Duration Judgments for Verbal Stimuli:
Effects of Emotion, Attention, and Memory Encoding

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Psychology

by

Laura Wendy Johnson

2014
In six experiments, this dissertation investigated duration judgments for verbal stimuli, testing predictions of information-processing models of time perception. Experiments 1, 2, and 3 explored the effects of low-valence, high-arousal taboo words on the perception of time. The results revealed that durations of taboo words were underestimated compared to neutral words in prospective timing tasks, including the temporal bisection task in Experiment 1 and the ordinality comparison procedure in Experiment 2. In both of these experiments, memory was better for taboo than neutral words in surprise free recall tasks. These results supported the predictions of the attentional gate model, which suggests that duration judgments are shorter when attention is directed away from time. In Experiments 1 and 2, emotion-linked memory encoding processes associated with taboo words may have been prioritized over the timing tasks, directing attention away from time and reducing duration judgments, as well as improving memory for taboo words. In Experiment 3, durations of taboo words were over-
estimated compared to neutral words in a retrospective duration reproduction task. This supported the predictions of the memory storage model, which suggests that retrospective duration judgments are longer for intervals in which greater quantities of non-temporal information are encoded. Experiments 4, 5, and 6 examined the effects of words displayed in novel combinations of fonts, sizes, and colors on the perception of time. Results from prospective temporal bisection tasks revealed that durations of words displayed in novel text styles were underestimated compared to words displayed in standard text styles. Similar results were found in Experiment 4, in which half of the words were displayed in novel text styles, and in Experiment 5, in which only one twelfth of the words were displayed in novel text styles. The similar effects of taboo words and novel text styles on duration judgments suggest that these effects are driven by attentional mechanisms. In Experiment 6, durations of novel shapes were judged to be longer than durations of novel words in the oddball task, providing tentative support for the hypothesis that words affect timing judgments differently than simple visual stimuli.
The dissertation of Laura Wendy Johnson is approved.

Dean Buonomano
Alan Castel
Aimee Drolet Rossi
Donald MacKay, Committee Chair

University of California, Los Angeles
2014
# TABLE OF CONTENTS

List of Tables ........................................................................... vii
List of Figures ........................................................................... viii
Acknowledgements ..................................................................... x
Vita ......................................................................................... xi
Chapter 1: Introduction and Overview ........................................ 1
  Methods Used in Time Research ........................................... 1
  Theoretical Models ............................................................... 3
  Emotion and Time ............................................................... 9
  Attention and Time ............................................................. 19
  Verbal Stimuli and Time ....................................................... 21
  Goals of the Present Studies ................................................. 22
Chapter 2: Duration Judgments for Taboo Words ....................... 25
  Experiment 1: Temporal Bisection Task ............................... 30
    Method ................................................................. 31
    Results ................................................................. 33
    Discussion ............................................................ 39
  Experiment 2: Ordinality Comparison Procedure .................. 44
    Method ................................................................. 45
    Results ................................................................. 47
    Discussion ............................................................ 49
  Experiment 3: Retrospective Duration Reproduction Task ....... 51
    Method ................................................................. 53
    Results ................................................................. 54
    Discussion ............................................................ 55
General Discussion .................................................................... 56
<table>
<thead>
<tr>
<th>Table Number</th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Summary of studies on emotion and time perception</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Characteristics of taboo and neutral words lists from Experiment 1</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Correlations between Experiment 1 timing and recall results and word ratings from Janschewitz (2008)</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>Characteristics of taboo and neutral words lists from Experiment 2</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>Correlations between Experiment 2 recall results and word ratings from Janschewitz (2008)</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>Correlations between Experiment 4 timing and recall results and word ratings from Janschewitz (2008)</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>Correlations between Experiment 5 timing and recall results and word ratings from Janschewitz (2008)</td>
<td>74</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1. Diagram of scalar expectancy theory .................................................. 5
2. Diagram of the attentional gate model ................................................ 8
3. Mean proportion of “long” responses by duration for neutral and taboo words in Experiment 1 ................................................................. 34
4. Psychometric functions for the mean proportion of “long” responses by duration for a typical participant in Experiment 1 ........................................ 37
5. Mean number of neutral and taboo words recalled in the surprise free recall test in Experiment 1 ................................................................. 38
6. Mean proportion of “long” responses by duration for words that were later recalled versus words that were not recalled in Experiment 1 ........................................ 39
7. Mean proportion of correct responses in the neutral-longer and taboo-longer conditions in Experiment 2 ......................................................... 47
8. Mean proportion of correct responses by duration pair for the neutral-longer and taboo-longer conditions in Experiment 2 ......................................... 48
9. Mean number of neutral and taboo words recalled in the surprise free recall test in Experiment 2 ................................................................. 49
10. Mean reproduced durations of taboo and neutral words in the retrospective timing task in Experiment 3 ......................................................... 54
11. Text styles for the standard and novel conditions in Experiment 4 ............ 64
12. Mean proportion of “long” responses by duration for the standard and novel conditions in Experiment 4 ......................................................... 65
13. Mean number of words recalled in the standard and novel conditions in the surprise free recall test in Experiment 4 ............................................... 68
14. Mean proportion of “long” responses by duration for the standard and oddball conditions in Experiment 5 ......................................................... 72
15 Mean percentage of words recalled in the standard and oddball conditions in the surprise free recall test in Experiment 5. ................................. 75

16 Mean proportion of “longer” responses for word and shape oddballs in Experiment 6. ................................................................. 81
ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Don MacKay for all of his guidance and encouragement over the past six years. I would also like to thank the members of my committee, Dean Buonomano, Alan Castel, and Aimee Drolet Rossi, for their valuable suggestions. In addition, I would like to thank all of the current and past members of CogFog and the Castel Lab Meeting for their feedback on this and all of my other research projects.

I would especially like to thank my research assistants, Yuqiao Cao, Gaudia Correia, Juliet Ernst, Leika Ilagan, Kelcey Nishihara, Yu Pan, Stephen Perez, Sylvia Quach, and Sara Warner, for collecting data for the experiments described in this dissertation. I would also like to thank Sadie Aguilar, Victoria Chala, Kelly Fazel, Amanda Hassid, Bryan Holbrook, Betty Huang, Scott Korchinski, Carly Martinetti, Zachary Sandoval, and Ben Taylor for their help with other experiments in our lab. Without the invaluable assistance of these hard-working students, none of this would have been possible.

Finally, I would like to thank my family, and especially my husband Joseph, for their constant love and support.
VITA

2008
B.A., Neuroscience and Linguistics
Pomona College
Claremont, California

2009
M.A., Psychology
University of California, Los Angeles

2009–2012
Teaching Assistant and Teaching Associate
Department of Psychology
University of California, Los Angeles

2012
C. Phil, Psychology
University of California, Los Angeles

2013–2014
Teaching Fellow
Department of Psychology
University of California, Los Angeles

2013
Teaching Practicum Program
Instructor: Introductory Psychology
Department of Psychology
University of California, Los Angeles

2014
Collegium of University Teaching Fellows
Instructor: Psychology of Time, Emotion, and Memory
Department of Psychology
University of California, Los Angeles

PUBLICATIONS


CHAPTER 1: INTRODUCTION AND OVERVIEW

It is relatively easy to think of anecdotal examples of occasions when emotion has seemed to influence our perception of time. Some people report terrifying situations, like car accidents, when time feels like it moves in slow motion. On the other hand, time often feels like it flies by when we are enjoying ourselves. So what is the relationship between emotion and time perception? Under what conditions does emotion make the passage of time feel subjectively slower or faster?

There is a large body of work in the research literature that investigates time perception in general. Much of this research makes use of psychophysical experiments. There is also an expanding literature investigating the neural bases of time perception. However, there is still relatively little work looking at the effects of emotion on time perception, although the field seems to be growing. In fact, there is surprisingly little research that addresses time perception using stimuli with any semantic content at all, with research on words being especially rare. This dissertation will review the existing literature on duration judgments for emotional stimuli, then present a series of experiments that were designed to investigate some of the underlying causes of the effects of emotional words on time perception.

Time perception is typically evaluated through the use of various types of duration judgments, with the most common types described below. When time subjectively seems to be moving very quickly, participants tend to underestimate the length of durations. In contrast, when time feels like it is moving at a slower rate, participants tend to overestimate the length of durations.

Methods Used in Time Research

Studies of time perception that ask participants to make duration judgments use one of two paradigms: prospective or retrospective (Brown, 2010; Hicks, Miller, & Kinsbourne, 1976). In prospective timing tasks, participants know before they begin that they will be
asked to make time judgments. In retrospective timing tasks, participants are not warned ahead of time that they will be asked to make any sort of time judgment, so the timing task comes as a surprise. Although retrospective timing tasks may be more easily applicable to real-life situations, the obvious downside is that only one data point can be collected per participant, since only the first timing judgment is a surprise, making all later judgments inherently prospective. In most of the time perception literature, retrospective tasks have been used for estimation of relatively long durations (e.g., minutes or hours), while prospective tasks have been used for shorter durations (Grondin, 2010).

There are several experimental methods commonly used in time perception studies. Participants can be asked to make a duration judgment verbally, by producing or reproducing a duration, or by comparing multiple durations (Zakay & Block, 1997). In verbal estimation tasks, participants are presented with a stimulus, and afterward are asked to estimate its duration in standard units of time (e.g., seconds, minutes, etc.). In duration production tasks, participants are asked to produce an interval of a duration specified by the experimenter. For example, the experiment could ask the participant “How long is 2 seconds?” and then the participant would attempt to hold down a button for exactly 2 seconds. In duration reproduction tasks, participants experience a stimulus, and then attempt to reproduce its duration, often by attempting to hold down a button for an equivalent duration. In comparison tasks, participants are presented with at least two stimuli and are asked to compare their durations. Depending on the specific task, they might be asked which stimulus was shorter or longer, or whether two intervals were the same length or different lengths. Another common variant is called the temporal bisection task. Participants begin by gaining extensive practice at recognizing intervals of two different durations, usually labeled “short” and “long”. The participants then view each stimulus and decide whether it was presented for an interval closer to the learned “short” or “long” duration. Most studies find that participants tend to underestimate durations when making prospective time judgments, particularly in duration reproduction tasks (Angrilli, Cherubini, Pavese, & Mantredini, 1997). Retrospec-


Theoretical Models

There is an extensive literature detailing various models that attempt to explain how the brain perceives and processes time (for reviews, see Buhusi & Meck, 2005; Grondin, 2010; Ivry & Schlerf, 2008; Wittmann & van Wassenhove, 2009). Generally speaking, these models can be divided into two categories: dedicated and intrinsic (Ivry & Schlerf, 2008). Dedicated models of time perception propose that there are neural mechanisms used specifically for timing. Variations of dedicated models include pacemaker-counter models (see, e.g., Gibbon, Church, & Meck, 1984) and oscillator models (see, e.g., Church & Broadbent, 1990). Intrinsic models suggest that, rather than using specific mechanisms dedicated to timing, the brain represents duration through intrinsic properties of other neural mechanisms. For example, the state-dependent network model says the brain can measure time based on time-dependent changes in the state of neural networks (Karmarkar & Buonomano, 2007). According to the energy readout model, duration can be measured based on the amount of neural activity used to represent a stimulus (Eagleman & Pariyadath, 2009). However, research on intrinsic models has tended to focus on very short durations, in the range of milliseconds. There is evidence that different mechanisms may be used to process time at different scales (Buonomano, 2007). The majority of research on time perception from a cognitive perspective has used longer durations, from approximately half a second to several seconds, and in some cases several minutes. In addition to timing mechanisms, these longer intervals may also require the use of other cognitive processes, such as attention and memory. Dedicated models, as well as variations that incorporate cognitive processes, will be the primary focus of this review.

Some of the most popular theories used to explain time perception are known as internal clock models (for review, see Wearden, 2001). One of the simpler internal clock
models is the pacemaker-accumulator model (Treisman, 1963). This model postulates an internal pacemaker that emits pulses at regular intervals, functioning as the “ticking” of the clock. The accumulator counts these pulses, giving the brain a way to estimate the passage of time. A greater number of pulses counted during a particular interval results in a longer duration judgment. Another important aspect of this model is a switch connecting the pacemaker and the accumulator. When the switch is on, triggered by a start signal that indicates incoming temporal information, pulses from the pacemaker are counted by the accumulator. When it is off, pulses are not counted. Variations in the accuracy of temporal judgments can be explained by changes in the speed of the pacemaker, which can vary based on arousal levels, or in the function of the switch.

