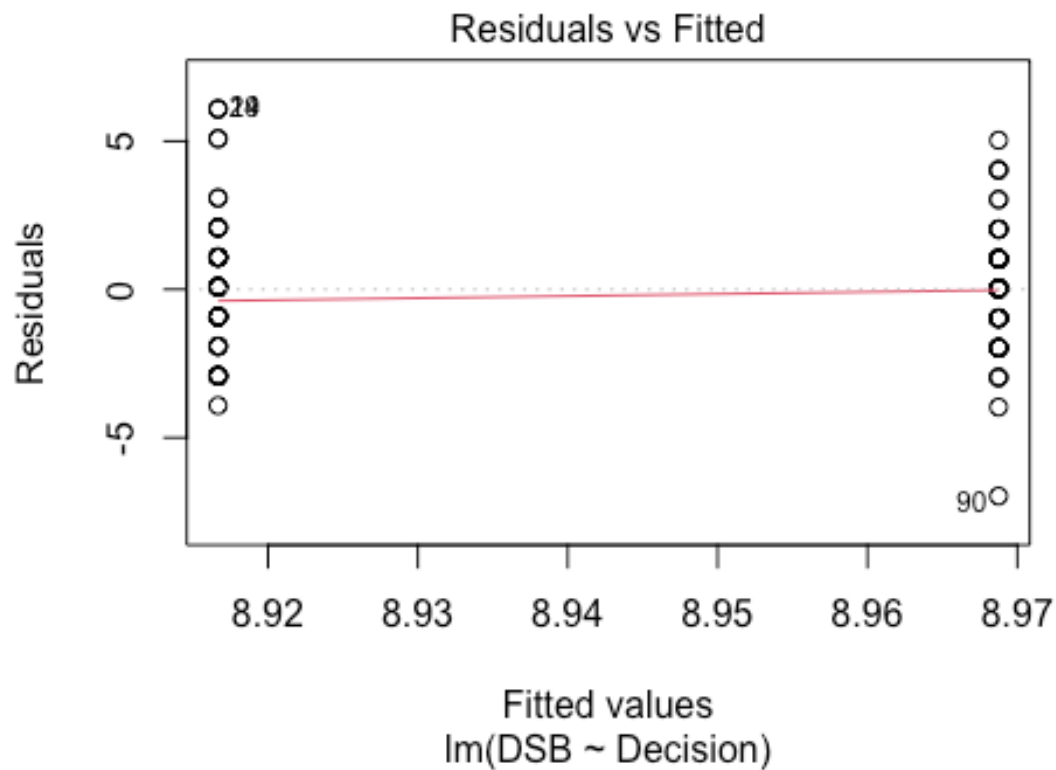
E5_DSB

Inferential statistics

Assumptions checking

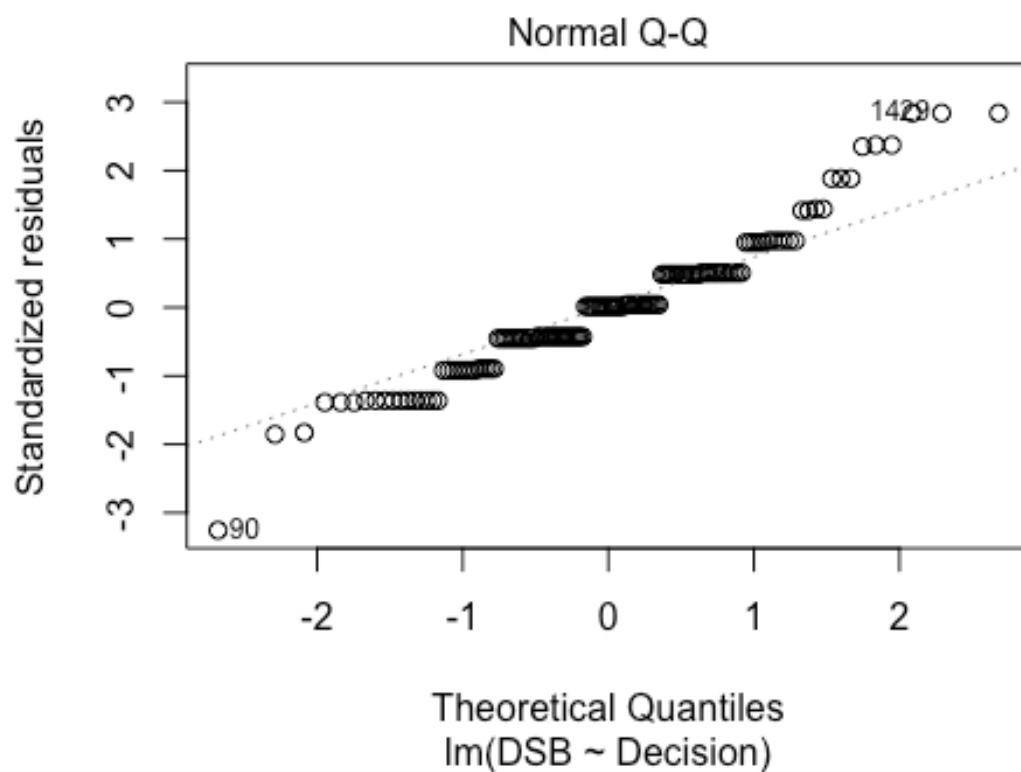
E5_model_DSB<- lm(DSB ~ Decision, data = E5_data_ctrl)

plot(E5_model_DSB, which=1)
```



```
E5_DSB_Levene<-leveneTest(DSB ~ Decision, data= E5_data_ctrl)
E5_DSB_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  1  0.8283 0.3644
##      134
E5_DSB_residuals<-residuals(lm(DSB ~ Decision, data = E5_data_ctrl))
shapiro.test(E5_DSB_residuals)
##
## Shapiro-Wilk normality test
```

```
##
## data: E5_DSB_residuals
## W = 0.95427, p-value = 0.0001709
plot(E5_model_DSB, which= 2)
```



```
E5_DSB_Utest<-wilcox.test(E5_data_ctrl$DSB~E5_data_ctrl$Decision,
var.equal=TRUE)
E5_DSB_Utest
##
## Wilcoxon rank sum test with continuity correction
##
```

```

## data: E5_data_ctrl$DSB by E5_data_ctrl$Decision
## W = 2449.5, p-value = 0.5216
## alternative hypothesis: true location shift is not equal to 0
Effect of moral decision types on delayed recall
Descriptive statistics
E5_data_delay<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Total_Recall', 'Delay'))
E5_data_delay$Subject_ID<-as.factor(E5_data_delay$Subject_ID)
E5_data_delay$Decision<-as.factor(E5_data_delay$Decision)
E5_data_delay$Delay<- factor(E5_data_delay$Delay, levels
=c('Day', 'Week', 'Month'))
E5_data_delay$Total_Recall<-as.numeric(E5_data_delay$Total_Recall)
E5_table_Decision <- E5_data_delay %>%
  group_by(Decision) %>%
  summarise( N = n(),
             M = mean(Total_Recall, na.rm= TRUE),
             SD = sd(Total_Recall, na.rm= TRUE),
             Min = min(Total_Recall, na.rm= TRUE),
             Median = median(Total_Recall, na.rm= TRUE),
             Max = max(Total_Recall, na.rm= TRUE))
E5_table_Decision<-APA.descr.table(E5_table_Decision)
E5_table_Decision

```

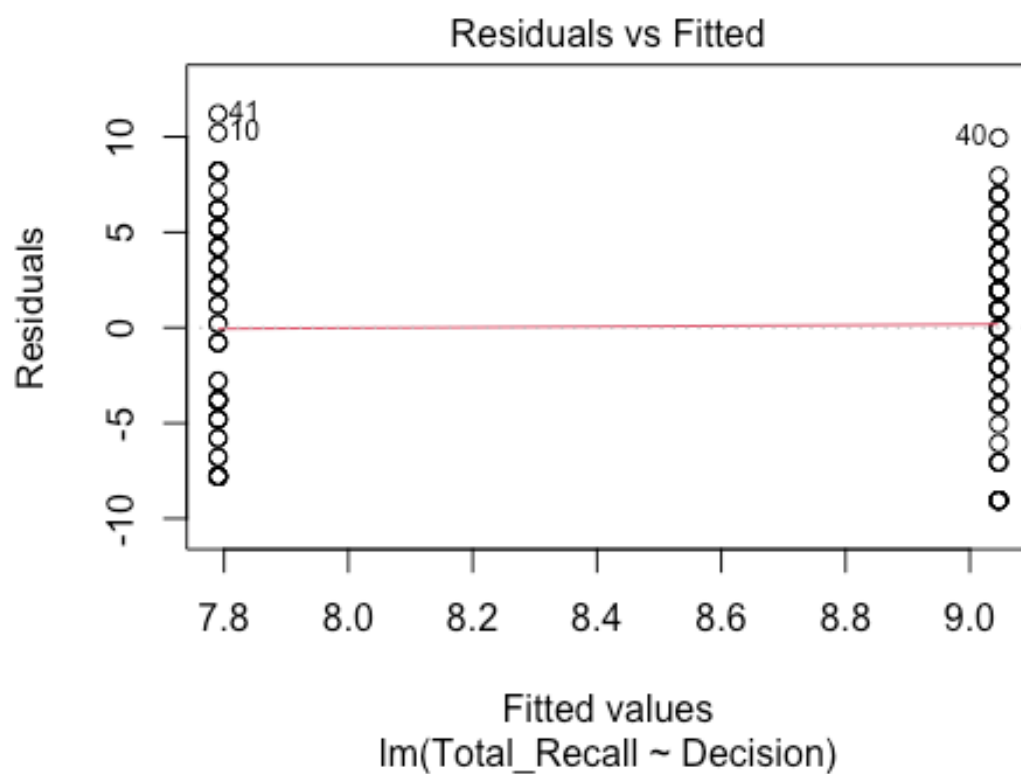
```

E5_table_Delay <- E5_data_delay %>%
  group_by(Delay) %>%
  summarise( N = n(), M = mean(Total_Recall), SD
= sd(Total_Recall),
  Min = min(Total_Recall),
  Median = median(Total_Recall),
  Max = max(Total_Recall))
E5_table_Delay<-APA.descr.table(E5_table_Delay)
E5_table_Delay
E5_Table_Descriptives<-E5_data_delay %>%
  group_by(Decision, Delay) %>%
  summarise( N = n(), M = mean(Total_Recall, na.rm= TRUE),
  SD = sd(Total_Recall,na.rm= TRUE),
  Min = min(Total_Recall, na.rm= TRUE),
  Median = median(Total_Recall, na.rm= TRUE),
  Max = max(Total_Recall, na.rm= TRUE))
## `summarise()` has grouped output by 'Decision'. You can override
using the `.groups` argument.
E5_Table_Descriptives<-APA.descr.table(E5_Table_Descriptives)%>%
  cols_label(Delay='Condition')
E5_Table_Descriptives
###Inferential statistics
Frequentist approach

```

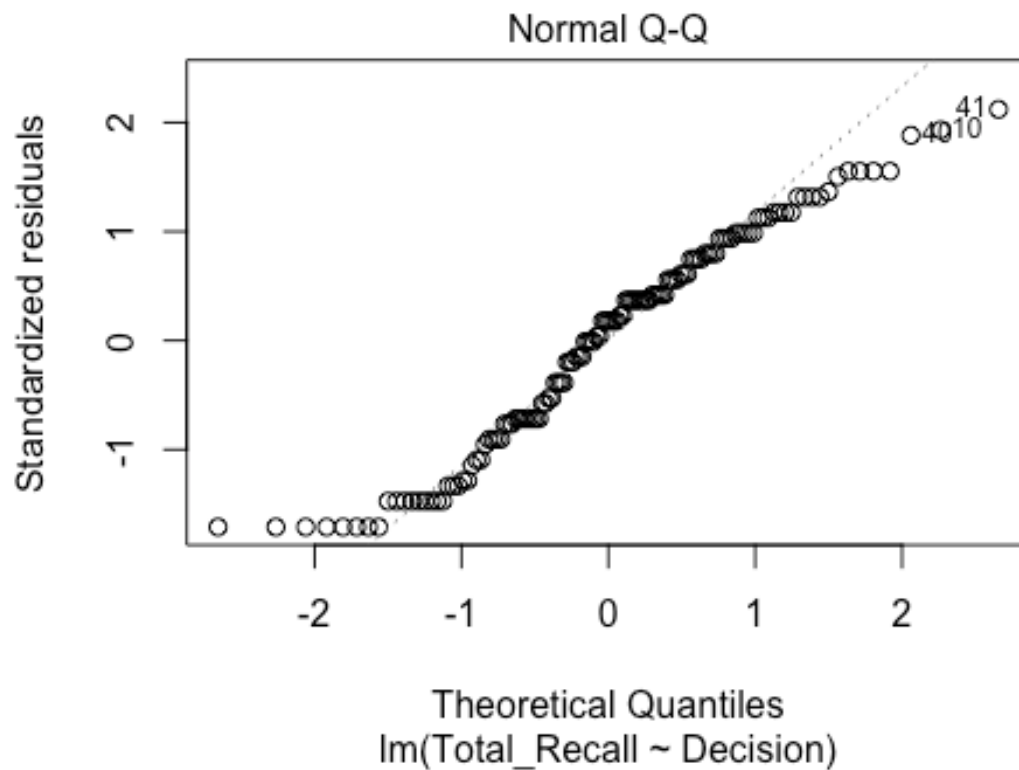
*Assumptions checking*