One of the most influential internal clock models is scalar expectancy theory (Gibbon, 1977; Gibbon et al., 1984), which expands on the pacemaker-accumulator model by adding some additional components. A diagram of scalar expectancy theory is shown in Figure 1. The pacemaker, switch, and accumulator function in the same way as in the pacemaker-accumulator model, in what is known as the clock stage. The memory stage includes a mechanism for storing mental representations of different durations, which can be stored in either working memory or long-term memory. Representations of durations that are only needed temporarily, in an experimental task for example, are stored in working memory, while representations of more frequently used intervals (e.g., seconds, hours, etc.) are stored as reference durations in long-term memory. The third component of scalar expectancy theory is the decision stage, which allows the comparison of different durations. This stage is necessary when trying to estimate the length of a time interval by comparing it to the reference durations stored in memory, or when directly comparing different durations in a timing task. One fundamental characteristic of scalar expectancy theory, and the source of its name, is the scalar property, which is the requirement that the standard deviation for temporal estimates increases linearly as a function of the mean. Evidence in support of the scalar property has been found in animals (Gibbon, 1977) and in human participants.
Cognitive models of time perception have proposed that processes like memory and attention play a central role in forming duration judgments. One example of a cognitive model that functions without an internal clock is the storage size model (Ornstein, 1969). According to this model, time is not measured directly, but estimates of temporal duration are reconstructed based on non-temporal information stored in memory. Events for which there is a greater quantity of information stored in memory are perceived as being longer than events associated with a smaller quantity of information. For example, in one of Ornstein’s (1969) studies, participants retrospectively estimated the lengths of intervals spent categorizing newly-learned dance moves. The durations of the conditions displaying more complex stimuli (i.e., a larger number of dance moves) were estimated to be longer than the conditions displaying relatively less complex stimuli (i.e., a smaller number of dance moves). Ornstein replicated this basic finding with a variety of other tasks.
In addition to the storage size model, there are several other memory-based cognitive models that do not include an internal clock (for reviews, see Block, 2003; Block & Zakay, 2008). One of these is the contextual change model, which suggests that the length of a duration is estimated retrospectively based on the number of contextual changes that are perceived (Block, 1982; Fraisse, 1963). These contextual changes can affect “process context”, which refers to internal cognitive context produced by specific kinds of processing, and “environmental context”, which refers to the context of the physical environment. According to this model, larger numbers of contextual changes result in longer duration judgments. The model assumes that context is automatically encoded into memory along with each event, so when it is time to retrieve the memory for the event, its duration can be estimated based on the number of different contextual situations associated with it. Since context typically changes more frequently over a long duration than a short duration, a larger number of contextual associations suggest a longer duration. A related model proposes that remembered durations are estimated not based on the amount of information stored or on the number of contextual changes, but rather the segmentation of the events encoded into memory during the interval being judged (Poynter, 1983, 1989). According to the segmentation model, greater segmentation of the remembered events helps to provide reference points. The durations between these reference points are inferred and then used to estimate time intervals.

Research supporting the storage size, contextual change, and segmentation models is based almost entirely on retrospective timing tasks (Lejeune, 1998). However, data from prospective timing tasks tend to show a pattern that is quite different. In these tasks, intervals that include larger quantities of non-temporal information are usually judged to be shorter than intervals containing less non-temporal information. Thomas and Weaver (1975) proposed a model involving attentional resources to explain these findings. In this model, attention is a limited resource that must be divided between timing mechanisms and processing of other non-temporal information. For the accumulator to count pulses from
the pacemaker, it is necessary for attention to be allocated to time. If attention is directed toward some other aspect of a stimulus, less is available to focus on timing, so some of the pulses from the pacemaker are not counted. When fewer pulses are counted, the interval is judged to be of a shorter duration.

The attentional gate model (Zakay & Block, 1997) combines aspects of cognitive timing models with scalar expectancy theory. A diagram of the attentional gate model is shown in Figure 2. The model includes a pacemaker, which emits pulses at a rate sensitive to arousal. Higher arousal levels cause the rate of the pacemaker to speed up, releasing more pulses within a given interval. The accumulator records the pulses, as described in previous models. A switch opens at the start of the relevant interval and closes at the end, signaled by external stimuli relevant to the task. Between the pacemaker and the switch, the attentional gate model adds a gate that is controlled by the amount of attention directed toward time. This gate is theoretically distinct from a switch, since it can be opened or closed to varying degrees, while a switch must be either completely open or completely closed. When more attention is allocated to temporal processing, the gate opens wider and more pulses can reach the counter. When attention is directed elsewhere and less is allocated to temporal processing, the gate is opened more narrowly and fewer pulses make it through. The need for both a switch and a gate in the model is somewhat controversial (Lejeune, 1998, 2000; Zakay, 2000), however, as it may be possible to describe the relevant data by attributing it to the function of the switch. As in scalar expectancy theory, the final pulse count can be stored in working memory and can then be compared to various other durations stored in either long-term or working memory, as required by the task.

Another model that does not require an internal clock is called the memory-age model (Block, 2003). This model attempts to incorporate attentional factors to explain the effects of attention on prospective timing judgments. According to the model, whenever attention is directed toward time, duration is estimated based on the apparent age of the previous act of attending to time. However, the model also claims that the apparent age of a memory
Figure 2. Diagram of the attentional gate model based on description from Zakay and Block (1997). Arousal levels influence the rate of the pacemaker, attention controls the width of the opening of the gate, and the switch is triggered by a start signal indicating incoming temporal information.

Increases according to a negatively accelerated power function, so checking more frequently means that larger changes in the apparent age are detected. This leads to longer duration estimates when more attentional resources are directed toward time, rather than toward non-temporal information. When less attention is directed toward time, the apparent age of the memory is checked less frequently. Due to the shape of the function as it approaches its asymptote, the changes in apparent age do not seem as large, so when they get added up it seems like the overall duration of the event was shorter.

Although the non-clock-based models are intriguing and should be kept in mind as plausible alternatives to clock-based models, the current studies were not intended to help differentiate between them. Because it is one of the most popular models and seems to fit the data well, the attentional gate model will be discussed with respect to the results of the current prospective studies in Experiments 1, 2, 4, 5, and 6. The results of the retrospective
study in Experiment 3 will be discussed with respect to the storage size model, since this is the simplest model that seems to fit the data appropriately.

**Emotion and Time**

A number of studies have addressed the effects of emotional stimuli on time perception (for reviews, see Droit-Volet & Gil, 2009; Droit-Volet & Meck, 2007). In many of the studies, the results are explained in terms of the effects of emotion on either arousal or attention (see, e.g., Buhusi & Meck, 2006; Burle & Casini, 2001). In internal clock models, higher levels of arousal can increase the speed of the pacemaker. When the pacemaker emits more pulses within a given interval, the duration of that interval is judged to be longer. This would predict that highly arousing emotional stimuli would cause durations to be overestimated, making time appear to move at a subjectively slower rate. Emotion may also affect time perception through attentional mechanisms, since emotional stimuli tend to be more attention-grabbing than neutral stimuli (see, e.g., Mathewson, Arnell, & Mansfield, 2008 for evidence from emotional words; Schimmack, 2005 for evidence from emotional pictures). According to cognitive timing models like the attentional gate model, when more attention is directed toward experimental stimuli, less is available for timing. This results in fewer pulses being counted within a given interval, so durations are judged to be shorter. This predicts that highly attention-grabbing emotional stimuli will cause durations to be underestimated, making time appear to move at a subjectively faster rate. Clearly, arousal and attention can have contrasting effects on duration judgments. A number of studies have attempted to determine the relationship between arousal and attention, often with varying interpretations. A selection of those studies, conducted using a variety of emotional stimuli and situations, are reviewed here. See Table 1 for a summary of the studies discussed.
<table>
<thead>
<tr>
<th>Study</th>
<th>Emotional Stimuli</th>
<th>Summary of Results</th>
<th>Suggested Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angrilli, Cherubini, Pavese, &amp; Mantredini</td>
<td>IAPS pictures</td>
<td>For low-arousal pictures, negative were judged as shorter than positive; for high-arousal pictures, negative were judged as longer than positive</td>
<td>Interaction between arousal and attention</td>
</tr>
<tr>
<td>(1997)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noulhiane, Mella, Samson, Ragot, &amp; Pouthas</td>
<td>IADS sounds</td>
<td>Negative sounds were judged as longer than positive; low-arousal sounds were judged as longer than high-arousal</td>
<td>Interaction between arousal and attention</td>
</tr>
<tr>
<td>(2007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tipples (2010)</td>
<td>Taboo and other emotional words</td>
<td>Taboo words were judged as shorter than neutral or other emotional words</td>
<td>Attention</td>
</tr>
<tr>
<td>Droit-Volet, Brunot, &amp; Niedenthal (2004)</td>
<td>Faces</td>
<td>Emotional faces were judged as longer than neutral faces</td>
<td>Arousal</td>
</tr>
<tr>
<td>Thayer &amp; Shiff (2010)</td>
<td>Eye contact with experimenter who was either scowling or smiling</td>
<td>Durations were judged as longer when the experimenter was scowling than smiling</td>
<td>Arousal</td>
</tr>
<tr>
<td>Gable &amp; Poole (2012)</td>
<td>Positive pictures</td>
<td>High-approach-motivation pictures (desserts) were judged to be shorter than neutral or low-approach-motivation pictures (flowers)</td>
<td>Approach motivation</td>
</tr>
<tr>
<td>Gil, Rousset, &amp; Droit-Volet (2009)</td>
<td>Food pictures</td>
<td>Food pictures were judged to be shorter than shapes; disliked foods were judged to be shorter than liked foods</td>
<td>Attention</td>
</tr>
</tbody>
</table>
Angrilli et al. (1997) studied the effects of emotional pictures on time perception. Participants were shown emotional pictures from the International Affective Picture System (IAPS: Lang, Bradley, & Cuthbert, 2008), manipulated according to both valence and

<table>
<thead>
<tr>
<th>Study</th>
<th>Emotional Stimuli</th>
<th>Summary of Results</th>
<th>Suggested Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stetson, Fiesta, &amp; Eagleman (2007)</td>
<td>31-meter free fall</td>
<td>Duration of the free fall was judged to be longer when experienced than when viewed</td>
<td>Arousal</td>
</tr>
<tr>
<td>Langer, Wapner, &amp; Werner (1961)</td>
<td>Riding on platform moving toward stairwell with potential for fall</td>
<td>Time felt slower when moving toward stairwell than when moving away</td>
<td>Arousal</td>
</tr>
<tr>
<td>Campbell &amp; Bryant (2007)</td>
<td>Skydiving</td>
<td>Ratings of fear were positively correlated with longer duration judgments for first skydiving experience</td>
<td>Arousal</td>
</tr>
<tr>
<td>Gil &amp; Droit-Volet (2009)</td>
<td>Ratings of depression</td>
<td>Participants experiencing higher levels of depression underestimated durations compared to non-depressed participants</td>
<td>Arousal</td>
</tr>
<tr>
<td>Bar-Haim, Kerem, Lamy, &amp; Zakay (2010)</td>
<td>Ratings of anxiety</td>
<td>Participants experiencing higher levels of anxiety showed greater overestimation of durations for threatening faces than non-anxious participants</td>
<td>Arousal</td>
</tr>
<tr>
<td>Twenge, Catanese, &amp; Baumeister (2003)</td>
<td>Social anxiety</td>
<td>Participants in social exclusion condition overestimated durations compared to participants in control condition</td>
<td>Focus on present time</td>
</tr>
<tr>
<td>Rudd, Vohs, &amp; Aaker (2012)</td>
<td>Feelings of awe</td>
<td>Awe expanded participants’ sense of time available</td>
<td>Focus on present time</td>
</tr>
</tbody>
</table>
arousal. The stimuli included images that were positive low-arousal (e.g., pets and babies), positive high-arousal (e.g., erotic scenes), negative low-arousal (e.g., spiders and rats), and negative high-arousal (e.g., bloody wounds). Participants viewed these images for durations of 2, 4, or 6 seconds, and were then asked to complete one of two tasks. In one condition, the participants verbally estimated the duration of each picture, while in the other condition they attempted to reproduce the duration of each picture by holding down a button for an identical time interval. In both conditions, durations were generally underestimated, with more extreme underestimates in the duration reproduction task. Analyses of both tasks revealed the same interaction between valence and arousal. For low-arousal images, durations of negative images were judged to be shorter than durations of positive images. In contrast, the opposite pattern was found for high-arousal images, with durations of negative images being judged to be longer than durations of positive images. Angrilli et al. explained these results by suggesting that duration judgments for low-arousal emotional stimuli are affected primarily by attentional mechanisms. They suggested that more attentional resources are directed toward negative than positive stimuli, resulting in shorter temporal judgments for low-arousal negative pictures than low-arousal positive pictures. Conversely, Angrilli et al. suggested that duration judgments for high-arousal emotional stimuli are affected primarily by arousal mechanisms. A stronger arousal reaction in response to negative than positive stimuli causes durations to seem longer for negative than positive pictures in the high-arousal condition.