```
E5_model_delay_recall<- lm(Total_Recall ~ Decision, data =
E5_data_delay)
plot(E5_model_delay_recall, which=1)
```



```
E5_delay_recall_Levene<-leveneTest(Total_Recall ~ Decision, data=
E5_data_delay)
E5_delay_recall_Levene
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  1  2.9643 0.08759 .
```

```
##      125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
E5_delay_recall_residuals<-residuals(lm(Total_Recall ~ Decision,
data = E5_data_delay))
shapiro.test(E5_delay_recall_residuals)
##
## Shapiro-Wilk normality test
##
## data:  E5_delay_recall_residuals
## W = 0.96099, p-value = 0.001033
plot(E5_model_delay_recall, which= 2)
```



*Bootstrapped model*

```
E5_model_inter <- lm(Total_Recall ~ Delay * Decision, data=
```

```
E5_data_delay)
```

```
E5_model_inter_boot <- Boot(E5_model_inter, R=10000)
```

```
Confint(E5_model_inter_boot, type= 'perc') %>%
```

```
  kable(digits= 3, caption= 'Bootstrapped regression model of total  
recall with 95% CIs') %>%
```

```
  kable_styling(full_width= FALSE)
```

Bootstrapped regression model of total recall with 95% CIs

Estimate

2.5 %

97.5 %

(Intercept)

10.143

7.950

12.210

DelayWeek

-1.619

-4.503

1.360

DelayMonth

-5.593

-8.729

-2.238

DecisionUtilitarian

1.497

-1.108

4.107

DelayWeek:DecisionUtilitarian

-0.171

-3.678

3.314

DelayMonth:DecisionUtilitarian

-1.047

-5.493

3.283

Regression model on t distribution assumption - provided for information.

```
model_parameters(E5_model_inter) %>% print_md()
```

Parameter	Coefficient	SE	95% CI	t(121)	p
(Intercept)	10.14	1.03	(8.10, 12.19)	9.83	<.001
Delay (Week)	-1.62	1.46	(-4.51, 1.27)	-1.11	0.269
Delay (Month)	-5.59	1.48	(-8.52, -2.67)	-3.79	<.001
Decision (Utilitarian)	1.50	1.40	(-1.27, 4.27)	1.07	0.287
Delay (Week) * Decision (Utilitarian)	-0.17	2.03	(-4.20, 3.86)	-0.08	0.933
Delay (Month) * Decision (Utilitarian)	-1.05	2.05	(-5.10, 3.01)	-0.51	0.610

```
E5_Figure<-ggplot(E5_data_delay, aes(x=Decision, y=Total_Recall,
```

```
shape=Decision, color=Delay, )) +
```

```
  labs(y= 'Number of ideas correctly recalled') +
```

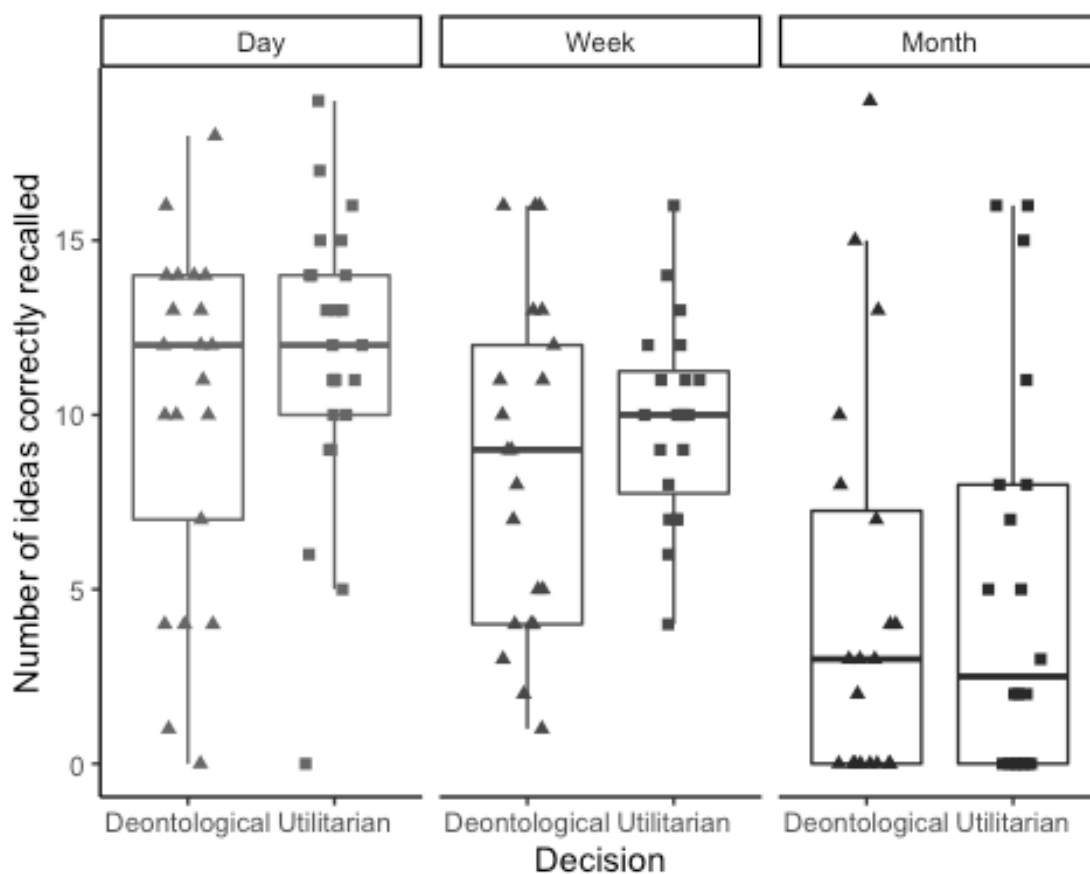
```
  facet_grid(. ~ Delay) +
```

```

geom_boxplot(outlier.shape = NA, show.legend
= FALSE) +
geom_jitter(width = 0.2, height = 0 ,
show.legend = FALSE) +
scale_shape_manual(values = c(17,15))+
scale_color_manual(values =
c('grey40','grey28', 'grey16')) +
theme_classic()

```

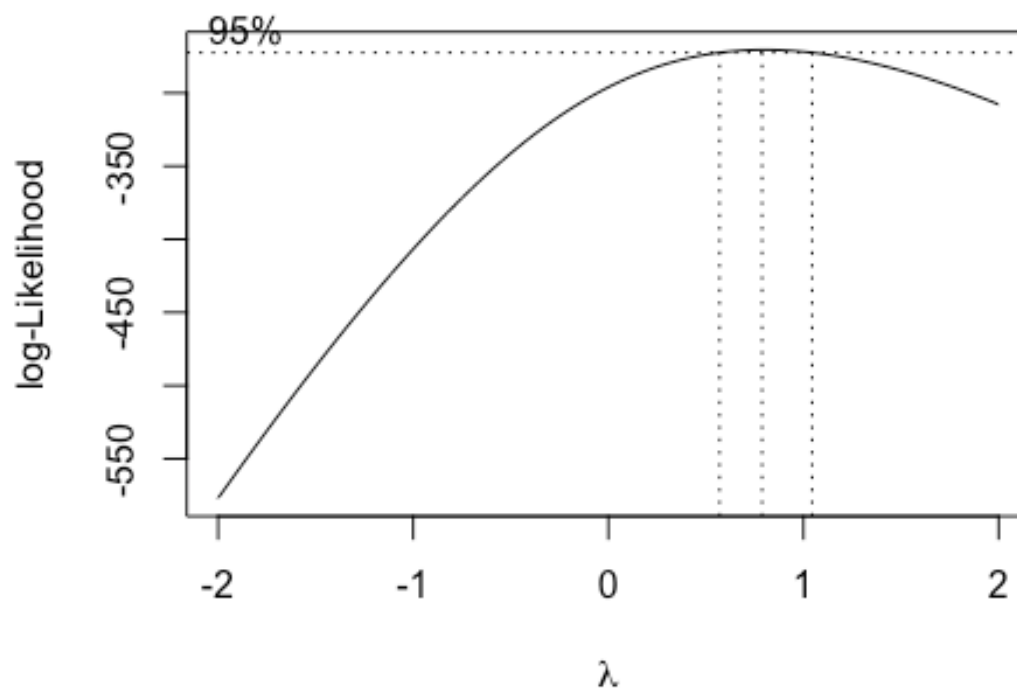
E5\_Figure



AIC & BIC

```
E5_data_delay$Total_Recall_tr<- E5_data_delay$Total_Recall + 1
```

```
E5_boxcox<-boxcox(lm(Total_Recall_tr~1, data= E5_data_delay))
```



```
E5_lambda <- E5_boxcox$x[which.max(E5_boxcox$y)]# give us the lambda
value required to normalise Total_Recall
```

```
E5_data_delay$Total_Recall_norm<-
```

```
E5_data_delay$Total_Recall^E5_lambda
```

```
E5_model_inter<- lm(Total_Recall_norm ~ Delay * Decision, data=
E5_data_delay)
```

```
E5_model_decision<- lm(Total_Recall_norm ~ Delay + Decision, data=
E5_data_delay)
```

```
E5_model_null<- lm(Total_Recall_norm ~ Delay, data= E5_data_delay)
```

```
AIC(E5_model_inter, E5_model_decision)
```

```
##           df      AIC
```

```
## E5_model_inter      7 605.0275
```

```
## E5_model_decision  5 601.3052
```

```
BIC(E5_model_inter, E5_model_decision)
```

```
##           df      BIC
```

```
## E5_model_inter      7 624.9368
```

```
## E5_model_decision  5 615.5262
```

```
AIC(E5_model_null, E5_model_decision)
```

```
##           df      AIC
```

```
## E5_model_null      4 601.2515
```

```
## E5_model_decision  5 601.3052
```

```
BIC(E5_model_null, E5_model_decision)
```

```
##           df      BIC
```

```
## E5_model_null      4 612.6282
```

```
## E5_model_decision  5 615.5262
```

Complementary analysis

```
# Rearranging from Item 1 ...
```

```
data_iI<-subset(Experiment5, TI == 'Delay_Recall', select =
```

```
c('Subject_ID', 'Decision', 'Delay', 'Item_1'))
```

```
data_iI$Item<- 1
data_iI<-rename(data_iI,Recall=Item_1)
data_iI<-rename(data_iI,Item_n=Item)
#...2...
data_iII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay','Item_2'))
data_iII$Item<- 2
data_iII<-rename(data_iII,Recall=Item_2)
data_iII<-rename(data_iII,Item_n=Item)
#...3...
data_iIII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay','Item_3'))
data_iIII$Item<- 3
data_iIII<-rename(data_iIII,Recall=Item_3)
data_iIII<-rename(data_iIII,Item_n=Item)
#...4...
data_iIV<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay','Item_4'))
data_iIV$Item<- 4
data_iIV<-rename(data_iIV,Recall=Item_4)
data_iIV<-rename(data_iIV,Item_n=Item)
#...5...
data_iV<-subset(Experiment5, TI == 'Delay_Recall', select =
```

```
c('Subject_ID', 'Decision', 'Delay', 'Item_5'))  
data_iV$Item<- 5  
data_iV<-rename(data_iV, Recall=Item_5)  
data_iV<-rename(data_iV, Item_n=Item)  
#...6...  
data_iVI<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay', 'Item_6'))  
data_iVI$Item<- 6  
data_iVI<-rename(data_iVI, Recall=Item_6)  
data_iVI<-rename(data_iVI, Item_n=Item)  
#...7...  
data_iVII<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay', 'Item_7'))  
data_iVII$Item<- 7  
data_iVII<-rename(data_iVII, Recall=Item_7)  
data_iVII<-rename(data_iVII, Item_n=Item)  
#...8...  
data_iVIII<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay', 'Item_8'))  
data_iVIII$Item<- 8  
data_iVIII<-rename(data_iVIII, Recall=Item_8)  
data_iVIII<-rename(data_iVIII, Item_n=Item)  
#...9...
```

```
data_iIX<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_9'))
data_iIX$Item<- 9
data_iIX<-rename(data_iIX, Recall=Item_9)
data_iIX<-rename(data_iIX, Item_n=Item)
#...10...
data_iX<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_10'))
data_iX$Item<- 10
data_iX<-rename(data_iX, Recall=Item_10)
data_iX<-rename(data_iX, Item_n=Item)
#...11...
data_iXI<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_11'))
data_iXI$Item<- 11
data_iXI<-rename(data_iXI, Recall=Item_11)
data_iXI<-rename(data_iXI, Item_n=Item)
#...12...
data_iXII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_12'))
data_iXII$Item<- 12
data_iXII<-rename(data_iXII, Recall=Item_12)
data_iXII<-rename(data_iXII, Item_n=Item)
```

```
#...13...
```

```
data_iXIII<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay','Item_13'))
```

```
data_iXIII$Item<- 13
```

```
data_iXIII<-rename(data_iXIII,Recall=Item_13)
```

```
data_iXIII<-rename(data_iXIII,Item_n=Item)
```

```
#...14...
```

```
data_iXIV<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay','Item_14'))
```

```
data_iXIV$Item<- 14
```

```
data_iXIV<-rename(data_iXIV,Recall=Item_14)
```

```
data_iXIV<-rename(data_iXIV,Item_n=Item)
```

```
#...15...
```

```
data_iXV<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay','Item_15'))
```

```
data_iXV$Item<- 15
```

```
data_iXV<-rename(data_iXV,Recall=Item_15)
```

```
data_iXV<-rename(data_iXV,Item_n=Item)
```

```
#...16...
```

```
data_iXVI<-subset(Experiment5, TI == 'Delay_Recall', select =  
c('Subject_ID', 'Decision', 'Delay','Item_16'))
```

```
data_iXVI$Item<- 16
```

```
data_iXVI<-rename(data_iXVI,Recall=Item_16)
```

```
data_iXVI<-rename(data_iXVI,Item_n=Item)
#...17...
data_iXVII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_17'))
data_iXVII$Item<- 17
data_iXVII<-rename(data_iXVII,Recall=Item_17)
data_iXVII<-rename(data_iXVII,Item_n=Item)
#...18...
data_iXVIII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_18'))
data_iXVIII$Item<- 18
data_iXVIII<-rename(data_iXVIII,Recall=Item_18)
data_iXVIII<-rename(data_iXVIII,Item_n=Item)
#...19...
data_iXIX<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_19'))
data_iXIX$Item<- 19
data_iXIX<-rename(data_iXIX,Recall=Item_19)
data_iXIX<-rename(data_iXIX,Item_n=Item)
#...20...
data_iXX<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_20'))
data_iXX$Item<- 20
```

```
data_iXX<-rename(data_iXX,Recall=Item_20)
data_iXX<-rename(data_iXX,Item_n=Item)
#...21...
data_iXXI<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_21'))
data_iXXI$Item<- 21
data_iXXI<-rename(data_iXXI,Recall=Item_21)
data_iXXI<-rename(data_iXXI,Item_n=Item)
#...22...
data_iXXII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_22'))
data_iXXII$Item<- 22
data_iXXII<-rename(data_iXXII,Recall=Item_22)
data_iXXII<-rename(data_iXXII,Item_n=Item)
#...23...
data_iXXIII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_23'))
data_iXXIII$Item<- 23
data_iXXIII<-rename(data_iXXIII,Recall=Item_23)
data_iXXIII<-rename(data_iXXIII,Item_n=Item)
#...24...
data_iXXIV<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay', 'Item_24'))
```

```

data_iXXIV$Item<- 24

data_iXXIV<-rename(data_iXXIV,Recall=Item_24)

data_iXXIV<-rename(data_iXXIV,Item_n=Item)

#...25...

data_iXXV<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay','Item_25'))

data_iXXV$Item<- 25

data_iXXV<-rename(data_iXXV,Recall=Item_25)

data_iXXV<-rename(data_iXXV,Item_n=Item)

#...26...

data_iXXVI<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay','Item_26'))

data_iXXVI$Item<- 26

data_iXXVI<-rename(data_iXXVI,Recall=Item_26)

data_iXXVI<-rename(data_iXXVI,Item_n=Item)

#...27...

data_iXXVII<-subset(Experiment5, TI == 'Delay_Recall', select =
c('Subject_ID', 'Decision', 'Delay','Item_27'))

data_iXXVII$Item<- 27

data_iXXVII<-rename(data_iXXVII,Recall=Item_27)

data_iXXVII<-rename(data_iXXVII,Item_n=Item)

#combining them altogether

E5_data_Item<-rbind(data_iI,data_iII, data_iIII,data_iIV,

```

```

data_iV,data_iVI, data_iVII, data_iVIII, data_iIX,data_iX,
    data_iXI,data_iXII, data_iXIII,data_iXIV,
data_iXV,data_iXVI, data_iXVII, data_iXVIII, data_iXIX,data_iXX,
    data_iXXI,data_iXXII, data_iXXIII,data_iXXIV,
data_iXXV,data_iXXVI, data_iXXVII)

#and setting variables

E5_data_Item$Recall<-as.numeric(E5_data_Item$Recall)
E5_data_Item$Item_n<-as.factor(E5_data_Item$Item_n)
E5_data_Item$Decision<- factor(E5_data_Item$Decision, levels
=c('Deontological','Utilitarian'))

# Count the occurrence of each items being recall at each delay and
for each decision group

E5_items_recall <- aggregate( Recall ~ Decision + Delay + Item_n,
    FUN = sum,
    data = E5_data_Item )

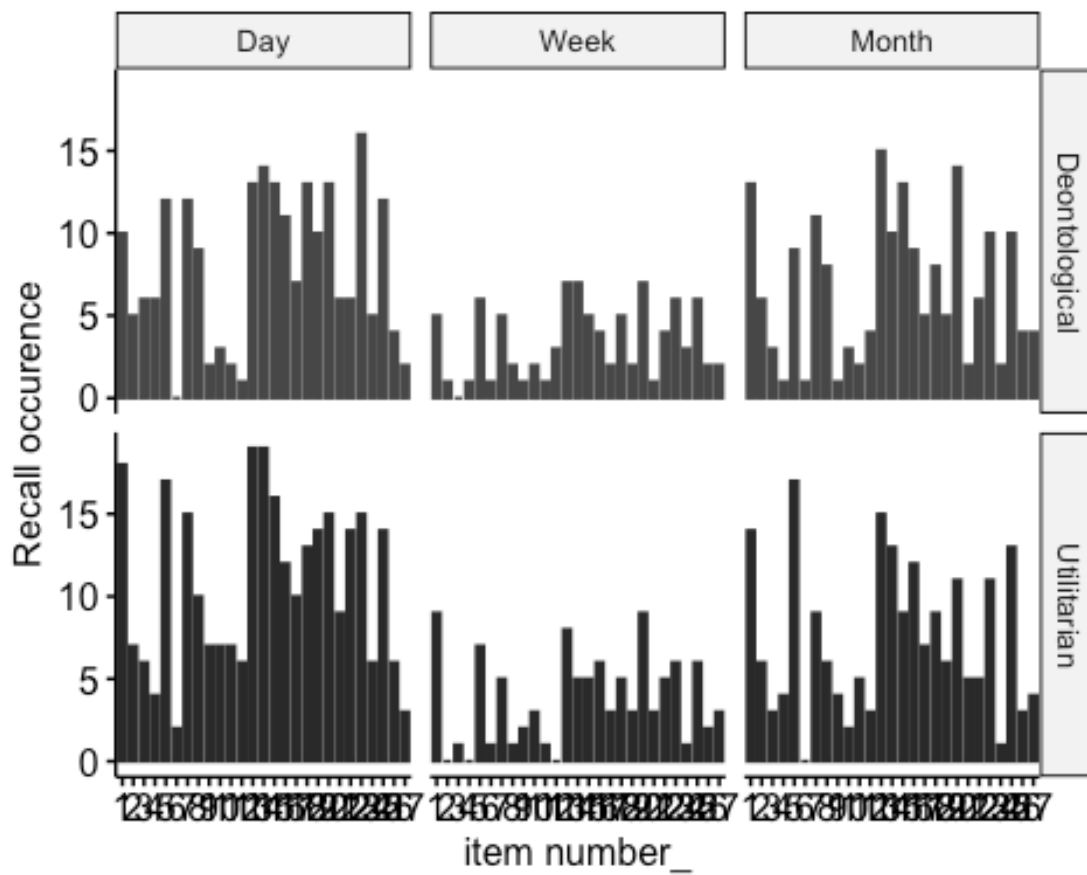
# Plotting the data

E5_plot_item_recall <- ggbarplot(E5_items_recall, x = 'Item_n', y =
'Recall',
    color = 'Decision', fill= 'Decision', palette =
c('grey28', 'grey16'),
    ylab = 'Recall occurrence', xlab = 'item number_',
position = position_dodge())

E5_Figure_item_recall_occurrence<-facet(E5_plot_item_recall,

```

```
facet.by = c('Decision',  
'Delay'),  
short.panel.labs = FALSE,  
panel.labs = list(  
  Delay = c('Day', 'Week',  
'Month'),  
  Decision =  
c('Deontological', 'Utilitarian')  
)  
) +  
  remove('legend')  
E5_Figure_item_recall_occurence
```




---

EXPERIMENT 6: The Effect of Decision-Making on Long-Term Forgetting as a Function of Learning

```
data(Experiment6, package = 'PhDCarolane')
```

```
# Factor variables
```

```
Experiment6[, c('Subject_ID', 'Decision', 'Retrieval_Practice')] <-  
sapply(Experiment6[, c('Subject_ID', 'Decision',  
'Retrieval_Practice')], as.factor)
```