Using similar methods, Noulhiane, Mella, Samson, Ragot, and Pouthas (2007) studied duration judgments for emotional sounds. Participants listened to sounds from the International Affective Digitized Sounds System (IADS; Bradley & Lang, 1999). These sounds were classified as positive low-arousal (e.g., laughing), positive high-arousal (e.g., erotic sounds), negative low-arousal (e.g., sobs), negative high-arousal (e.g., woman crying), and neutral (e.g., street noises). These sounds were played for durations of 2–6 seconds, and participants were asked to either verbally estimate the duration or reproduce it by holding down a but-
ton. Results were consistent across both tasks. Participants judged all types of emotional sounds to be longer than neutral sounds at durations of 2–4 seconds, but not longer. Effects of valence and arousal were found only at the 2-second duration. The negative sounds were judged to be longer than the positive sounds, while the low-arousal sounds were judged to be longer than the high-arousal sounds. Unlike in Angrilli et al. (1997), there was no interaction between valence and arousal. Somewhat confusingly, Noulhiane et al. suggested that the main effect of valence was caused by arousal mechanisms, while the main effect of arousal was dependent on attentional mechanisms. They explained that emotional stimuli generally increase the pacemaker rate by increasing arousal, but that negative sounds increase the pacemaker rate more than positive sounds. This explains the longer duration estimates for emotional than neutral stimuli, as well as the longer duration judgments for negative than positive sounds. They explained the main effect of arousal using attentional mechanisms, suggesting that the high-arousal sounds captured attention more than the low arousal sounds. This means that during the high-arousal sounds there was less attention available for the timing task than during the low-arousal sounds, resulting in shorter duration estimates for the high-arousal sounds.

Tipples (2010) investigated duration judgments for emotional words using the temporal bisection task. In this task, participants learned to identify “short” 400 ms intervals and “long” 1600 ms intervals. Six categories of words were used in this experiment: positive low-arousal (e.g., rainbow), positive high-arousal (e.g., ecstasy), negative low-arousal (e.g., mildew), negative high-arousal (e.g., assault), taboo (e.g., shit), and neutral (e.g., chair). Participants were asked to indicate whether each word was shown for a duration that was closer to the “short” or the “long” interval. The results showed that participants gave a smaller proportion of “long” responses for taboo words than for any of the other categories, suggesting that the durations of taboo words were underestimated compared to all of the other word types. The bisection point, defined as the duration where participants were equally likely to respond “long” and “short”, was significantly higher for taboo words than
for any of the other categories, again suggesting that time felt subjectively faster during the presentation of taboo words. Tipples explained these results using purely attentional mechanisms. He explained that taboo words, but not the other word types, caused participants to direct their attention away from the timing task and toward word meaning, leading to reduced duration estimates.

Droit-Volet, Brunot, and Niedenthal (2004) looked at time perception using images of emotional faces as stimuli. They used the temporal bisection task, training participants to recognize durations of 400 and 1600 ms. After the training phase, the participants were shown neutral, angry, happy, and sad faces and were asked to decide whether each face was shown for a duration closer to the learned “short” or “long” interval. The results indicated that participants consistently overestimated the durations of the angry, happy, and sad faces compared to the neutral faces. This overestimation increased with longer durations, which is consistent with arousal-based explanations of emotional time perception. Other experiments have found overestimation of durations associated with negative faces, but not positive faces. For example, Thayer and Schiff (2010) asked participants to make eye contact with an experimenter, then reproduce the duration of the time spent making eye contact. Participants overestimated the duration when the experimenter was scowling, but not when he or she was smiling.

Other studies have manipulated emotion by placing participants in emotional situations. Stetson, Fiesta, and Eagleman (2007) studied time perception during a frightening event by asking participants to estimate the duration of a 31-meter free fall before landing in a safety net. They estimated the duration while watching others complete the free fall, then estimated the duration of their own free fall after experiencing it themselves. The estimates were significantly longer when they recalled their own fall than when watching others, suggesting that frightening events do feel subjectively longer, which is consistent with arousal-based explanations. In another experiment, Langer, Wapner, and Werner (1961) studied the effects of imminent danger on participants’ ability to produce a 5-second interval. The par-
Participants were blindfolded and stood on a platform with wheels. In the dangerous condition, participants were asked to produce the interval while the platform was being pushed toward the top of a stairwell. In the non-dangerous condition, the platform was being moved away from the stairs. The results indicated that time felt subjectively slower in the dangerous condition compared to the non-dangerous condition, consistent with arousal-based explanations of emotion and time perception. In a study on novice skydivers, Campbell and Bryant (2007) found that ratings of fear were positively correlated with the length of judgments of the duration of participants’ first skydiving experience. Again, the higher arousal state induced by fear was associated with increased overestimation of durations.

The effects of emotion on duration judgments seem to vary immensely depending on the type of emotional stimuli used. The results of studies that placed participants in frightening situations (Campbell & Bryant, 2007; Langer et al., 1961; Stetson et al., 2007) or showed them emotional facial expressions (Droit-Volet et al., 2004; Thayer & Schi, 2010) could be explained entirely by arousal effects: Due to the increase in the speed of the internal pacemaker during the emotional conditions, time seemed subjectively slower. Tipple’s (2010) results using emotional words could be explained entirely by attentional effects: The taboo words were more attention-grabbing than the other words, leaving less attention to be allocated to the timing task and causing time to seem subjectively faster. However, Angrilli et al.’s (1997) findings using emotional pictures and Noulhiane et al.’s (2007) findings using emotional sounds were a bit more complicated. Angrilli et al. found an interaction between valence and arousal, while Noulhiane et al. found main effects of both valence and arousal, but no interaction between them. They each proposed different mechanisms to account for their findings, using different combinations of arousal- and attention-based explanations. Very few of the studies in the emotional time perception literature offer convincing explanations for why some aspects of emotional stimuli affect time perception via arousal mechanisms and others via attentional mechanisms. These explanations are usually presented in a manner that seems very post-hoc and at times circular. Much more research
is needed to clarify the relationships between attention, arousal, and the perception of time.

Rather than asking participants to make temporal judgments for emotional stimuli, other experiments have tested for effects of the participants’ emotional states or moods on time perception (for review, see Droit-Volet, 2013). For example, Gil and Droit-Volet (2009) found that participants experiencing high levels of depression and sadness, as evaluated by the Beck Depression Inventory, tended to underestimate durations in a temporal bisection task to a greater extent than control participants did. They attributed this to a general slowing of the internal clock caused by depressive mood. In contrast, Bar-Haim, Kerem, Lamy, and Zakay (2010) found that participants with high levels of anxiety, as evaluated by the State-Trait Anxiety Inventory, tended to overestimate durations for threatening faces more than participants with low levels of anxiety. Bar-Haim et al. attributed this effect primarily to an increase in the rate of the internal clock caused by the generally higher levels of arousal produced by anxiety. In another study, Twenge, Catanese, and Baumeister (2003) induced anxiety in their participants by manipulating social exclusion. Participants gathered in the same room and interacted with each other for a while, then were asked to list the names of the other participants they would most want to work with on a future task. The participants in the social exclusion condition were told that no one had selected them so they would need to work alone, while the participants in the social inclusion condition were told that multiple people had selected them. When asked to judge the length of 40- and 80-second intervals, participants in the social exclusion condition overestimated the duration significantly more than the socially included participants. Finally, in one of the few studies manipulating positive mood, Rudd, Vohs, and Aaker (2012) found that feelings of awe expanded participants’ sense of the amount of time available to them.

While most studies focus on arousal and attention as the main mechanisms by which emotion can influence time perception, a small subset of them addresses other possibilities. Approach-motivation theories have been proposed in an attempt to explain the effects of motivational intensity on attention, which can then influence time perception (Gable &
Harmon-Jones, 2010a, 2008; Gable & Poole, 2012). While levels of motivational intensity are usually correlated with arousal, motivational intensity specifically refers to the drive to act. Positive stimuli high in motivational intensity usually provoke approach responses, often in pursuit of a goal. Negative stimuli high in motivational intensity can provoke withdrawal responses (e.g., disgust) or approach responses (e.g., anger), depending on the specific stimulus. Stimuli low in motivational intensity are described as being goal-irrelevant. High motivational intensity tends to promote attentional narrowing, for both positive (Gable & Harmon-Jones, 2008) and negative stimuli (Gable & Harmon-Jones, 2010b). This attentional narrowing may be helpful in situations with high motivational intensity because it allows people to focus their attention on the goal-relevant information and avoid irrelevant distractions.

Gable and Poole (2012) studied the effects of high and low motivational intensity on time perception, focusing mainly on positive emotional states. They manipulated motivation and emotion by using three different types of pictures: neutral (shapes), positive low-approach-motivation (flowers), and positive high-approach-motivation (pictures of desserts). Using the temporal bisection task, they asked participants to determine whether each picture was shown for durations closer to the learned “short” 400 ms interval or “long” 1600 ms interval. The results showed that participants consistently underestimated the durations of the high-approach-motivation pictures compared to the other pictures types. This suggested that approach motivation, and not simply positive valence, was responsible for the underestimation effects. In a separate study, Gil, Rousset, and Droit-Volet (2009) also looked at the effect of food pictures on duration judgments. Participants were asked to complete a temporal bisection task showing pictures of liked foods, disliked foods, and neutral shapes. Gil et al. found that participants underestimated the durations of food pictures relative to neutral shapes, but that the effect was greater for foods the participants disliked than for foods they liked. Disliked foods are not high in approach motivation, so this finding cannot be explained entirely by motivational factors. Rather than finding greater underestimation
of durations for high-approach-motivation stimuli like Gable and Poole, Gil et al. found the largest underestimation effects for the disgusting foods, which likely elicit withdrawal motivation. Rather than proposing motivational effects, Gil et al. claimed that because the food pictures were low-arousal, their findings were driven primarily by attentional mechanisms.

Gable and Poole (2012) explained that subjective speeding of time during events associated with high motivational intensity might directly aid in the pursuit of goals. If it feels like less time has passed, a person might be willing to spend more time persisting in goal-related behaviors. This idea is supported by other research showing that people are happy when time seems to move quickly: When participants believe that time has moved more quickly than expected when performing a task, they rate that task as more positive and enjoyable (Sackett, Meyvis, Nelson, Converse, & Sackett, 2010). In an additional experiment, Gable and Poole tested whether the subjective speeding of time is specific to positive motivation by manipulating motivational direction, using both positive and negative images that were high in motivational intensity. They used the temporal bisection task while showing participants positive high-approach-motivation pictures and negative high-withdrawal-motivation pictures. The results indicated that participants judged positive high-approach-motivation pictures as being shown for less time relative to the negative high-withdrawal-motivation pictures. The results suggested that the effects of motivational intensity on time perception are specific to positive stimuli, which elicit approach motivation. Models of time perception and emotion that are based entirely on attention would predict that both positive and negative motivational intensity would lead to the underestimation of durations, since both have been shown to promote attentional narrowing (Gable & Harmon-Jones, 2010a, 2008). This narrowing of attention would direct attention toward processing the emotional stimuli and away from the timing task, resulting in underestimation of time. The positive and negative pictures were matched based on arousal, so arousal cannot explain the difference between the two conditions. The findings from this experiment suggest that attention and arousal are not enough to explain the effects of emotion on time perception. Approach motivation
may play a role as well, especially for positive stimuli.