```
Experiment6$Decision[Experiment6$Decision == 'No_decision'] <- 'No
```

```
decision'
```

```
Experiment6$Decision<- factor(Experiment6$Decision, levels= c('No
decision','Deontological','Utilitarian')) # I put the No decision
Level as the Level of reference.
```

```
Experiment6$Retrieval_Practice[Experiment6$Retrieval_Practice ==
'With_RP']<- 'With RP'
```

```
Experiment6$Retrieval_Practice[Experiment6$Retrieval_Practice ==
'Without_RP']<- 'Without RP'
```

```
Experiment6$Retrieval_Practice<-
factor(Experiment6$Retrieval_Practice, levels= c('Without RP','With
RP')) # I put the Without RP Level as the Level of reference.
```

```
#Numerical variables
```

```
Experiment6[, c('DSF', 'DSB', 'Total_Recall')] <-
sapply(Experiment6[, c('DSF', 'DSB', 'Total_Recall')], as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by
coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by
coercion
```

```
Experiment6$Total_Recall<-replace(Experiment6$Total_Recall,
Experiment6$Total_Recall=='NaN', NA)
```

*# The demographics file non included here shows that 7 participants are over 50.*

*# They are identified under the Subject\_ID 47, 66, 83, 98, 146, 158, 163*

```
Experiment6<-Experiment6[Experiment6$Subject_ID != 47 &
Experiment6$Subject_ID != 66 & Experiment6$Subject_ID != 83 &
Experiment6$Subject_ID != 98 & Experiment6$Subject_ID != 146 &
Experiment6$Subject_ID != 158 & Experiment6$Subject_ID != 163, ]
Experiment6%>%
```

```
  group_by(Subject_ID)%>%
```

```
  filter(Delay=='Month', is.na(Total_Recall))
```

```
## # A tibble: 13 x 65
```

```
## # Groups:   Subject_ID [13]
```

```
##   Subject_ID Administration Condition Decision
```

```
Retrieval_Pract... TI      Delay
```

```
##   <chr>      <chr>      <chr>      <fct>      <fct>
```

```
<chr> <chr>
```

```
##   1 8          In_person      Decision    Deontolog... With RP
```

```
Delay... Month
```

```
##   2 36          In_person      No decisi... No decisi... With RP
```

```
Delay... Month
```

```
##   3 65          In_person      No decisi... No decisi... With RP
```

```
Delay... Month
```

```

## 4 75      In_person      Decision      Utilitari... With RP
Delay... Month

## 5 101     In_person      No decisi... No decisi... With RP
Delay... Month

## 6 103     In_person      Decision      Deontolog... With RP
Delay... Month

## 7 106     In_person      Decision      Deontolog... With RP
Delay... Month

## 8 107     In_person      No decisi... No decisi... With RP
Delay... Month

## 9 110     In_person      No decisi... No decisi... Without RP
Delay... Month

## 10 111    In_person      Decision      Deontolog... Without RP
Delay... Month

## 11 128    In_person      Decision      Deontolog... Without RP
Delay... Month

## 12 130    In_person      No decisi... No decisi... Without RP
Delay... Month

## 13 131    In_person      Decision      Deontolog... Without RP
Delay... Month

## # ... with 58 more variables: DSF <dbl>, DSB <dbl>, Total_Recall
<dbl>,

## #   Forgetting <chr>, Item_1 <chr>, Item_2 <chr>, Item_3 <chr>,

```

Item\_4 <chr>,  
## # Item\_5 <chr>, Item\_6 <chr>, Item\_7 <chr>, Item\_8 <chr>,  
Item\_9 <chr>,  
## # Item\_10 <chr>, Item\_11 <chr>, Item\_12 <chr>, Item\_13 <chr>,  
Item\_14 <chr>,  
## # Item\_15 <chr>, Item\_16 <chr>, Item\_17 <chr>, Item\_18 <chr>,  
Item\_19 <chr>,  
## # Item\_20 <chr>, Item\_21 <chr>, Item\_22 <chr>, Item\_23 <chr>,  
Item\_24 <chr>,  
## # Item\_25 <chr>, Item\_26 <chr>, Item\_27 <chr>,  
Recall\_Order\_Item\_1 <chr>,  
## # Recall\_Order\_Item\_2 <chr>, Recall\_Order\_Item\_3 <chr>,  
## # Recall\_Order\_Item\_4 <chr>, Recall\_Order\_Item\_5 <chr>,  
## # Recall\_Order\_Item\_6 <chr>, Recall\_Order\_Item\_7 <chr>,  
## # Recall\_Order\_Item\_8 <chr>, Recall\_Order\_Item\_9 <chr>,  
## # Recall\_Order\_Item\_10 <chr>, Recall\_Order\_Item\_11 <chr>,  
## # Recall\_Order\_Item\_12 <chr>, Recall\_Order\_Item\_13 <chr>,  
## # Recall\_Order\_Item\_14 <chr>, Recall\_Order\_Item\_15 <chr>,  
## # Recall\_Order\_Item\_16 <chr>, Recall\_Order\_Item\_17 <chr>,  
## # Recall\_Order\_Item\_18 <chr>, Recall\_Order\_Item\_19 <chr>,  
## # Recall\_Order\_Item\_20 <chr>, Recall\_Order\_Item\_21 <chr>,  
## # Recall\_Order\_Item\_22 <chr>, Recall\_Order\_Item\_23 <chr>,

```
## # Recall_Order_Item_24 <chr>, Recall_Order_Item_25 <chr>,
## # Recall_Order_Item_26 <chr>, Recall_Order_Item_27 <chr>
Experiment6<-Experiment6[Experiment6$Subject_ID != 8 &
Experiment6$Subject_ID != 36 & Experiment6$Subject_ID != 65 &
Experiment6$Subject_ID != 75 & Experiment6$Subject_ID != 101 &
Experiment6$Subject_ID != 103 & Experiment6$Subject_ID != 106 &
Experiment6$Subject_ID != 107 & Experiment6$Subject_ID != 110 &
Experiment6$Subject_ID != 111 & Experiment6$Subject_ID != 128 &
Experiment6$Subject_ID != 130 & Experiment6$Subject_ID != 131, ]
Post-decision recall
E6_data_ctrl<-subset(Experiment6, Delay == 'Immediate', select =
c('Subject_ID', 'Decision', 'Total_Recall', 'Retrieval_Practice',
'DSF', 'DSB'))
```

### Descriptive statistics

```
E6_post_recall_descriptives<-E6_data_ctrl %>%
  group_by(Retrieval_Practice, Decision) %>%
  summarise( N = n(),
             M = mean(Total_Recall, na.rm= TRUE),
             SD = sd(Total_Recall, na.rm= TRUE),
             Min = min(Total_Recall, na.rm= TRUE),
             Median = median(Total_Recall,na.rm= TRUE),
             Max = max(Total_Recall, na.rm= TRUE))
```

## `summarise()` has grouped output by 'Retrieval\_Practice'. You can override using the `.groups` argument.

```
E6_post_recall_descriptives<-
```

```
APA.descr.table(E6_post_recall_descriptives)%>%
```

```
  cols_label(Decision='Condition')
```

```
E6_post_recall_descriptives
```

**Inferential statistics**

```
E6_post_recall_model<- lm(Total_Recall ~ Retrieval_Practice *
```

```
Decision, data= E6_data_ctrl)
```

*Assumptions checking*

```
dwt(E6_post_recall_model)
```

```
## lag Autocorrelation D-W Statistic p-value
```

```
## 1 0.001483586 1.993892 0.968
```

```
## Alternative hypothesis: rho != 0
```

```
leveneTest(E6_post_recall_model)
```

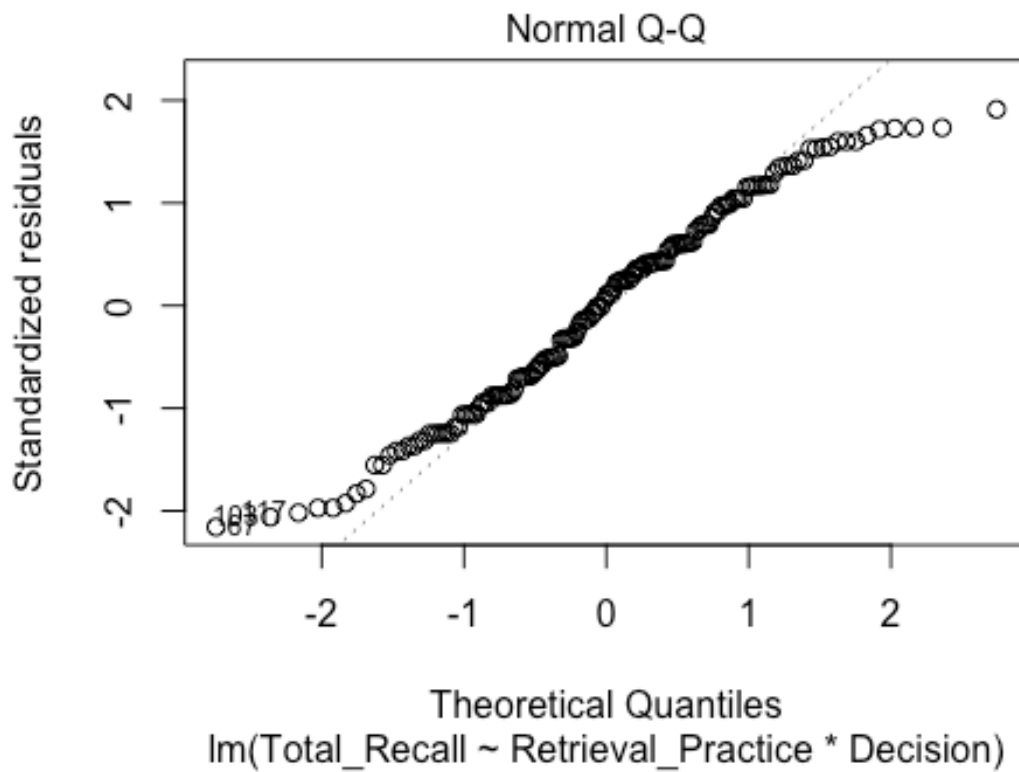
```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
## Df F value Pr(>F)
```

```
## group 5 0.9228 0.4678
```

```
## 158
```

```
plot(E6_post_recall_model, which=2)
```



```
E6_post_recall_SW<-shapiro.test(residuals(E6_post_recall_model))
```

```
E6_post_recall_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: residuals(E6_post_recall_model)
```

```
## W = 0.97629, p-value = 0.006453
```

*Bootstrapped model's results*

*# The boot function does not work if the dataset from which the model is based has missing data.*

*# There one missing data in Total Recall but two in DSF and DSB, so I am creating a new subset without DSF and DSB for the bootstrap otherwise it would not run.*

```
E6_post_recall_data<-subset(Experiment6, Delay == 'Immediate',
select =
c('Subject_ID','Decision','Retrieval_Practice','Total_Recall'))
E6_post_recall_data<- na.omit(E6_post_recall_data)
E6_post_recall_model_boot<- lm(Total_Recall ~ Retrieval_Practice *
Decision, data= E6_post_recall_data)
E6_post_recall_model_boot<- Boot(E6_post_recall_model_boot, R=10000)
Confint(E6_post_recall_model_boot, type= 'perc')%>%
  kable(digits= 3, caption= 'Bootstrapped regression model of post-
decision recall with 95%CIs')%>%
  kable_styling(full_width= FALSE)
```

Bootstrapped regression model of post-decision recall with 95%CIs

Estimate

2.5 %

97.5 %

(Intercept)

13.704

11.600

15.893

Retrieval\_PracticeWith RP

-1.324

-4.423

1.680

DecisionDeontological

-0.624

-3.565

2.210

DecisionUtilitarian

0.106

-3.072

3.170

Retrieval\_PracticeWith RP:DecisionDeontological

0.890

-3.307

5.078

Retrieval\_PracticeWith RP:DecisionUtilitarian

0.257

-3.826

4.375

*Parametric regression model's results* Provided for information only, not reported in the experiment manuscript.

```
model_parameters(E6_post_recall_model) %>% print_md()
```

Parameter	Coefficient	SE	95% CI	t(158)	p
(Intercept)	13.70	1.06	(11.62, 15.79)	12.98	< .001
Retrieval Practice (With RP)	-1.32	1.47	(-4.22, 1.57)	-0.90	0.368
Decision (Deontological)	-0.62	1.52	(-3.63, 2.38)	-0.41	0.683
Decision (Utilitarian)	0.11	1.60	(-3.05, 3.26)	0.07	0.947
Retrieval Practice (With RP) *	0.89	2.08	(-3.22, 5.00)	0.43	0.670
Decision (Deontological)					
Retrieval Practice (With RP) *	0.26	2.13	(-3.96, 4.47)	0.12	0.904
Decision (Utilitarian)					

Digit Span-Forward

### Descriptive statistics

```
E6_DSF_descriptives<-E6_data_ctrl %>%
  group_by(Retrieval_Practice, Decision) %>%
  summarise( N = n(),
             M = mean(DSF, na.rm= TRUE),
             SD = sd(DSF, na.rm= TRUE),
             Min = min(DSF, na.rm= TRUE),
             Median = median(DSF, na.rm= TRUE),
             Max = max(DSF, na.rm= TRUE))
```

```
## `summarise()` has grouped output by 'Retrieval_Practice'. You can  
override using the `.groups` argument.
```

```
E6_DSF_descriptives<-APA.descr.table(E6_DSF_descriptives)
```

```
E6_DSF_descriptives
```

```
Inferential statistics
```

```
E6_DSF<- lm(DSF ~ Retrieval_Practice * Decision, data= E6_data_ctrl)
```

```
Assumptions checking
```

```
leveneTest(E6_DSF)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##           Df F value Pr(>F)
```

```
## group     5  0.8859 0.4921
```

```
##           158
```

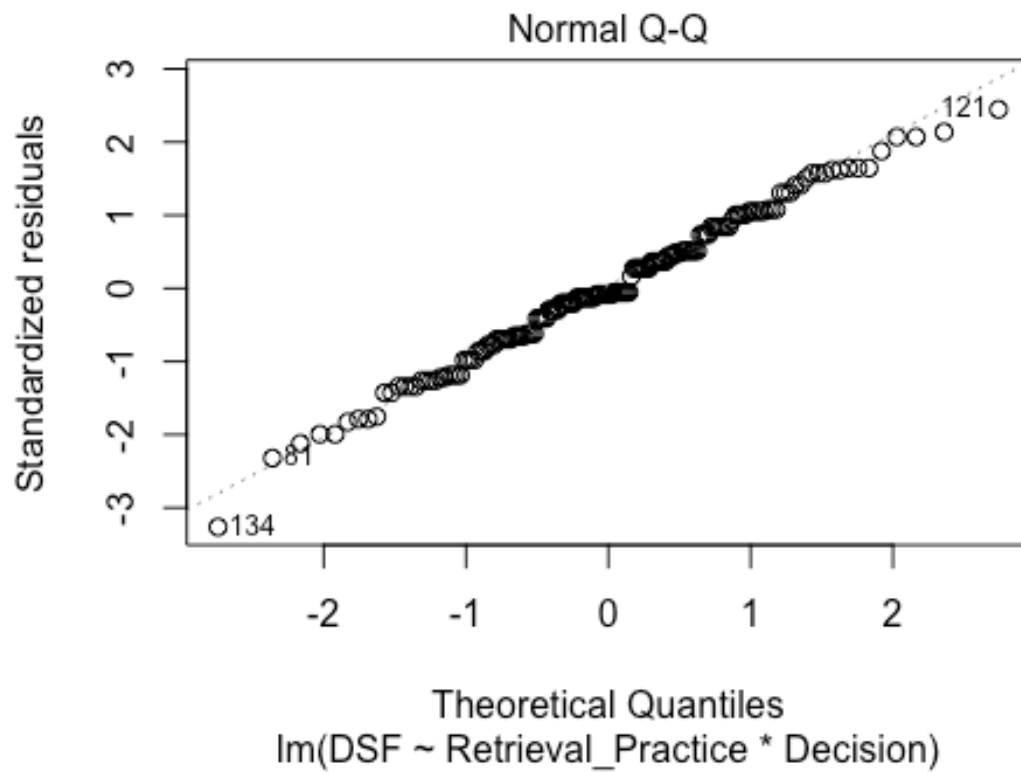
```
dwt(E6_DSF)
```

```
## lag Autocorrelation D-W Statistic p-value
```

```
## 1 -0.1547873 2.306367 0.056
```

```
## Alternative hypothesis: rho != 0
```

```
plot(E6_DSF, which=2)
```



```
E6_DSF_SW<-shapiro.test(residuals(E6_DSF))
```

```
E6_DSF_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: residuals(E6_DSF)
```

```
## W = 0.99225, p-value = 0.5247
```

*Regression model*

```
model_parameters(E6_DSF) %>% print_md()
```

Parameter	Coefficient	SE	95% CI	t(158)	p
(Intercept)	10.15	0.35	(9.47, 10.83)	29.38	<.001
Retrieval Practice (With RP)	-0.63	0.48	(-1.58, 0.32)	-1.31	0.191
Decision (Deontological)	0.21	0.50	(-0.77, 1.20)	0.43	0.671
Decision (Utilitarian)	-0.43	0.52	(-1.47, 0.60)	-0.83	0.407
Retrieval Practice (With RP) *	0.37	0.68	(-0.98, 1.71)	0.54	0.590
Decision (Deontological)					
Retrieval Practice (With RP) *	1.14	0.70	(-0.24, 2.52)	1.64	0.104
Decision (Utilitarian)					

Digit Span-Backward

### Descriptive statistics

```
E6_DSB_descriptives<-E6_data_ctrl %>%
```

```
  group_by(Retrieval_Practice, Decision) %>%
```

```
  summarise( N = n(),
```

```
             M = mean(DSB, na.rm= TRUE),
```

```
             SD = sd(DSB, na.rm= TRUE),
```

```
             Min = min(DSB, na.rm= TRUE),
```

```
             Median = median(DSB, na.rm= TRUE),
```

```
             Max = max(DSB, na.rm= TRUE))
```

## `summarise()` has grouped output by 'Retrieval\_Practice'. You can override using the `.groups` argument.

```
E6_DSB_descriptives<-APA.descr.table(E6_DSB_descriptives)%>%
```

```
  cols_label(Decision='Condition')
```

```
E6_DSB_descriptives
```

**Inferential statistics**

```
E6_DSB<- lm(DSB ~ Retrieval_Practice * Decision, data= E6_data_ctrl)
```

*Assumptions checking*

```
leveneTest(E6_DSB)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group  5  1.2406 0.2927
```

```
##      158
```

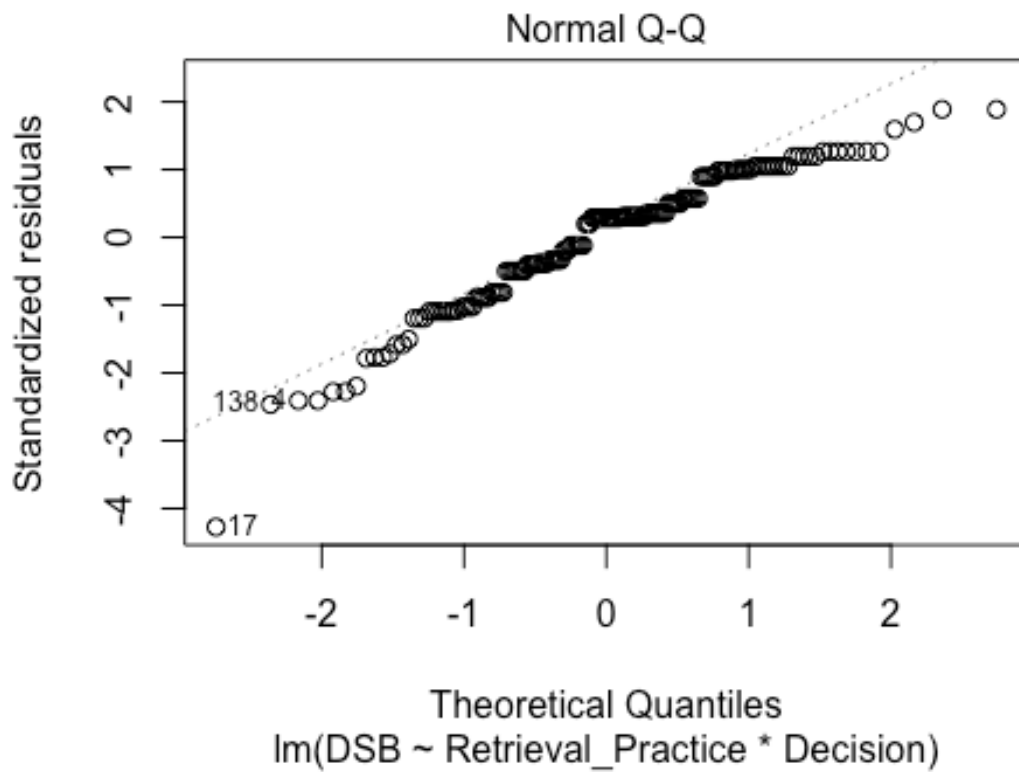
```
dwt(E6_DSB)
```

```
## lag Autocorrelation D-W Statistic p-value
```

```
##  1      0.02193735      1.947289      0.704
```

```
## Alternative hypothesis: rho != 0
```

```
plot(E6_DSB, which=2)
```



```
shapiro.test(residuals(E6_DSB))
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: residuals(E6_DSB)
```

```
## W = 0.9466, p-value = 7.28e-06
```

*Bootstrapped model's results*

*# The boot function does not work if the dataset from which the model is based has missing data.*

*# There one missing data in Total Recall and two in DSF and DSB, so*

*I am creating a new subset without missing data the bootstrap.*

```
E6_data_DSB<-subset(Experiment6, Delay == 'Immediate', select =
c('Subject_ID', 'Decision', 'Retrieval_Practice', 'DSB'))
E6_data_DSB<- na.omit(E6_data_DSB)
E6_DSB_model_boot<- lm(DSB ~ Retrieval_Practice * Decision, data=
E6_data_DSB)
E6_DSB_model_boot<- Boot(E6_DSB_model_boot, R=10000)
Confint(E6_DSB_model_boot, type= 'perc')%>%
  kable(digits= 3, caption= 'Bootstrapped regression model of DSB
with 95%CIs')%>%
  kable_styling(full_width= FALSE)
```

Bootstrapped regression model of DSB with 95%CIs

Estimate

2.5 %

97.5 %

(Intercept)

9.556

9.000

10.053

Retrieval\_PracticeWith RP

-0.383

-1.227

0.426

DecisionDeontological

-0.276

-1.113

0.589

DecisionUtilitarian

0.159

-0.553

0.901

Retrieval\_PracticeWith RP:DecisionDeontological

0.587

-0.581

1.771

Retrieval\_PracticeWith RP:DecisionUtilitarian

0.250

-0.795

1.328

For information, the results for the regression model on t distribution assumption are provided below.