Attention and Time

There has been a tremendous amount of research on the effects of attention on the perception of time, and, in fact, there are entire edited books on the subject (see, e.g., Nobre & Coull, 2010). One of the most robust findings in the timing literature concerns the effects of dual-task conditions, where one of the tasks requires time judgments and the other task requires participants to attend to non-temporal information. Brown (2008) reviewed 57 articles on this subject, with a total of 77 experiments, and found that 70 of those experiments showed interference effects, with performance on timing tasks being disrupted by various non-temporal distractor tasks. In each of these experiments, subjects performed timing tasks alone and at the same time as other non-temporal tasks. Temporal judgments given during the dual-task condition are typically shorter than when there is only one task. This pattern has been shown with a wide range of distractor tasks, including mirror drawing (Brown, 1985), loudness discrimination (Grondin & Macar, 1992) brightness discrimination (Casini, Macar, & Grondin, 1992), the Stroop task (Sawyer, Meyers, & Huser, 1994), and proofreading (Brown & Stubbs, 1992). Other studies have manipulated the difficulty of non-temporal distractor tasks (see, e.g., Block, 1992; Brown, 1985, 2008, 2010), and have found that more difficult distractors tend to cause durations to be underestimated to a greater extent in the timing task. Distractor task difficulty is typically defined based on the level of processing demands, with more difficult tasks requiring more attentional resources. For example, Brown (1985) manipulated difficulty of a perceptual-motor task by asking participants in the easy condition to trace the outline of a star, and asking participants in the difficult condition to trace the outline of the star while viewing it in a mirror. Both conditions required the same movements, but the difficult condition required considerably more cognitive resources to complete, leaving fewer available for the concurrent timing task. Other studies have demonstrated that the interference effect can be attenuated with practice.
as performance of the distractor task becomes more automatic (see, e.g., Brown, 1998; Brown & Bennett, 2002). As participants gain experience with the non-temporal task, its effect on concurrent duration judgments is considerably reduced.

Other research has provided support for the idea that allocation of attentional resources can influence the perception of time by examining the effects of multiple tasks on errors in duration judgments. For example, Brown and West (1990) showed participants four letters displayed on a computer screen, and depending on the condition, asked them to monitor the durations of one, two, three, or all four of the letters for a later duration reproduction task. The results showed that dividing attention between greater numbers of targets led to larger errors in time judgments, suggesting that the attention required for temporal judgments is drawn from a limited pool of attentional resources. Another experiment used the attentional sharing procedure, in which participants were instructed to divide their attention in specific proportions between temporal and non-temporal tasks (Brown, 2010). For example, in one condition they were asked to direct 25% of their attention to time and 75% to another task. In the conditions that asked participants to direct less attention toward time, duration judgments were less accurate.

Time feels subjectively faster when attention is divided between multiple tasks. The results from attentional studies suggest that, on some level, the saying “time flies when you’re having fun” is essentially true, although perhaps it should say something more along the lines of “time flies when you’re not thinking about time”. The old saying “a watched pot never boils” is essentially true as well: When more attention is directed toward time, it feels longer. Cahoon and Edmonds (1980) tested this saying quite literally, asking participants to watch a pot of water on a hot plate and tell the experimenter when it started to boil. The experimenter left the room and returned after a 240-second interval, then asked the participants to estimate the length of the time that he or she was away. The group of participants who had been asked to watch the water boiling gave significantly longer estimates than a control group that had not been asked to watch the pot of water. A different study
published the same year also tested the literal interpretation of the saying: Block, George, Reed (1980) set up a similar experiment, but sometimes the liquid was allowed to boil while the participants watched, and sometimes it was not. Participants judged the duration to be longer if the water never boiled than if it did.

**Verbal Stimuli and Time**

Surprisingly few studies have investigated duration judgments using verbal stimuli, but several of the existing studies have focused on the effects of word length and word frequency on duration judgments. Hochhaus, Swanson, and Carter (1991) found that high-frequency words were judged to be longer than low-frequency words using two different tasks, which included a scale asking participants to rate word durations from 1 (shortest) to 4 (longest), and an ordinality comparison task asking participants to decide which of two words in a pair was displayed for longer. Warm and McCray (1969) discovered that words with more letters were perceived as having shorter durations, and that the size of this effect was greater for low frequency words than for high frequency words. These results make sense when interpreted in terms of the processing resources required for reading. Familiar words are identified more quickly than unfamiliar words (Pierce, 1963), and shorter words are identified more quickly than longer words (McGinnies, Comer, & Lacey, 1952), suggesting that that familiar and shorter words are easier to read. Low frequency words are relatively unfamiliar and require more resources to process, leaving fewer attentional resources available for the timing task. With less attention directed toward time, fewer pulses are counted by the accumulator, which results in shorter duration judgments. Reber, Zimmermann, and Wurtz (2004) found that words were judged to be longer than pronounceable non-words, which in turn were judged to be longer than unpronounceable non-words. The words were easiest to process, followed by the pronounceable non-words, followed by the unpronounceable non-words, and each condition influenced duration judgments accordingly. The same explanation can be applied to words with more letters: They require more resources to read, leaving less
available for the timing task, which results in shorter duration judgments. Because of this relationship between processing and duration judgments, some studies have actually used duration judgments as an index of the processing resources required in a task (see, e.g., Chastain & Ferraro, 1997).

In a task that investigated relative word familiarity in a different way, Rhodes and McCabe (2009) studied the effects of expertise by showing football-specific words to participants with high and low levels of football knowledge. The participants with high levels of football knowledge judged football-related words as being displayed for significantly longer than unrelated words, while duration judgments from participants with low levels of football knowledge were not affected by word category. This effect can be explained based on football experts' relatively more fluent perception of football-relevant stimuli. If the football experts are able to process football words more easily due to familiarity, then more resources are left over to direct toward the timing task, resulting in longer duration judgments for the more fluent words.

Finally, as described above, Tipples (2010) compared taboo, neutral, and other emotional words in the temporal bisection task. Compared to the neutral words, participants underestimated durations for the taboo words. The increased levels of attention directed toward the taboo words reduced the resources available for the timing task, leading to shorter duration judgments. Nearly all of the experiments investigating the effects of words on duration judgments have found the same pattern of effects: Words that are easier to process are judged to last longer than words that are more difficult to process. This pattern of results supports the predictions made by theories that propose effects of attention on duration judgments, such as the attentional gate and memory age models.

Goals of the Present Studies

The present studies explored duration judgments for verbal stimuli, and attempted to test the predictions of information-processing models of time perception, such as the at-
tentional gate and storage size models. Although other theoretical models are interesting, plausible, and certainly worthy of further study, these were the only models directly tested in the present research. Experiments 1, 2, and 3 examined the effects of taboo words on duration judgments, while Experiments 4, 5, and 6 examined the effects of novelty on duration judgments, as manipulated by the appearance of the text. Experiments 1, 2, and 3 attempted to explain the effects of emotional words on duration judgments by suggesting attentional mechanisms, and Experiments 4, 5, and 6 attempted to replicate some of those effects by manipulating novelty rather than word emotion.

Experiment 1 utilized a temporal bisection task with taboo and neutral words, followed by a surprise free recall test for all of the words in the task. The goal was to replicate the finding that durations for taboo words are underestimated compared to neutral words, as well as provide a chance to test the hypothesis that the timing effects are partially driven by emotion-linked memory encoding processes. Experiment 2 used an ordinality comparison task with taboo and neutral words, followed by a surprise free recall test. The purpose was to provide support for the results of the previous experiment by replicating the findings using a different task. While the first two experiments used prospective timing tasks, Experiment 3 used a retrospective timing task. Participants were asked to make a surprise duration reproduction judgment for either a taboo or a neutral word, depending on the experimental condition. The goal was to test the prediction that in retrospective tasks, unlike in prospective tasks, durations for taboo words are overestimated compared to neutral words. Based on the storage size model, durations for stimuli that require more processing resources are judged to be longer in retrospective timing tasks. In contrast, durations for stimuli that require more processing resources are judged to be shorter in prospective timing tasks, as predicted by the attentional gate model. The same factors that cause durations to be underestimated in prospective tasks can therefore cause them to be overestimated in retrospective tasks.

Experiment 4 used a temporal bisection task with neutral words displayed in standard
and novel text styles, which were made up of various combinations of colors, fonts, and sizes. By displaying words in attention-grabbing text styles, this experiment attempted to replicate the results of Experiment 1, and test the hypothesis that the effects of taboo words on the timing task were primarily caused by attentional, and not exclusively emotional, mechanisms. Experiment 5 used a temporal bisection task very similar to the previous experiment, except that the words in the novel text condition appeared far less frequently. The goal was to design an experiment that was more closely comparable to the oddball paradigm (see, e.g., Tse, Intriligator, Rivest, & Cavanagh, 2004), which typically shows that the durations of novel and uncommon, or “oddball”, stimuli tend to be overestimated compared to standard stimuli. Experiment 6 used the oddball task with both words and simple shapes as stimuli. Standard words and shapes were medium-sized and appeared in black, while the oddball stimuli appeared in various colors, fonts, and sizes. The purpose was to reconcile the findings of Experiments 4 and 5 with the existing literature on the effects of novelty on duration judgments.
CHAPTER 2: DURATION JUDGMENTS FOR TABOO WORDS

As reviewed in Chapter 1, the effect of taboo words on duration judgments has been shown to be in the opposite direction compared to the effects of nearly all other types of emotional stimuli. Tipples (2010) found that durations of taboo words, which are high-arousal and very negative, were underestimated compared to other word types. Many other studies have found that durations for other types of emotional stimuli are overestimated compared to neutral stimuli. For example, dangerous situations (Langer et al., 1961), high-arousal negative pictures (Angrilli et al., 1997), images of emotional faces (Droit-Volet et al., 2004), and negative sounds (Noulhiane et al., 2007) have resulted in longer duration judgments when compared to neutral stimuli. Because only one published study has examined the effects of taboo words on duration judgments, and especially because its results were so different compared to other studies in the field, it seems worthwhile to attempt to replicate this finding and explore it further. The three studies described in this chapter will examine the effects of taboo words on duration judgments. Experiment 1 used a temporal bisection task, Experiment 2 used an ordinality comparison procedure, and Experiment 3 used a retrospective duration reproduction task. These different tasks are described in more detail below.

Experiment 1 compared the effects of taboo and neutral words on duration judgments in the temporal bisection task. The temporal bisection task trains participants to learn to recognize two different standard time intervals at the beginning of the experiment. After the training phase, the participants view stimuli displayed for durations equivalent to the learned standard intervals, as well as for a range of durations in between the standard intervals. The participants must answer whether each stimulus is displayed for a duration closer to the shorter standard interval or closer to the longer standard interval. This task was originally developed for use with animal subjects, including pigeons (Stubbs, 1968), and rats (Church & Deluty, 1977), but it is now one of the most commonly used methods in studies of human time perception. By analyzing the participants’ responses, it is possible to calculate the
bisection point, also known as the point of subjective equality. This is the stimulus duration where participants are equally likely to respond “short” and “long”. The bisection point represents the duration that participants subjectively feel is halfway between the longer and shorter standard intervals. Animal studies usually find the bisection point to be at the geometric mean between the longer and shorter standard intervals (Church & Deluty, 1977), but with human participants the bisection point is often closer to the arithmetic mean (Wearden, 1991). However, this difference may be explained by the tendency for the experimenters running animal studies to select standard intervals that vary by a smaller ratio (often 2:1) than in human studies (often 4:1). In a study with human participants, Wearden and Ferrara (1996) found that larger ratios between the longer and shorter standard intervals tended to produce bisection points near the arithmetic mean, while smaller ratios tended to produce bisection points near the geometric mean. Each of the temporal bisection studies in this dissertation used longer and shorter standard intervals that differed by a 4:1 ratio, so bisection points near the arithmetic mean were predicted.