```
model_parameters(E6_DSB) %>% print_md()
```

Parameter	Coefficient	SE	95% CI	t(158)	p
(Intercept)	9.56	0.28	(9.00, 10.11)	33.78	< .001

Parameter	Coefficient	SE	95% CI	t(158)	p
Retrieval Practice (With RP)	-0.38	0.39	(-1.16, 0.39)	-0.97	0.331
Decision (Deontological)	-0.28	0.41	(-1.08, 0.53)	-0.68	0.500
Decision (Utilitarian)	0.16	0.43	(-0.69, 1.00)	0.37	0.711
Retrieval Practice (With RP) *	0.59	0.56	(-0.51, 1.69)	1.05	0.294
Decision (Deontological)					
Retrieval Practice (With RP) *	0.25	0.57	(-0.88, 1.38)	0.44	0.663
Decision (Utilitarian)					

Delayed recall

```
E6_data_delay<-subset(Experiment6, Delay == 'Month', select =
c('Subject_ID', 'Decision', 'Total_Recall', 'Retrieval_Practice'))
```

**Descriptive statistics**

```
E6_delay_descriptives<-E6_data_delay %>%
  group_by(Retrieval_Practice, Decision) %>%
  summarise( N = n(),
             M = mean(Total_Recall, na.rm= TRUE),
             SD = sd(Total_Recall, na.rm= TRUE),
             Min = min(Total_Recall, na.rm= TRUE),
             Median = median(Total_Recall, na.rm= TRUE),
             Max = max(Total_Recall, na.rm= TRUE))
```

## `summarise()` has grouped output by 'Retrieval\_Practice'. You can override using the `.groups` argument.

```
E6_delay_descriptives<-APA.descr.table(E6_delay_descriptives)%>%
```

```
  cols_label(Decision='Condition')
```

```
E6_delay_descriptives
```

**Inferential statistics**

```
contrasts(E6_data_delay$Decision)<-cbind('Any Decision - No
```

```
Decision'=c(-1,0.5,0.5),
```

```
      'Utilitarian - Deontological'=c(0,-1,+1))
```

```
E6_delay_model<- lm(Total_Recall ~ Retrieval_Practice * Decision,
```

```
data= E6_data_delay)
```

*Assumptions checking*

```
leveneTest(E6_delay_model)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##           Df F value Pr(>F)
```

```
## group     5  3.0784 0.01112 *
```

```
##           159
```

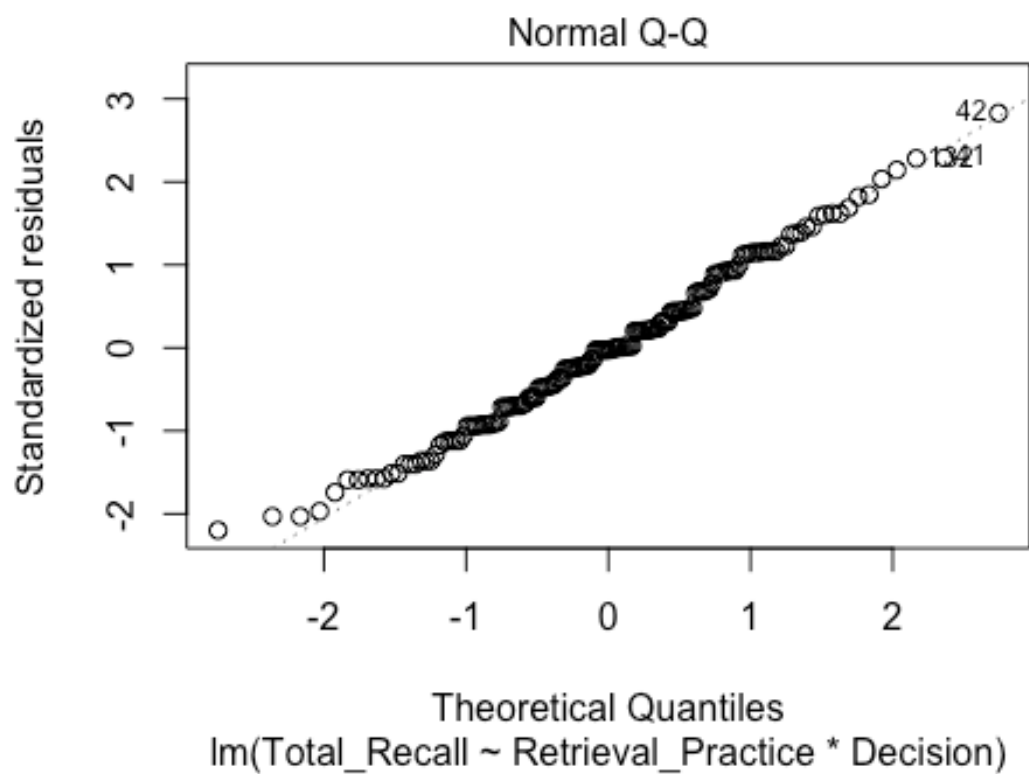
```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
E6_delay_dwt<-dwt(E6_delay_model)
```

```
E6_delay_dwt
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.09303649 1.79963 0.17
## Alternative hypothesis: rho != 0
plot(E6_delay_model, which=2)
```



```
E6_delay_SW<-shapiro.test(residuals(E6_delay_model))
```

```
E6_delay_SW
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: residuals(E6_delay_model)
```

```
## W = 0.99105, p-value = 0.3905
```

*Bootstrapped model's results*

```
E6_delay_model_boot<- Boot(E6_delay_model, R=10000)
```

```
Confint(E6_delay_model_boot, type= 'perc')%>%
```

```
  kable(digits= 3, caption= 'Bootstrapped regression model of total  
recall with 95%CIs')%>%
```

```
  kable_styling(full_width= FALSE)
```

Bootstrapped regression model of total recall with 95%CIs

Estimate

2.5 %

97.5 %

(Intercept)

4.749

3.913

5.610

Retrieval\_PracticeWith RP

4.416

3.087

5.733

DecisionAny Decision - No Decision

0.638

-0.436

1.723

DecisionUtilitarian - Deontological

1.028

-0.076

2.150

Retrieval\_PracticeWith RP:DecisionAny Decision - No Decision

-1.093

-3.024

0.731

Retrieval\_PracticeWith RP:DecisionUtilitarian - Deontological

-0.997

-2.628

0.589

Parametric regression model's results for information

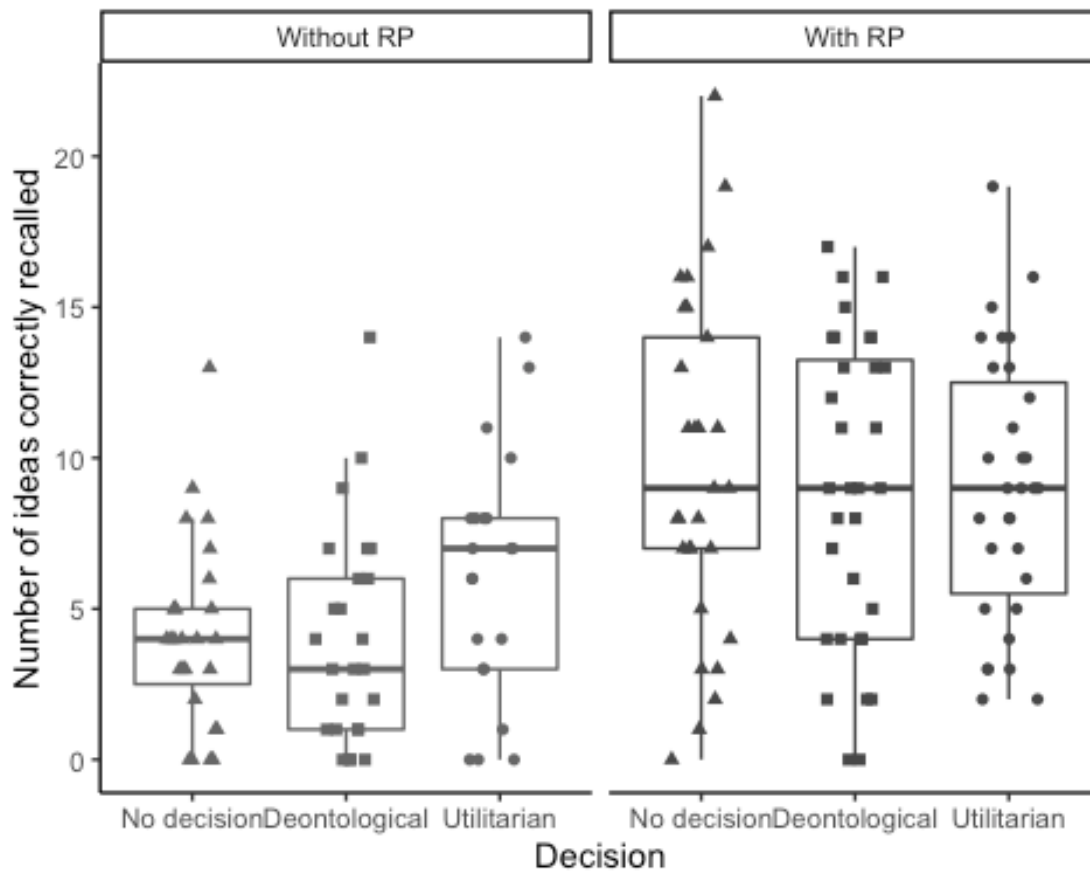
`model_parameters(E6_delay_model) %>% print_md()`

Parameter	Coefficient	SE	95% CI	t(159)	p
(Intercept)	4.75	0.52	(3.71, 5.78)	9.06	< .001
Retrieval Practice (With RP)	4.42	0.70	(3.03, 5.80)	6.30	< .001
Decision (Any Decision - No Decision)	0.64	0.72	(-0.79, 2.06)	0.88	0.378

Parameter	Coefficient	SE	95% CI	t(159)	p
Decision (Utilitarian - Deontological)	1.03	0.66	(-0.27, 2.33)	1.56	0.121
Retrieval Practice (With RP) *	-1.09	0.98	(-3.03, 0.85)	-1.11	0.267
Decision (Any Decision - No Decision)					
Retrieval Practice (With RP) *	-1.00	0.87	(-2.71, 0.71)	-1.15	0.251
Decision (Utilitarian - Deontological)					

```
E6_Figure<-ggplot(E6_data_delay, aes(x=Decision, y=Total_Recall,
shape=Decision, color=Retrieval_Practice, )) +
  labs(y= 'Number of ideas correctly
recalled') +
  facet_grid(. ~ Retrieval_Practice) +
  geom_boxplot(outlier.shape = NA, show.legend
= FALSE) +
  geom_jitter(width = 0.2, height = 0 ,
show.legend = FALSE) +
  scale_shape_manual(values = c(17,15,16))+
  scale_color_manual(values =
c('grey40', 'grey28', 'grey16')) +
  theme_classic()
```

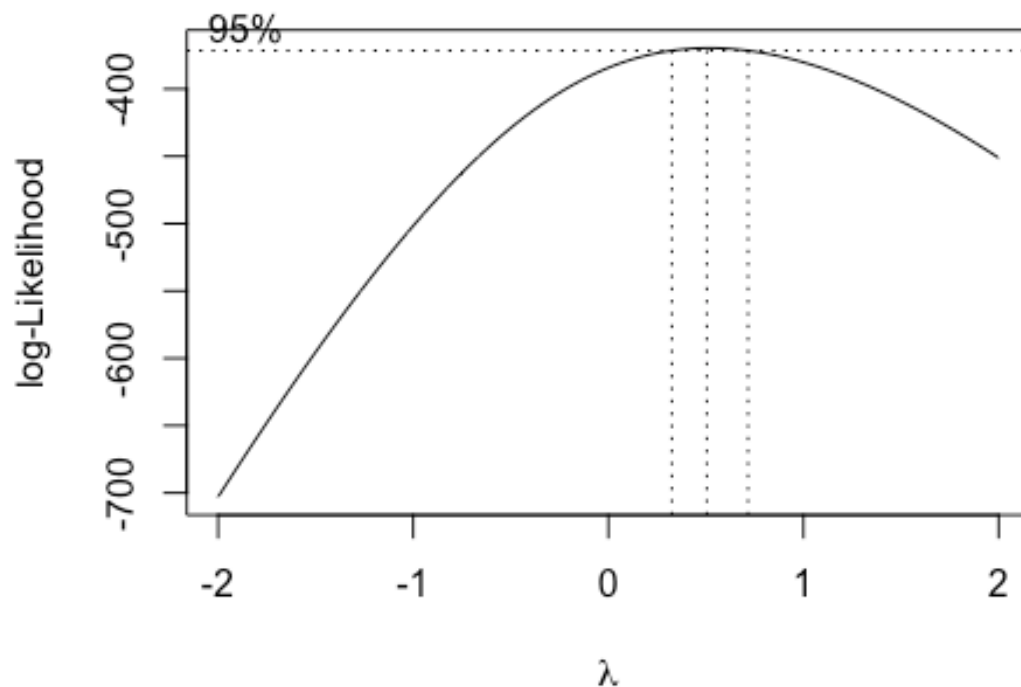
E6\_Figure



AIC & BIC