By analyzing the results of the temporal bisection task at different durations, it is also possible to test some specific predictions of internal clock models. According to the attentional gate model, both arousal and attention can influence duration judgments. Arousal is hypothesized to affect duration judgments by increasing the rate of pulses emitted by the pacemaker, while attention influences the number of pulses counted by the accumulator, perhaps by causing the opening and closing of an attentional gate. Many studies have investigated methods of differentiating between the relative effects of arousal and attention on duration judgments (see, e.g., Buhusi & Meck, 2006; Burle & Casini, 2001). One of the simplest ways to address this, however, is to compare the size of the effect at different durations (Gil & Droit-Volet, 2009). If the stimulus of interest affects duration judgments by changing the rate of the pacemaker, effects should be larger at longer durations than at shorter durations. This occurs because the effects of the rate difference compound over longer durations. Alternatively, if the stimulus of interest affects duration judgments through
attentional mechanisms, the effect should be constant across durations. When task responses are plotted as a function of duration, the attentional effects can be detected by comparing the intercepts of the functions for different experimental conditions. Clock rate effects are reflected in the slope of the functions.

Experiment 2 used the ordinality comparison procedure, which is simpler than the temporal bisection task in several ways. In this task, participants view pairs of words presented sequentially, with the two words being displayed for slightly different durations. The participants are then asked to decide which of the two words was displayed for longer. The extensive training procedure required for the temporal bisection task is not necessary for the ordinality comparison procedure, since there are no standard durations to learn. The ordinality comparison procedure is not as demanding in terms of memory, since participants only need to remember the durations of the displayed words for long enough to compare the two words in each pair.

In contrast to Experiments 1 and 2, Experiment 3 used a retrospective timing task. As described in Chapter 1, the distinction between prospective and retrospective tasks is one of the most important divisions in the timing literature. In prospective timing tasks, participants are told before they begin that the task requires them to attend to time. In retrospective timing tasks, however, participants are not told to attend to time, so duration judgments come as a surprise. Experimental results typically vary depending on this framing of the timing task, and several theoretical models make different predictions for prospective versus retrospective duration judgments (for reviews, see Block & Zakay, 2001, 2008).

In a meta-analysis, Block and Zakay (1997) reviewed 20 timing experiments, revealing that prospective judgments tend to be longer and less variable than retrospective judgments. Because participants are instructed to attend to time in prospective but not retrospective tasks, more attentional resources are directed toward time during prospective than retrospective tasks, during which attention to time is incidental. This leads to longer duration judgments for prospective than retrospective timing tasks. The greater variability in ret-
rospective judgments suggests that participants use a variety of methods to estimate the length of remembered durations, and that the accuracy of these estimates varies depending on specific task conditions. In their analysis, Block and Zakay found that processing difficulty influenced prospective but not retrospective judgments, while duration length and stimulus complexity influenced retrospective but not prospective judgments. Based on these moderator variables, they concluded that attentional models are necessary for explaining the results from prospective tasks, while memory-based models are necessary for explaining the results from retrospective tasks. The finding that processing difficulty influenced prospective judgments was consistent with the attentional gate model. The finding that stimulus complexity influenced retrospective judgments was consistent with the memory storage and contextual change models.

Several studies have investigated the different effects of prospective and retrospective tasks on duration judgments by comparing them directly as conditions in the same experiment. For example, Miller, Hicks, and Willette (1978) asked participants to study a list of words for a later recall test. There were multiple trials, in which the participants were instructed to either study the word list or to relax. Some were asked to give prospective judgments for the lengths of these trials, while others were asked to give retrospective judgments. In the prospective condition, the study trials were judged to be shorter than the relaxation trials, with judgments for study trials becoming longer when they were given later in the experiment, after multiple study trials. The interpretation was that repetition decreased the attentional resources required for the study task, so more attention was available for timing later in the experiment, which resulted in longer duration judgments. The longest judgments were for the relaxation trials, since there was no distracting study task occurring at the same time. In the retrospective condition, the duration judgments for study trials were shorter when they were given later in the experiment. The interpretation of this result was that retrospective judgments were based on memory for the processing that occurred. Due to repetition, the studying that occurred during the later trials required less cognitive
processing, so fewer memories were encoded. This resulted in fewer memories available for use as cues for estimating retrospective durations, so the later trials were judged to be shorter than earlier trials.

In another study that used both prospective and retrospective judgments, Zakay and Block (2004) asked participants to complete a task requiring them to resolve syntactic ambiguity of sentences, then asked them to make duration judgments for the time spent on the task. In the prospective condition the participants knew ahead of time that they would be asked to judge this duration, while the participants in the retrospective condition did not. In the prospective condition, durations of the ambiguity detection task were judged to be shorter than durations of a reading task without syntactic ambiguity that was used as a control. In the retrospective condition, durations of the ambiguity detection task were judged to be longer than the control task. Similarly, Zakay and Feldman (1993) asked participants to give prospective or retrospective duration judgments for the time spent studying a list of words. The words were presented in an easy recall condition with the words blocked by semantic category, and in a difficult recall condition with the words presented in a random order. The results indicated that prospective judgments were not influenced by recall condition. However, retrospective judgments were longer for the easy recall condition than for the difficult recall condition. This finding supports the memory storage model because the condition in which more memories were stored was judged to be longer.

The three experiments in Chapter 2 attempted to replicate this pattern of results at a smaller time scale. In the experiments reviewed above, durations for more complex stimuli were typically underestimated in prospective tasks and overestimated in retrospective tasks. Rather than comparing these types of judgments when applied to the durations for entire tasks, the current experiments tested them at the level of reading single words. If reading a taboo word requires more processing resources than reading a neutral word, its duration will tend to be underestimated compared to neutral words in prospective tasks due to the additional attentional demands that direct attention away from the primary timing task. For
the same reason, its duration will tend to be overestimated compared to neutral words in retrospective tasks because this same additional processing leads to the formation of a greater number of memories, which can then be used as cues when forming duration judgments.

**Experiment 1: Temporal Bisection Task**

Experiment 1 was based on Tipples (2010), with two modifications and several new analyses. Experiment 1 used only taboo and neutral words, while Tipples included a variety of emotional words. This caused taboo words to make up a much larger proportion of the total number of words displayed (i.e., one half instead of one sixth). One of the primary goals of Experiment 1 was to attempt to replicate the results of Tipples, which showed that durations of taboo words were underestimated compared to other word types.

Another important change was the addition of a surprise memory test at the end of the experiment. Attention-based explanations of the effects of emotion on time perception claim that the durations of emotional stimuli tend to be underestimated because they draw attention away from the timing task. However, if the attention is directed away from timing, what is it being used for? Presumably it is being used to process the emotional stimuli in some way, but few studies in the time perception literature address this directly. One possibility is that attentional resources are being used for the encoding of the emotional stimuli into memory. The addition of a free recall task is a simple test to see if this is plausible. If attentional resources are being diverted away from timing and toward the encoding of emotional stimuli, then this explanation predicts that the durations of taboo words will be underestimated compared to neutral words, but that more taboo than neutral words will be recalled in the free recall test.

If the results of Experiment 1 replicate Tipples (2010) as expected, taboo words will be underestimated compared to neutral words. The bisection point, or point of subjective equality should also occur significantly later for taboo than neutral words. The addition of the free recall test after the timing task allows some interesting analyses that will help test
the hypothesis that memory encoding processes affect performance on the timing task. If the durations for words that are later recalled are underestimated compared to words that are not recalled, it will provide support for the idea that emotion-linked memory encoding processes distract attention away from the timing task, leading to underestimation of durations.

Method

Participants. Thirty-five undergraduates (18–23 years, $M = 19.9$ years, $SD = 1.3$; 14 male, 21 female) from the University of California, Los Angeles (UCLA) received partial course credit for participating. All participants spoke English as their native language.

Materials. The materials consisted of 91 taboo and 91 neutral words from Janschewitz (2008). Taboo words, including profanities, insults, and sexual references, were included in the task due to their emotional content. Presentation of taboo words has been shown to increase skin conductance, a measure of sympathetic nervous system activity and emotional arousal (LaBar & Phelps, 1998). Table 2 shows the mean numbers of letters and syllables and mean ratings of personal use, familiarity, offensiveness, tabooness, valence, arousal, and imagery, as reported by Janschewitz. Mean number of letters, number of syllables, and familiarity ratings did not differ significantly between lists. The taboo words were rated as significantly more negative, more arousing, more offensive, and more taboo than the neutral words. The taboo words were rated as significantly less imageable, and they received significantly lower personal use ratings than the neutral words. In order to control for the possibility that taboo words are a cohesive semantic category, the neutral words all belonged to the category of household objects and activities (e.g., *table*, *scrub*). The words were displayed in black 60-point Arial font.

Procedure. During the initial phase of the experiment, participants were trained to recognize the difference between two standard durations: the “short” 400 ms duration
Table 2

*Characteristics of Taboo and Neutral Words Lists from Experiment 1*

<table>
<thead>
<tr>
<th></th>
<th>Taboo Words</th>
<th>Neutral Words</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>t(180) p-value</td>
</tr>
<tr>
<td>Number of Letters</td>
<td>5.59 1.94</td>
<td>5.73 1.73</td>
<td>0.483 .630</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>1.76 0.83</td>
<td>1.71 0.78</td>
<td>0.367 .714</td>
</tr>
<tr>
<td>Personal Use</td>
<td>3.62 1.50</td>
<td>4.67 1.12</td>
<td>5.385 &lt; .001*</td>
</tr>
<tr>
<td>Familiarity</td>
<td>4.78 1.46</td>
<td>5.13 0.95</td>
<td>1.914 .057</td>
</tr>
<tr>
<td>Offensiveness</td>
<td>2.78 1.12</td>
<td>1.02 0.04</td>
<td>14.983 &lt; .001*</td>
</tr>
<tr>
<td>Tabooness</td>
<td>4.86 1.46</td>
<td>1.05 0.14</td>
<td>24.779 &lt; .001*</td>
</tr>
<tr>
<td>Valence</td>
<td>3.54 1.01</td>
<td>5.11 0.21</td>
<td>14.504 &lt; .001*</td>
</tr>
<tr>
<td>Arousal</td>
<td>4.36 0.96</td>
<td>1.56 0.30</td>
<td>26.520 &lt; .001*</td>
</tr>
<tr>
<td>Imageability</td>
<td>4.57 1.68</td>
<td>7.22 1.13</td>
<td>12.469 &lt; .001*</td>
</tr>
</tbody>
</table>

*Note.* All values from norms collected by Janschewitz (2008). All ratings were made on a 9-point scale: Personal use was rated from 1 (I never use this word) to 9 (I use this word all the time), familiarity was rated from 1 (I never encounter this word) to 9 (I encounter this word all the time), offensiveness was rated from 1 (I am not all offended by this word) to 9 (this word is very offensive to me), tabooness was rated from 1 (not at all taboo) to 9 (very taboo), valence was rated from 1 (strongly negative) to 9 (strongly positive), arousal was rated from 1 (not at all arousing) to 9 (very arousing), and imageability was rated from 1 (doesn’t bring an image to mind) to 9 (brings a vivid image to mind).

and the “long” 1600 ms duration. They viewed eight randomly generated strings of letters presented in the short and long durations in an alternating sequence (i.e., short, long, short, long, etc.). Before beginning, the participants were told that the durations would occur in this sequence. After each letter string, a series of question marks cued the participants to respond by pressing a labeled key to indicate whether the letter string had been displayed for the short or the long duration. The “F” key was labeled “short” and the “J” key was labeled “long”. After each answer, feedback was displayed on the screen for 2 seconds before the next trial began automatically. The trials were separated by an interval that randomly varied in a range of 1–3 seconds. In the second part of the training phase, the participants
viewed another set of eight randomly generated letter strings, but this time the durations were presented in a random order instead of alternating. Participants responded in the same way and continued to receive feedback about each answer.

After completing the training phase, participants began the main phase of the experiment. Each word on the list was presented once for one of seven possible durations, including the short and long durations from the training phase as well as five intermediate durations: 400, 600, 800, 1000, 1200, 1400, and 1600 ms. Assignment of words to durations was counterbalanced between participants. The words were presented in a randomized order. After each word, question marks cued the participants to respond by pressing a labeled key to indicate whether the word had been displayed for a length of time closer to either the short or the long duration learned during the training phase. Feedback was not given and the next trial began automatically after an interstimulus interval that varied randomly in a range of 1–3 seconds.

After completing the temporal bisection task, participants were given a surprise free recall test. They were asked to recall and type in as many words as possible from the timing task. The test was untimed and participants were allowed to list the words in any order.

Results

Figure 3 shows the mean proportion of “long” responses for taboo and neutral words for each of the seven durations. A 2 (word type: taboo vs. neutral) × 7 (duration) repeated measures analysis of variance (ANOVA) revealed a significant main effect of word type, indicating more “long” responses for neutral than for taboo words, $F(1, 34) = 6.324, \eta_p^2 = .157, p = .017$. There was also a significant main effect of duration, with more “long” responses as durations increased, $F(6, 204) = 228.108, \eta_p^2 = .870, p < .001$. There was no significant interaction, $F(6, 204) = 1.725, \eta_p^2 = .048, p = .117$.

This analysis was repeated without the practiced “short” and “long” durations, revealing the same pattern of results. The 2 (word type: taboo vs. neutral) × 5 (duration)
repeated measures ANOVA revealed a significant main effect of word type indicating more “long” responses for neutral than taboo words, $F(1, 34) = 4.939, \eta^2_p = .127, p = .033$. There was a significant main effect of duration with more “long” responses for longer durations, $F(4, 136) = 187.525, \eta^2_p = .847, p < .001$, but no significant interaction, $F(4, 136) = 1.733, \eta^2_p = .048, p = .146$.

To test internal clock speed predictions, the effect of word type was compared for the shorter half of durations used in the experiment (i.e., 400, 600, and 800 ms) versus the longer half of durations used in the experiment (i.e., 1200, 1400, and 1600 ms). When averaging across the shorter half of durations, the mean proportion of “long” responses was significantly higher for neutral words ($M = 0.17, SD = .14$) than for taboo words ($M = 0.13, SD = .13$), $t(34) = 3.175, p = .003$. However, when averaging across the longer half of duration, the
The mean proportion of “long” responses did not differ significantly for neutral ($M = 0.85, SD = .14$) and taboo words ($M = 0.85, SD = .17$), $t(34) = 0.147, p = .884$. This suggests that the effect of word type was driven by the differences at shorter durations.

The 1000 ms duration was of particular interest because it was equally close to the “short” and “long” response durations. Overall, participants showed a bias toward responding ”long” to this duration, with a mean proportion of 0.61 ($SD = 0.19$) “long” responses, which was significantly greater than the chance proportion of 0.50, $t(34) = 3.571, p = .001$. They were significantly more likely to respond “long” for neutral words ($M = 0.65, SD = 0.19$) than for taboo words ($M = 0.57, SD = 0.23$), $t(34) = 2.503, p = .017$.

Summing across all duration conditions, the mean proportion of “long” responses was 0.53 ($SD = .10$) for neutral words and 0.50 ($SD = .10$) for taboo words. One-sample $t$-tests were used to compare these values to the expected proportion of 0.50. Participants chose “long” for the neutral words at a rate marginally greater than 0.50, $t(34) = 1.656, p = .107$, while the proportion of “long” responses for taboo words did not differ significantly from 0.50, $t(34) = 0.142, p = .888$. Thus, while durations of taboo words were underestimated compared to neutral words, duration judgments for taboo words were actually closer to the presented durations, with durations of neutral words possibly being overestimated.

The mean proportion of “long” responses was calculated by item and compared with the ratings given in Janschewitz (2008). This revealed significant positive correlations between the mean proportion of “long” responses and number of letters, number of syllables, and ratings of personal use and familiarity. There were significant negative correlations between the mean proportion of “long” responses and ratings of tabooess and arousal. The mean proportion of “long” responses was not significantly correlated with ratings of offensiveness, valence, or imageability. Correlation coefficients and $p$-values are displayed in Table 3.

The bisection point, or point of subjective equality, was calculated for both taboo and neutral words for each participant by plotting the psychometric function of the stimulus
Table 3
Correlations Between Experiment 1 Timing and Recall Results and Word Ratings from Janschewitz (2008)

<table>
<thead>
<tr>
<th></th>
<th>Mean Proportion of “Long” Responses</th>
<th>Probability of Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation coefficient</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td></td>
<td>$r(180)$</td>
<td>$p$-value</td>
</tr>
<tr>
<td>Number of Letters</td>
<td>.147</td>
<td>.048*</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>.178</td>
<td>.016*</td>
</tr>
<tr>
<td>Personal Use</td>
<td>.168</td>
<td>.016*</td>
</tr>
<tr>
<td>Familiarity</td>
<td>.168</td>
<td>.024*</td>
</tr>
<tr>
<td>Offensiveness</td>
<td>-.114</td>
<td>.125</td>
</tr>
<tr>
<td>Tabooneess</td>
<td>-.181</td>
<td>.014*</td>
</tr>
<tr>
<td>Valence</td>
<td>.061</td>
<td>.411</td>
</tr>
<tr>
<td>Arousal</td>
<td>-.194</td>
<td>.009*</td>
</tr>
<tr>
<td>Imageability</td>
<td>.096</td>
<td>.097</td>
</tr>
</tbody>
</table>

durations versus the proportion of “long” responses. The bisection point represents the duration where participants choose “long” half of the time and “short” half of the time. The statistical package R (R Core Team, 2012) with library psyphy (Knoblauch, 2012) was used to fit the data with a cumulative normal curve using a generalized linear model. See Figure 4 for an example of a psychometric function for a typical participant. Two participants were excluded from this analysis because it was not possible to calculate the bisection point from their functions. The mean bisection point was 1010 ms ($SD = 178$) for taboo words and 971 ms ($SD = 165$) for neutral words. A paired $t$-test revealed a strong trend suggesting that the bisection point was reached later for taboo than for neutral words, $t(32) = 1.991$, $p = .055$. A later bisection point indicates relative underestimation of durations because it means that participants are willing to respond “short” at longer
durations. This analysis suggests that, on average, participants perceived the durations of taboo words to be approximately 40 ms shorter than the neutral words.

Figure 4. Psychometric functions for the mean proportion of “long” responses by duration for a typical participant in Experiment 1. The left panel shows the function for neutral words and the right panel shows the function for taboo words. The bisection point is marked with a vertical line in each panel.

Figure 5 shows the mean numbers of taboo and neutral words recalled in the surprise free recall test. Participants recalled significantly more taboo than neutral words, $t(34) = 10.172$, $p < .001$. Rate of recall was not influenced by display duration, $F(6, 204) = 1.753$, $\eta_p^2 = .049$, $p = .110$. Recall of specific words was compared with the ratings given in Janischewitz (2008). Recall was positively correlated with ratings of offensiveness, tabooness, and arousal, and negatively correlated with valence and imageability. Recall was not significantly correlated with number of letters, number of syllables, or ratings of personal use or familiarity. Correlation coefficients and $p$-values are displayed in Table 3.

An additional analysis compared the proportions of “long” responses that were given for words that were recalled versus words that were not recalled, which are displayed in
**Figure 5.** Mean number of neutral and taboo words recalled in the surprise free recall test in Experiment 1. Error bars represent standard error of the mean.

Figure 6. It was not possible to perform an ANOVA on these data due to missing cases (e.g., if a participant did not recall any words that were shown for 400 ms, all of his or her data would need to be excluded from the analysis). The statistical package R (R Core Team, 2012) with library *lme4* (Bates, Maechler, & Bolker, 2012) was used to perform a linear mixed effects analysis of the relationship between recall condition and the frequency of “long” responses. Recall condition and duration (without the interaction term) were entered into the model as fixed effects. The intercepts for subjects and by-subject random slopes for recall condition and duration were entered as random effects. The p-values were obtained using likelihood ratio tests of the full model compared with the model without the variable of interest. The results revealed a trend indicating that there were fewer “long” responses for words that were recalled than for words that were not recalled, $\chi^2(1) = 2.957, p = .085$. The same analysis was repeated for duration, which revealed a significant effect, with more “long” responses for longer durations, $\chi^2(1) = 84.206, p < .001$. There was no significant interaction between recall condition and duration, $\chi^2(1) = 0.131, p = .717$. 

38
Discussion

The results from Experiment 1 replicated the basic finding from Tipples (2010): Durations for taboo words were underestimated compared to neutral words. In the temporal bisection task, participants responded “long” less frequently for taboo words than for neutral words. In addition, the bisection point, defined as the duration where participants selected “long” half of the time, was higher for taboo than neutral words, which again suggests that durations of taboo words were relatively underestimated. The bisection point also allows us to quantify the underestimation effect, suggesting that taboo words were underestimated by approximately 40 ms on average. This is comparable to the results from the Tipples study, which found that the mean bisection points were 1000 ms for taboo words and 971 ms for
neutral words, while Experiment 1 found that the mean bisection points were 1010 ms for taboo words and 971 ms for neutral words.

One of the primary differences between Experiment 1 and Tipples (2010) was the relative frequency of the appearance of taboo words in the task. In Experiment 1 half of the words were taboo, while in the Tipples study only one sixth of the words were taboo, with the remaining words made up of neutral and non-taboo emotional words. The comparable results of the two studies suggest that the effect of taboo words on duration judgments in the Tipples study was not simply due to their appearance being surprising or relatively unusual within the task. The underestimation effect persisted even when taboo words appeared very frequently, suggesting that the individual words remained attention-grabbing even though category of taboo words was no longer surprising.

Previous experiments using the temporal bisection task have revealed a general bias toward “long” responses, and that appeared to be the case in Experiment 1, with a greater number of “long” than “short” responses throughout the task. This suggests that there was a general overestimation of durations, as has been reported elsewhere in the timing literature. The underestimation of durations of taboo words compared to neutral words actually meant that the responses for taboo words were closer to being correct. One possible explanation is that the emotion associated with the taboo words served to counteract the usual overestimation effect, bringing the proportion of “long” responses closer to the optimal rate. Another alternate possibility is that emotion-linked encoding processes lead to more accurate representations of the temporal context associated with taboo words than neutral words. Previous studies have demonstrated that other aspects of context are recalled more accurately for taboo than neutral words, including word color, word location, and occurrence in the task (MacKay & Ahmetzanov, 2005; MacKay et al., 2004). The display duration for a word is another aspect of context, so it is possible that recall for duration information could also be better for taboo than neutral words. However, most of the existing literature has focused on comparing the duration judgments for emotional and neutral stimuli, with the
assumption that judgments for neutral stimuli form a baseline. The judgments for emotional stimuli are then described in terms of deviations from that baseline. The hypothesis that the shorter duration judgments for taboo than neutral words reflect more accurate recall of durations needs to be tested further, but is supported by the results of Experiment 1.

The use of a range of durations in the temporal bisection task made it possible to test predictions made by internal clock models. These models suggest that if timing effects are driven by changes in arousal, which alters the rate of the pacemaker, the effects should be larger at longer durations than at shorter durations. This occurs because the effects of rate compound over time. When comparing the proportion of “long” responses at the shorter and longer durations used in Experiment 1, the results revealed the opposite pattern, with a significant effect of word type for the shorter durations but not for the longer durations. This finding is more consistent with attention-based explanations for the effect of taboo words on duration estimations. If the initial reading of the word distracts attention away from the timing task, reducing the number of pulses counted by the accumulator, then its effect should be consistent across durations. However, the proportion of time that is lost due to this initial distraction is much larger for shorter durations than longer durations. In other words, if every taboo word is underestimated by 40 ms, this 40 ms makes up a considerably larger proportion of the total duration of a 400 ms stimulus than a 1600 ms stimulus. This effect is therefore large enough to affect responses in the lower range of presented durations but not the upper range.

The other major way that Experiment 1 differed from previous work was through the addition of the surprise free recall test after the timing task. The results revealed that participants recalled an average of nearly four times as many taboo words as neutral words. This is not surprising, and better recall for taboo than neutral words has been shown in a variety of tasks in previous experiments (see, e.g., Guillet & Arndt, 2009; Hadley & MacKay, 2006; MacKay et al., 2004). The finding that taboo words were both underestimated and better recalled compared to neutral words provides some support for the idea that addi-
tional processing of taboo words causes their durations to be underestimated compared to neutral words. The automatic emotion-linked memory encoding processes that occur when viewing taboo words (MacKay & Ahmetzanov, 2005) may distract attention away from the timing task, resulting in underestimation of durations. Further support for this hypothesis is provided by the analysis that suggested that the durations of the specific words that were recalled in the memory test tended to be underestimated compared to the words that were not recalled. However, this analysis only revealed a trend in the predicted direction ($p = .085$), so strong conclusions are not warranted. Further research is necessary to explore the relationship between memory encoding and the attentional processes that interact with duration judgments.