```
E6_data_delay$Total_Recall_tr <- E6_data_delay$Total_Recall + 1
```

```
E6_boxcox <- boxcox(lm(Total_Recall_tr ~ 1, data = E6_data_delay))
```



```
E6_boxcox
```

```
## $x
```

```
## [1] -2.00000000 -1.95959596 -1.91919192 -1.87878788 -1.83838384
```

```
-1.79797980
```

```
## [7] -1.75757576 -1.71717172 -1.67676768 -1.63636364 -1.59595960
```

```
-1.55555556
```

```
## [13] -1.51515152 -1.47474747 -1.43434343 -1.39393939 -1.35353535
```

```
-1.31313131
```

```
## [19] -1.27272727 -1.23232323 -1.19191919 -1.15151515 -1.11111111
```

```
-1.07070707
```

## [25] -1.03030303 -0.98989899 -0.94949495 -0.90909091 -0.86868687  
-0.82828283

## [31] -0.78787879 -0.74747475 -0.70707071 -0.66666667 -0.62626263  
-0.58585859

## [37] -0.54545455 -0.50505051 -0.46464646 -0.42424242 -0.38383838  
-0.34343434

## [43] -0.30303030 -0.26262626 -0.22222222 -0.18181818 -0.14141414  
-0.10101010

## [49] -0.06060606 -0.02020202 0.02020202 0.06060606 0.10101010  
0.14141414

## [55] 0.18181818 0.22222222 0.26262626 0.30303030 0.34343434  
0.38383838

## [61] 0.42424242 0.46464646 0.50505051 0.54545455 0.58585859  
0.62626263

## [67] 0.66666667 0.70707071 0.74747475 0.78787879 0.82828283  
0.86868687

## [73] 0.90909091 0.94949495 0.98989899 1.03030303 1.07070707  
1.11111111

## [79] 1.15151515 1.19191919 1.23232323 1.27272727 1.31313131  
1.35353535

## [85] 1.39393939 1.43434343 1.47474747 1.51515152 1.55555556  
1.59595960

## [91] 1.63636364 1.67676768 1.71717172 1.75757576 1.79797980

1.83838384

## [97] 1.87878788 1.91919192 1.95959596 2.00000000

##

## \$y

## [1] -703.0634 -693.9738 -684.9447 -675.9784 -667.0770 -658.2428

-649.4780

## [8] -640.7851 -632.1666 -623.6250 -615.1632 -606.7840 -598.4902

-590.2850

## [15] -582.1714 -574.1529 -566.2326 -558.4142 -550.7012 -543.0974

-535.6065

## [22] -528.2324 -520.9792 -513.8510 -506.8518 -499.9859 -493.2575

-486.6711

## [29] -480.2309 -473.9412 -467.8064 -461.8307 -456.0183 -450.3734

-444.9001

## [36] -439.6020 -434.4832 -429.5468 -424.7965 -420.2351 -415.8655

-411.6902

## [43] -407.7113 -403.9307 -400.3500 -396.9702 -393.7921 -390.8160

-388.0422

## [50] -385.4700 -383.0987 -380.9274 -378.9545 -377.1782 -375.5964

-374.2067

## [57] -373.0065 -371.9928 -371.1626 -370.5123 -370.0387 -369.7380

-369.6065

## [64] -369.6404 -369.8358 -370.1886 -370.6951 -371.3511 -372.1528

```
-373.0963
```

```
## [71] -374.1775 -375.3929 -376.7384 -378.2106 -379.8058 -381.5205
```

```
-383.3513
```

```
## [78] -385.2949 -387.3480 -389.5076 -391.7706 -394.1342 -396.5954
```

```
-399.1517
```

```
## [85] -401.8004 -404.5389 -407.3650 -410.2762 -413.2703 -416.3451
```

```
-419.4987
```

```
## [92] -422.7290 -426.0341 -429.4122 -432.8616 -436.3804 -439.9672
```

```
-443.6204
```

```
## [99] -447.3384 -451.1197
```

```
E6_lambda <- E6_boxcox$x[which.max(E6_boxcox$y)]
```

```
E6_lambda # give us the lambda value required to normalise
```

```
Total_Recall
```

```
## [1] 0.5050505
```

```
# corresponds to a square root function
```

```
E6_data_delay$Total_Recall_norm<-
```

```
E6_data_delay$Total_Recall^E6_lambda
```

```
E6_model_inter<- lm(Total_Recall_norm ~ Retrieval_Practice *
```

```
Decision, data= E6_data_delay)
```

```
E6_model_decision<- lm(Total_Recall_norm ~ Retrieval_Practice +
```

```
Decision, data= E6_data_delay)
```

```
E6_model_null<- lm(Total_Recall_norm ~ Retrieval_Practice, data=
```

```
E6_data_delay)
```

```
AIC(E6_model_inter, E6_model_decision)
```

```
##           df      AIC
```

```
## E6_model_inter    7 484.1833
```

```
## E6_model_decision  5 481.6816
```

```
BIC(E6_model_inter, E6_model_decision)
```

```
##           df      BIC
```

```
## E6_model_inter    7 505.9249
```

```
## E6_model_decision  5 497.2113
```

```
AIC(E6_model_null, E6_model_decision)
```

```
##           df      AIC
```

```
## E6_model_null     3 479.5684
```

```
## E6_model_decision  5 481.6816
```

```
BIC(E6_model_null, E6_model_decision)
```

```
##           df      BIC
```

```
## E6_model_null     3 488.8862
```

```
## E6_model_decision  5 497.2113
```

Complementary analyses

Omnibus comparisons

```
E6_data_decision<-E6_data_delay
```

```
E6_data_decision$Condition<-E6_data_decision$Decision
```

```
E6_data_decision$Condition<-as.character(E6_data_decision$Condition)
```

```
E6_data_decision$Condition[E6_data_decision$Condition ==
```

```
'Deontological']<- 'Decision'
```

```

E6_data_decision$Condition[E6_data_decision$Condition ==
'Utilitarian']<- 'Decision'
E6_data_decision$Condition<- factor(E6_data_decision$Condition,
levels =c('No decision', 'Decision'))
E6_Omnibus_any_decision<-glm(Total_Recall ~
Condition*Retrieval_Practice, data=E6_data_decision, family =
poisson)
summary(E6_Omnibus_any_decision)
##
## Call:
## glm(formula = Total_Recall ~ Condition * Retrieval_Practice,
##      family = poisson, data = E6_data_decision)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.386  -1.520  -0.055   1.244   3.486
##
## Coefficients:
##
##                                     Estimate Std. Error z
value
## (Intercept)                       1.41369    0.09492
14.894
## ConditionDecision                   0.19139    0.11565

```

1.655

```
## Retrieval_PracticeWith RP          0.85022    0.11222
```

7.576

```
## ConditionDecision:Retrieval_PracticeWith RP -0.26516    0.13688
```

-1.937

```
##                                     Pr(>|z|)
```

```
## (Intercept)                        < 2e-16 ***
```

```
## ConditionDecision                   0.0980 .
```

```
## Retrieval_PracticeWith RP          3.55e-14 ***
```

```
## ConditionDecision:Retrieval_PracticeWith RP  0.0527 .
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
```

```
## Null deviance: 638.20 on 164 degrees of freedom
```

```
## Residual deviance: 514.62 on 161 degrees of freedom
```

```
## AIC: 1082.7
```

```
##
```

```
## Number of Fisher Scoring iterations: 5
```

```
E6_data_decision<-subset(E6_data_delay, , select = c('Subject_ID',
'Decision', 'Total_Recall', 'Retrieval_Practice'))
```

```
E6_data_decision<-subset(E6_data_delay,Decision!= "No decision")
```

```
E6_data_decision$Decision<-factor(E6_data_decision$Decision, levels=
c('Deontological', 'Utilitarian'))
```

```
E6_data_decision %>%
```

```
  group_by(Decision, Retrieval_Practice) %>%
```

```
  summarise( N = n(), M = mean(Total_Recall), SD =
```

```
sd(Total_Recall),
```

```
    Min = min(Total_Recall),
```

```
    Median = median(Total_Recall),
```

```
    Max = max(Total_Recall))
```

```
## `summarise()` has grouped output by 'Decision'. You can override
using the `.groups` argument.
```

```
## # A tibble: 4 x 8
```

```
## # Groups:   Decision [2]
```

```
##   Decision      Retrieval_Practice    N     M    SD   Min Median
```

```
Max
```

```
##   <fct>          <fct>          <int> <dbl> <dbl> <dbl> <dbl>
```

```
<dbl>
```

```
## 1 Deontological Without RP          25  4.04  3.59    0     3
```

```
14
```

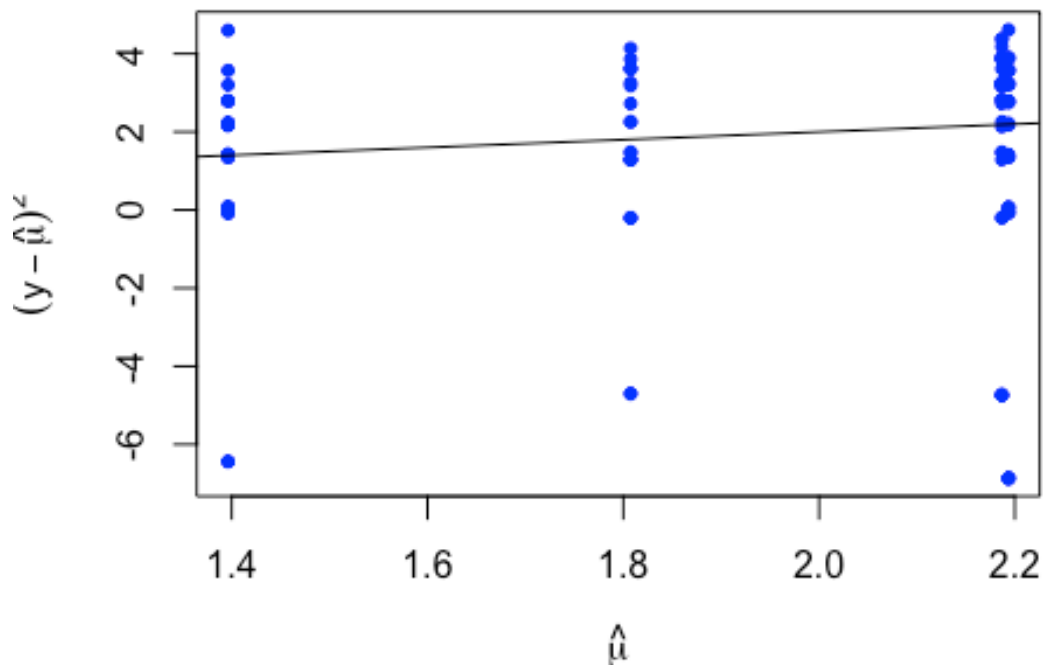
```
## 2 Deontological With RP             32  8.91  5.09    0     9
```