In one of the few studies specifically investigating the relationship between time perception and memory encoding, Dirnberger et al. (2012) showed participants pairs of emotional pictures selected from the IAPS database (Lang et al., 2008) and asked them to judge which of the two was shown for longer. Afterwards, they were given a recognition memory test. As in other studies looking at time judgments for emotional pictures (Angrilli et al., 1997), participants tended to overestimate the durations of high-arousal negative pictures. In addition, the pictures associated with overestimated durations were the same pictures that were more likely to be recognized in the memory test. Dirnberger et al. suggested that the increased arousal in response to the emotional pictures caused an accelerated rate of processing, which helped with memory encoding in the same trials where the durations were overestimated. This experiment differed from Experiment 1, since it found overestimation of the durations of emotional stimuli, rather than underestimation. However, it is interesting that a relationship between timing performance and memory was discovered, although it is not entirely clear why the stimuli with overestimated durations were recalled better in Dirnberger et al., while the stimuli with underestimated durations were recalled better in Experiment 1.

The correlational analysis that the compared the probability of recall for each of the
words in Experiment 1 with the ratings given in Janschewitz (2008) primarily revealed the expected results, indicating that the more strongly emotional words tended to be recalled more frequently. Words that were rated as more offensive or more taboo were more likely to be recalled. Words that had more negative valence ratings and higher arousal ratings were also more likely to be recalled. The more surprising result was the negative correlation between imageability and probability of recall, since typically more highly imageable words are easier to recall (see, e.g., Dewhurst & Conway, 1994). However, in the set of words used in Experiment 1, taboo words were rated as significantly less imageable than the neutral words. The higher frequency of recall and lower imageability ratings for taboo words probably drove this correlation. In fact, when the correlational analysis was repeated for neutral words only, the relationship between recall and imageability was in the expected direction, $r(89) = .294$, $p = .005$.

Some of the correlations comparing performance on the timing task with the ratings and numbers of letters and syllables from Janschewitz (2008) were more surprising, given previous findings in the literature. There were positive correlations between the proportion of “long” responses and the numbers of letters and syllables. This contrasts with previous research suggesting that durations for longer words tend to be underestimated compared to shorter words (Warm & McCray, 1969). The analysis also revealed positive correlations between the proportion of “long” responses and ratings of personal use and familiarity. This provides support for previous findings that indicate that more familiar words are judged to be longer than less familiar words (Hochhaus et al., 1991; Warm & McCray, 1969). Less familiar and less frequent words may require more processing resources to read, which leaves fewer resources available to process time, resulting in shorter duration judgments. Finally, the significant negative correlations between the proportion of “long” responses and ratings of taboo-ness and arousal were in the expected direction. Durations for higher arousal and more taboo words were relatively more likely to be underestimated. Ratings of offensiveness and valence were less effective as predictors of which words would be underestimated.
Experiment 2: Ordinality Comparison Procedure

The goal of Experiment 2 was to attempt to replicate the findings of Experiment 1 using the ordinality comparison procedure (see, e.g., Allman & Meck, 2012). In this task, participants viewed pairs of words presented one-at-a-time for slightly different durations, and were asked to determine whether the first or the second word was displayed for longer. Each pair was made up of one taboo word and one neutral word. Then, after the timing task, participants were asked to recall as many words as possible in a surprise free recall task.

One advantage of the ordinality comparison procedure over the temporal bisection task is that it does not require participants to remember durations for an extended time. In the temporal bisection task, they must attempt to remember the standard “short” and “long” durations throughout the entire experiment. In the ordinality comparison procedure, they only need to remember the durations of the two words long enough to decide which is longer, making the results less susceptible to effects of memory decay.

If the results of Experiment 2 replicate the findings of Experiment 1 and Tipples (2010), participants will choose the correct answer less frequently in the taboo-longer condition than in the neutral-longer condition, suggesting that durations of taboo words are underestimated compared to neutral words. In other words, participants will have trouble correctly saying that the taboo word is longer because sometimes they will actually perceive it as being shorter than the neutral word. When the neutral word is actually longer, participants will have an easier time correctly answering that it is longer. Furthermore, this effect will be larger at the shorter durations tested, as found in Experiment 1. If durations of taboo words are underestimated due to emotion-linked memory encoding processes, more taboo than neutral words will be recalled in the surprise free recall test.
Method

Participants. Forty UCLA undergraduates (18–23 years, \( M = 20.1 \) years, \( SD = 1.28 \); 12 male, 28 female) received partial course credit for participating. All participants spoke English as their native language.

Materials. The materials consisted of 92 taboo and 92 neutral words from Janschewitz (2008). Unlike in Experiment 1, the neutral words came from a list of category-unrelated words. Table 4 shows the mean numbers of letters and syllables and mean ratings of personal use, familiarity, offensiveness, tabooness, valence, arousal, and imagery, as reported by Janschewitz. Mean number of letters, number of syllables, and familiarity ratings did not differ significantly between lists. The taboo words were rated as being significantly more negative, more arousing, more offensive, and more taboo than the neutral words. The taboo words were rated as being significantly less imageable, and they received significantly lower personal use ratings than the neutral words. The words were displayed in black 60-point Arial font.

Procedure. For each trial, participants viewed one taboo and one neutral word presented sequentially. They were asked to determine which word was shown on the computer screen for longer, indicating their choice by pressing a key labeled either “first word” or “second word”. In half of the trials the taboo word was displayed for longer, and in the other half the neutral word was displayed for longer. Taboo and neutral words were paired randomly. Assignment of word pairs to the taboo-longer and neutral-longer conditions was counterbalanced between subjects. The order of presentation for the words in each pair was also counterbalanced between subjects.

The following pairs of durations were used, with durations differing according to a 4:5 ratio: 600 and 750 ms, 650 and 812 ms, 700 and 875 ms, and 750 and 937 ms (as used in an ordinality comparison task by Xuan, Zhang, He, & Chen, 2007). The first word was preceded
Table 4  
*Characteristics of Taboo and Neutral Words Lists from Experiment 2*

<table>
<thead>
<tr>
<th></th>
<th>Taboo Words</th>
<th>Neutral Words</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Number of Letters</td>
<td>5.58</td>
<td>1.94</td>
<td>5.66</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>1.75</td>
<td>0.83</td>
<td>1.80</td>
</tr>
<tr>
<td>Personal Use</td>
<td>3.64</td>
<td>1.51</td>
<td>4.17</td>
</tr>
<tr>
<td>Familiarity</td>
<td>4.79</td>
<td>1.46</td>
<td>4.78</td>
</tr>
<tr>
<td>Offensiveness</td>
<td>2.77</td>
<td>1.12</td>
<td>1.07</td>
</tr>
<tr>
<td>Tabooness</td>
<td>4.83</td>
<td>1.49</td>
<td>1.10</td>
</tr>
<tr>
<td>Valence</td>
<td>3.54</td>
<td>1.01</td>
<td>5.05</td>
</tr>
<tr>
<td>Arousal</td>
<td>4.34</td>
<td>0.99</td>
<td>1.66</td>
</tr>
<tr>
<td>Imageability</td>
<td>4.54</td>
<td>1.69</td>
<td>6.16</td>
</tr>
</tbody>
</table>

*Note.* All values from norms collected by Janschewitz (2008). All ratings were made on a 9-point scale: Personal use was rated from 1 (I never use this word) to 9 (I use this word all the time), familiarity was rated from 1 (I never encounter this word) to 9 (I encounter this word all the time), offensiveness was rated from 1 (I am not at all offended by this word) to 9 (this word is very offensive to me), tabooness was rated from 1 (not at all taboo) to 9 (very taboo), valence was rated from 1 (strongly negative) to 9 (strongly positive), arousal was rated from 1 (not at all arousing) to 9 (very arousing), and imageability was rated from 1 (doesn’t bring an image to mind) to 9 (brings a vivid image to mind).

by a 500 ms blank screen. The first word was shown for its assigned duration, followed by an interstimulus interval of 200 ms, then the second word for the corresponding duration. After another 200 ms blank screen, a series of question marks prompted the participants to press a key to indicate which word had been displayed for longer. The key press triggered the start of the next trial. This process was repeated for each of the 92 word pairs.

After completing the timing task, participants were given a surprise free recall test, with the same procedure described in Experiment 1.
Results

Figure 7 shows the mean proportions of correct responses in the neutral-longer and taboo-longer conditions. A paired $t$-test revealed that participants were significantly more likely to choose the correct answer in the neutral-longer condition than in the taboo-longer condition, $t(39) = 2.714, p = .010$. Another paired $t$-test showed that the proportion of correct answers did not significantly differ when the longer word was displayed first ($M = .74, SD = 0.13$) than when it was displayed second ($M = .74, SD = 0.14$), $t(39) = 0.131, p = .896$.

![Figure 7. Mean proportion of correct responses in the neutral-longer and taboo-longer conditions in Experiment 2. Error bars represent standard error of the mean.](image)

In order to investigate the effect of the specific durations used, the data were analyzed using a 2 (condition: taboo-longer vs. neutral-longer) $\times$ 4 (duration pair: 600-750 ms vs. 650-812 ms vs. 700-875 ms vs. 750-937 ms) repeated measures ANOVA. This revealed a significant main effect of condition such that participants were significantly more likely to choose the correct answer in the neutral-longer condition than in the taboo-longer condition, $F(1, 39) = 7.533, \eta^2_p = .162, p = .009$. There was also a significant main effect of duration pair, with participants choosing the correct answer more frequently for the longer duration
pairs than for the shorter duration pairs, $F(3, 117) = 6.306, \eta_p^2 = .139, p = .001$. There was a trend suggesting an interaction between condition and duration pair, $F(3, 117) = 2.212, \eta_p^2 = .054, p = .090$. Paired samples $t$-tests indicated that there was a significant difference between the taboo-longer and neutral-longer conditions in the 600-750 ms pair, $t(39) = 3.341, p = .002$, and the 700-875 ms pair, $t(39) = 2.836, p = .007$, a marginal effect in the 650-812 ms pair, $t(39) = 1.639, p = .109$, but no difference in the 750-937 ms pair, $t(39) = 0.694, p = .492$. Figure 8 shows the mean proportion of correct responses by duration pair for the neutral-longer and taboo-longer conditions.

![Figure 8](image)

*Figure 8.* Mean proportion of correct responses by duration pair for the neutral-longer and taboo-longer conditions in Experiment 2. Error bars represent standard error of the mean.

Figure 9 shows the mean numbers of taboo and neutral words recalled in the surprise free recall test. Significantly more taboo than neutral words were recalled, $t(39) = 11.483, p < .001$. Recall of specific words was compared with the ratings given in Janschewitz (2008).
Recall was positively correlated with ratings of familiarity, offensiveness, taboorness, and arousal, and negatively correlated with valence. Recall was not significantly correlated with number of letters, number of syllables, ratings of personal use, or imageability. Correlation coefficients and p-values are displayed in Table 5.

![Figure 9](image.png)

**Figure 9.** Mean number of neutral and taboo words recalled in the surprise free recall test in Experiment 2. Error bars represent standard error of the mean.

**Discussion**

The results supported the prediction that the durations of taboo words would be underestimated compared to neutral words, in agreement with Experiment 1 and Tipples (2010). Because the task was forced-choice, whenever participants failed to correctly label the taboo word as being shown for longer in the taboo-longer condition, they were incorrectly labeling the neutral word as longer. The finding that participants were significantly less likely to choose the correct answer in the taboo-longer condition than in the neutral longer condition suggested that they frequently mistakenly perceived the taboo word as being displayed for less time than the neutral word. The durations of taboo words were underestimated significantly more often than the durations of neutral words, suggesting that
Table 5  
*Correlations Between Experiment 2 Recall Results and Word Ratings from Janschewitz (2008)*

<table>
<thead>
<tr>
<th></th>
<th>Probability of Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td></td>
<td>$r(182)$</td>
</tr>
<tr>
<td>Number of Letters</td>
<td>-.079</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>-.081</td>
</tr>
<tr>
<td>Personal Use</td>
<td>-.030</td>
</tr>
<tr>
<td>Familiarity</td>
<td>.151</td>
</tr>
<tr>
<td>Offensiveness</td>
<td>.612</td>
</tr>
<tr>
<td>Tabooness</td>
<td>.678</td>
</tr>
<tr>
<td>Valence</td>
<td>-.447</td>
</tr>
<tr>
<td>Arousal</td>
<td>.681</td>
</tr>
<tr>
<td>Imageability</td>
<td>-.063</td>
</tr>
</tbody>
</table>

time felt subjectively faster in the presence of taboo words.