```
17
```

```
## 3 Utilitarian   Without RP          21  6.10  4.07    0     7
```

```
14
```

```
## 4 Utilitarian With RP          31  8.97  4.44    2    9
19
E6_Omnibus_decision<-glm(Total_Recall ~ Decision*Retrieval_Practice,
data=E6_data_decision, family = quasipoisson)
E6_overdis<-
plot(log(fitted(E6_Omnibus_decision)),log((E6_data_decision$Total_Re
call-
fitted(E6_Omnibus_decision))^2),xlab=expression(hat(mu)),ylab=expres
sion((y-hat(mu))^2),pch=20,col="blue")
abline(0,1)
```



```
summary(E6_Omnibus_decision)

##

## Call:
## glm(formula = Total_Recall ~ Decision * Retrieval_Practice,
family = quasipoisson,
## data = E6_data_decision)
##

## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2205  -1.4469   0.0108   1.2608   3.8573

##

## Coefficients:
##                                     Estimate Std. Error
t value
## (Intercept)                       1.3962    0.1645
8.488
## DecisionUtilitarian                 0.4113    0.2200
1.869
## Retrieval_PracticeWith RP          0.7905    0.1914
4.129
## DecisionUtilitarian:Retrieval_PracticeWith RP -0.4044    0.2604
-1.553

##                                     Pr(>|t|)
```

```

## (Intercept) 1.47e-13 ***
## DecisionUtilitarian 0.0644 .
## Retrieval_PracticeWith RP 7.32e-05 ***
## DecisionUtilitarian:Retrieval_PracticeWith RP 0.1235
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be
2.732781)
##
## Null deviance: 403.19 on 108 degrees of freedom
## Residual deviance: 333.73 on 105 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
E6_data_decision$Total_Recall_norm<-
E6_data_decision$Total_Recall^E6_lambda
summary(lm(Total_Recall_norm ~ Retrieval_Practice*Decision, data =
E6_data_decision))
##
## Call:
## lm(formula = Total_Recall_norm ~ Retrieval_Practice * Decision,
## data = E6_data_decision)

```



##

## Residual standard error: 1.025 on 105 degrees of freedom

## Multiple R-squared: 0.1822, Adjusted R-squared: 0.1588

## F-statistic: 7.796 on 3 and 105 DF, p-value: 9.502e-05

## Appendix 2

### *Experiment 2, Dilemma Story and Scoring Form*

Idea n°	Text	Required
1	In 2016,	*2016 OR *two years ago
2	Lemannis	*Lemannis
3	bacteria were detected	*bacteria
4	in the lake	*lake
5	of Geneva, Switzerland	*Geneva OR *Switzerland
6	These bacteria cause infections in body organs	*organs
7	leading to death.	indication that the bacteria are lethal
8	The government wants to launch	*government
9	a screening campaign	indication that an action will be taken to screen people
10	to detect the infection	indication that the target of the screening is the infection
11	when it is still curable.	indication that the infection is curable at some point
12	However, the risk	indication of a probability
13	of false-positive results	words related to an inaccurate diagnosis
14	is 1 over 3	statistics equal to 1:3
15	and the cure is fatal	indication that the cure is lethal
16	for uninfected people.	Indication of absence of sickness
17	After a pilot study	Indication of primary investigation
18	100 people	*100
19	were screened positive,	Indication of positive result
20	70 people were	*70
21	true-positive	indication of an accurate diagnosis
22	and recovered after receiving the cure,	indication of remission
23	30 people	*30
24	were false-positive	indication of an inaccurate diagnosis
25	and died after receiving the cure.	indication of death

## Appendix 3

*Percentage of Utilitarian Moral Decision Previously Obtained in Moore et al. (2008)*

*and Harrison et al. (2008) for The Pre-Selected Dilemmas*

Dilemma	Authors	Percentage of utilitarian
---------	---------	---------------------------

---

The Bike week	Moore et al., 2008	51%
The Nobel Prize	Moore et al., 2008	54%
The Vaccine	Moore et al., 2008	58%
The Best Friend	Harrison et al., 2008	59%

---

## Appendix 4

*Lexical Characteristic of The Pre-Selected Dilemmas from Moore et al., (2008) and Harrison et al. (2008) Compared to The Story C of The WMS-IV.*

Lexical characteristics	Moore et al., 2008			Harrison et al., 2008	WMS-IV, 2010
	The Bike week	The Nobel Prize	The Vaccine	The Best Friend	Story C
Total sentences	5	6	6	5	5
Mean sentence length	24.6	19.16	19.33	17	17.2
Total words	123	115	116	85	86
Total syllables	156	153	155	101	124
Mean syllables per words	1.27	1.33	1.34	1.19	1.44
Flesch reading ease index*	74.42	74.87	73.85	66.91	67.55
Flesch-Kincaird index**	8.99	6	7.76	8.15	8.11

## Appendix 5

### *Moral Dilemmas Set*

The Bike Week, Original Version (Moore et al., 2008)

*“You are an expert motorcycle rider and you have gone on vacation in order to participate in Bike Week. Thousands of other motorcycle riders from across the country have come to ride in this event. As you are riding down the road in the front of*

*a large group of other riders, you see that someone up ahead is losing control of his bike. As you watch him fall, you know that you could easily ride around the fallen rider but the riders behind you will run over each other trying to avoid the crash. To prevent a large pile up you could jump from your bike, allowing the bike to slide into the fallen rider, but this would also certainly kill the fallen rider. “*

What are you going to do?	
<input type="checkbox"/> I jump from my bike	<input type="checkbox"/> I ride around the rider

#### Bike Week, Modified Version with Added Details

*“You are an expert motorcycle rider and you have gone on vacation in order to participate in Bike Week. Two thousand a hundred other motorcycle riders from across Eastern Europe have come to ride in this event. As you are riding down the road in the front of a large group of other riders, you see a young man 600 meters ahead is losing control of his yellow bike. As you watch him fall, you know that you could easily ride around the fallen rider but the riders behind you will run over each other trying to avoid the crash. To prevent a large pile up you could jump from your bike, allowing the bike to slide into the fallen rider, but this would also certainly kill the fallen rider.”*

What are you going to do?	
<input type="checkbox"/> I jump from my bike	<input type="checkbox"/> I ride around the rider

#### The Nobel Prize, Original Version (Moore et al., 2008)

*“You and a fellow researcher have discovered a powerful new energy source that is cheap, safe, and clean. You realize that this could lead to the elimination of pollution and poverty around the world. However, your colleague wants to sell this*

*discovery to the military, which will turn it into a powerful new weapon. You can prevent this by releasing a flammable gas in the lab, so that when he turns on his computer, it will cause an explosion. This will also cause the death of your colleague. Everyone will think that it was an unfortunate lab accident, and the discovery will not be sold to those who might create a weapon out of it.”*

What are you going to do?	
<input type="checkbox"/> I release the gas	<input type="checkbox"/> I let my colleague sell our discovery

#### The Nobel Prize, Modified Version with Added Details

*“You and a fellow researcher, Professor Joe Grant, have discovered a powerful new energy source that is cheap, safe, and clean. You realize that this could lead to the elimination of pollution and poverty around the world. However, your colleague wants to sell this discovery to the US army, which will turn it into a powerful new weapon. You can prevent this by releasing propane gas in the lab, so that when he turns on his computer, it will cause an explosion within two seconds. This will also cause the death of your colleague. Everyone will think that it was just a lab accident, and the discovery will not be sold to those who might create a weapon out of it.”*

What are you going to do?	
<input type="checkbox"/> I release the gas	<input type="checkbox"/> I let my colleague sell our discovery

#### The Vaccine, Original Version (Moore et al., 2008)

*“You are a medical researcher and you have developed two substances in your laboratory. You know that one of them is a vaccine, but you don’t know which because both of the vials have been mislabelled as vaccine. You also know that the other*

*substance is deadly. Once you figure out which substance is the vaccine you can create more to save thousands of lives. You could allow another doctor to inject two*

What are you going to do?	
<input type="checkbox"/> I administrate the deadly injection	<input type="checkbox"/> I let thousands of people die

*unsuspecting patients, who need the vaccine, with these substances without warning them or the doctor that one of the substances is deadly. One person will live, the other will die, and you will be able to start saving lives with your vaccine.”*

#### The Vaccine, Modified Version with Added Details

*“You are a medical researcher in immunology at the University of Berlin, and you have developed two substances in your laboratory. You know that one of them is a vaccine, but you don’t know which because both of the vials have been mislabelled as vaccine by the new intern. You also know that the other substance causes organ infections. Once you figure out which substance is the vaccine you can create more to save thousands of lives. You could allow your colleague, Dr Joe Grant, to inject two unsuspecting patients, who need the vaccine, with these substances without warning them or the doctor that one of the substances is deadly. One person will live, the other will die, and you will be able to start saving lives with your vaccine.”*

What are you going to do?	
<input type="checkbox"/> I administrate the deadly injection	<input type="checkbox"/> I let a thousand of people died

#### The Best Friend, Original Version (Harrison et al., 2008)

*“You have a childhood friend. This friend explains you that he has committed a crime and asks you to promise that you will never tell anybody. One day you discover that an innocent man has been accused of the crime and you ask your friend to confess.*

*Your friend refuses to do so. If you say nothing to the police, the innocent man will be imprisoned. If you speak, you will break your promise and your friend will go to jail.”*

What are you going to do?	
<input type="checkbox"/> I keep my promise	<input type="checkbox"/> I denounce my best friend

The Best Friend, Modified Version with Added Details

*“You have a childhood friend, Joe. This friend explains you that he has committed a burglary and asks you to promise that he will never tell anybody. One day you discover that a young, innocent man has been accused of the crime, and you ask your friend to confess the next morning. Your friend refuses to do so. If you say nothing to the police, the innocent man will be imprisoned. If you speak, you will break your promise and your friend will go to jail.”*

What are you going to do?	
<input type="checkbox"/> I keep my promise	<input type="checkbox"/> I denounce my best friend

## Appendix 6

### Scoring Form of Modified Nobel Prize Dilemma Recall

Idea n°	Text	Required
1	You and a fellow researcher,	indication of a co worker
2	<i>Professor</i>	* Professor
3	<i>Joe</i>	Joe OR variant of the name
4	<i>Grant,</i>	*Grant
5	have discovered	Indication of a discovery
6	a powerful	* powerful
7	new energy source	Indication of a source of energy
8	that is cheap, and safe, clean.	Give at least one out of the three adjectives
9	You realise that this could lead to the elimination	Indication of an eradication
10	of pollution	indication of pollution
11	and poverty	indication of low economic or social conditions
12	around the world.	Indication of global consequences
13	However, your colleague wants to sell this discovery	indication of an intention to sell
14	to the <i>US</i>	US, USA, American
15	<i>army,</i>	Indication of military
16	which will turn it into a powerful new weapon.	indication that it will be weaponised
17	You can prevent this	Indication of prevention
18	by releasing	Mentioning the action of realising gas
19	<i>propane</i> gas	*propane
20	in the lab,	Indication of a workspace
21	so that when he turns on his	Mentioning an activation (switch on, start, etc.)
22	computer,	*computer
23	it will cause an explosion	Indication of an explosion
24	<i>within two seconds.</i>	*two seconds
25	This will also cause the death of your colleague.	Indication of the colleague dying
26	Everyone will think that it was just a lab accident,	Indication of a presumed accident
27	and the discovery will not be sold to those who might create a weapon out of it.	Indication that the sell will be avoided