The difference between the taboo-longer and neutral-longer conditions also showed a tendency to be larger for the shorter duration pairs in Experiment 2. This is similar to the results of Experiment 1, which showed an effect of word type for the shorter durations only. This is inconsistent with arousal-based explanations for timing effects, but consistent with attention-based effects. If taboo words distract attention away from the timing task, this brief distraction time makes up a larger proportion of the display time of words shown for shorter durations. In Experiment 2, this lost time may have made it more difficult to correctly identify a taboo word as being longer, especially when it was displayed for a relatively shorter duration.

The results also supported the prediction that more taboo than neutral words would be recalled, leaving open the possibility that attention allocated to emotion-linked memory encoding processes was part of the cause of the duration underestimation. The correlational
analysis that the compared the probability of recall for each of the words in Experiment 2 with the ratings given in Janschewitz (2008) revealed the expected results, indicating that the more strongly emotional words tended to be recalled more frequently, with higher arousal ratings and negative valence ratings leading to higher rates of recall. Words that were rated as being more familiar, more offensive, or more taboo were also more likely to be recalled.

In addition to Experiment 2, our lab ran several other studies using the ordinality comparison procedure that are not reported here. Whenever any other variables were added to the experiment, we were unable to replicate the primary result that participants are significantly less likely to answer correctly in the taboo-longer condition than in the neutral-longer condition, although the means were nearly always in the predicted direction. For example, one version added confidence ratings for each response, and another added a condition asking participants to indicate which word in each pair was displayed for a shorter time, rather than a longer time. This particular result may have been difficult to replicate because the word pairs were an additional source of variance. Experiment 1 and previous literature (e.g., Hochhaus et al., 1991; Warm & McCray, 1969) have shown that various characteristics of words, including length and familiarity, can influence duration judgments. Although the taboo and neutral lists were matched on as many of these characteristics as possible, the individual word pairs were not. The temporal bisection task may be a better method to use when studying duration judgments for words because it collects responses for individual words rather than pairs, avoiding any variance due to the specific combinations of words in the pairs.

**Experiment 3: Retrospective Duration Reproduction Task**

In Experiment 3, participants were instructed to study a list of words as they appeared on the computer screen. No timing task was mentioned. Then, after viewing the seventh word, which was either neutral or taboo depending on the condition, they were asked to attempt to reproduce the duration of the final word by holding down the mouse button for
an equivalent length of time.

A pilot version of this study was designed as a modified ordinality comparison task, where participants were asked to study a list of words, but after only two words had been displayed, they were asked to determine whether the first or the second word had been displayed for longer. However, the data revealed that participants nearly always selected the first word, regardless of word emotion or actual duration. This tendency to remember the first of two time intervals as being longer than the second is called the positive time-order effect, and it is actually a common finding in retrospective timing research (Block, 2003). The longer list of words used in Experiment 3 was an attempt to avoid this bias. During the pilot experiment, participants reported that the task seemed extremely difficult, so a duration reproduction task was used so that participants would not need to recall the durations of multiple words to provide an answer.

In duration reproduction tasks, different methods have led to different results. Asking participants to press a button to indicate the start of the interval they are reproducing and press it again to indicate the end of the interval has led to the most accurate reproduction results, but asking participants to continuously hold down a button throughout the interval has shown less variable results (Mioni, Stablum, McClintock, & Grondin, 2014). In the current study, minimizing variability was of key importance, especially because only one measurement could be obtained per participant due to the surprise nature of the retrospective task. The purpose of the experiment was to compare retrospective duration reproduction judgments for taboo and neutral words, so absolute accuracy was of less importance. For this reason, the current study used the duration reproduction methods that asked participants to continuously hold down the mouse button for an interval equivalent to their memory of the duration of the final word.

The same theoretical models used to explain the underestimation of durations for taboo words in prospective timing tasks predict overestimation of durations for taboo words when they are seen in retrospective timing tasks. In prospective timing tasks, participants
tend to underestimate the duration of taboo words compared to neutral words, possibly because the taboo words capture attentional resources that would normally be directed toward the timing task. This extra attention directed toward taboo words may lead to the formation of more memories. Then, when participants attempt to make retrospective judgments based on memory, they give longer duration judgments for taboo words because they remember more.

Method

Participants. Sixty UCLA undergraduates (18–23 years, $M = 20.0$ years, $SD = 1.5$; 6 male, 54 female) received partial course credit for participating. All participants spoke English as their native language.

Materials. The materials included five neutral filler words ($pencil$, $tree$, $cow$, $water$, and $table$). In the neutral condition, the participants were asked to retrospectively reproduce the duration of the word $seat$, and in the taboo condition the participants were asked to retrospectively reproduce the duration of the word $fuck$. This pair of words was matched based on number of letters and syllables, as well as familiarity ratings given in Janschewitz (2008). All words were shown in black 60-point Arial font.

Procedure. Participants were instructed to read and study a series of words. At the beginning of the task, they were not given any instructions related to time. All participants saw the same words, but the order differed depending on condition, which was manipulated between subjects. In the neutral condition, participants saw the word $fuck$, followed by the five filler words, then $seat$. In the taboo condition, participants saw the word $seat$, followed by the five filler words, then $fuck$. The final word was always displayed for 1000 ms, while the other words varied between 800 and 1200 ms. After all seven words were displayed, instructions appeared on the computer screen asking participants to reproduce the duration
of the last word by holding down the mouse button for the same length of time that the word was shown on the screen. After completing this task, participants were asked to rate how confident they were about their response using a scale from 1 (not confident) to 5 (very confident).

Results

The mean reproduced durations for the taboo and neutral word conditions are displayed in Figure 10. An independent-samples $t$-test revealed that the reproduced durations were significantly longer in the taboo condition than in the neutral condition, $t(58) = 2.279$, $p = .026$. Reproduced durations in the taboo condition were significantly overestimated compared to the correct duration of 1000 ms, $t(29) = 2.084$, $p = .046$, while reproduced durations in the neutral condition did not differ significantly from the correct 1000 ms duration, $t(29) = 0.924$, $p = .363$. Confidence ratings in the taboo ($M = 2.93$, $SD = 0.78$) and neutral conditions ($M = 3.03$, $SD = 1.10$) did not differ significantly, $t(58) = 0.406$, $p = .406$.

Figure 10. Mean reproduced durations of taboo and neutral words in the retrospective timing task in Experiment 3. Error bars represent standard error of the mean.
Discussion

The results of Experiment 3 supported the prediction that durations of taboo words would retrospectively be judged to be longer than neutral words. On the surface this may seem to contradict the results of Experiments 1 and 2, which showed that durations of taboo words were underestimated compared to neutral words in prospective timing tasks. However, the difference between prospective and retrospective tasks is one of the most important distinctions in the time perception literature, and it leads to different predictions for the effects of attention-grabbing stimuli on duration judgments. In prospective tasks, which ask participants to intentionally attend to time, stimuli that attract attention reduce the resources available for timing, causing durations to feel subjectively shorter. In retrospective tasks, which do not require participants to attend to time, time judgments rely on memory for incidentally encoded non-temporal information that can be used in an attempt to reconstruct duration length. Experiments 1 and 2 found better recall for taboo than neutral words, and other studies have demonstrated better recall for features associated with taboo than neutral words, including font color in a color-naming task (MacKay et al., 2004), as well as incidental task-irrelevant features like screen location (MacKay & Ahmetzanov, 2005). These recall results provide evidence that extra processing and memory encoding resources are directed toward taboo words. According to the memory storage model, if a larger amount of non-temporal information is encoded during the interval being estimated, duration judgments are longer. Due to emotion-linked memory encoding processes, more contextual information is encoded for taboo than neutral words. When participants attempt to make retrospective duration judgments, more of this information is accessible for taboo than neutral words. They use the amount of recalled information as a cue when generating duration judgments, since they did not attend to time during the interval being estimated. Because participants are able to remember more contextual information associated with taboo than neutral words, they retrospectively judge the durations of taboo words as being longer.
Retrospective duration judgments require participants to rely on their memories to generate estimates of the intervals they are judging. Further evidence for the role of memory in the formation of retrospective duration judgments comes from studies of participants with memory impairments. For example, amnesic patients tend to give shorter retrospective estimates of durations than control participants (see, e.g., Perbal, Pouthas, & Van der Linden, 2000). In a retrospective duration reproduction task, amnesic patient H.M. underestimated time intervals greater than 20 s compared to control participants, but performed normally in timing tasks with shorter durations (Richards, 1973). Patients with Alzheimer’s disease also tend to underestimate durations compared to age-matched control participants (El Haj, Moroni, Samson, Fasotti, & Allain, 2013). These are exactly the results predicted by the memory storage model: If larger amounts of encoded information result in longer retrospective duration judgments, then it follows that people with difficulty encoding new memories should produce shorter retrospective duration judgments than control participants. Finally, recent research on the neural mechanisms underlying retrospective duration judgments has proposed that the hippocampus plays a role (MacDonald, 2014; MacDonald, Meck, Sakata, & Fortin, 2014), which provides more support for the idea that memory is an essential component of retrospective timing tasks.

**General Discussion**

The results from Experiments 1, 2, and 3 provided evidence suggesting that taboo words influence duration judgments. In the prospective timing tasks in Experiments 1 and 2, durations for taboo words were underestimated compared to neutral words, which supports the predictions of the attentional gate model. In the retrospective timing task in Experiment 3, durations for taboo words were overestimated compared to neutral words, which supports the predictions of the memory storage model. The results of these three studies replicated and extended the findings of Tipples (2010), expanding upon the existing literature on the effects of emotional words on the perception of time.
One interesting question that remains is why, in prospective timing tasks, the durations for taboo words are underestimated compared to neutral words, while the durations for other emotional stimuli are usually overestimated compared to neutral stimuli. There are at least a couple of possible explanations, although the answer is not yet clear. One possibility is that taboo words simply do not evoke a strong enough arousal response to cause the overestimation effects seen with pictures and sounds. The effects of pictures and sounds on duration judgments are typically described as being caused by arousal-mediated changes in the rate of the pacemaker, or internal clock. Although taboo words do receive significantly higher arousal ratings than neutral words, and have been shown to affect physiological measures of arousal like skin conductance, it is possible that this is not a strong enough emotional reaction to change the rate of the pacemaker. Another possibility concerns the amount of time spent processing a word compared to a picture or sound. When a word appears, it takes a certain amount of time to read and comprehend it, but afterward there is no need to continue consciously processing it. A picture is a much more complex stimulus, so when it appears, there is much more to process. Imagine being shown a gruesome, emotionally negative picture. For at least a few seconds, you continue to “experience” it as you view it, and you may even wish to look away or close your eyes so that it cannot continue to disturb you. This is the experience that typically results in overestimation of durations. Now imagine being shown an offensive taboo word. You read it once, and you may react to it emotionally, but you probably do not continue to react to it for the entire length of time that it is on the screen. It seems unlikely that you would read it over and over again, and you probably would not feel a strong desire to look away or close your eyes to avoid viewing the word. Subjectively, the experiences of viewing an emotional picture and reading a taboo word seem very different. It is possible that the prolonged experience of reacting to emotional pictures leads to overestimation of durations, while the quicker experience of reading a taboo word momentarily distracts attention away from the primary timing task, which leads to underestimation of durations, but does not cause the extended
increase in arousal necessary to cause overestimation. Further research is necessary to test this explanation experimentally.

An additional factor that could account for some of the differences between taboo words and other emotional stimuli involves the ways in which conditions are controlled in experiments. When using words as stimuli, it is relatively easy to control for a large number of features, such as length, familiarity, and visual appearance. Even when comparing categories of words that are very different, such as taboo and neutral words, it is still possible to control for many different features. However, when using other types of stimuli, such as pictures, controlling for these features becomes much more complicated. It is much more difficult to define and quantify factors like familiarity and complexity, both of which could plausibly influence duration judgments. Studies that make use of pictures from databases like IAPS have access to lists of norms and ratings that are associated with each picture, which certainly help control for these factors to some extent. However, it is still very difficult to control for differences between picture conditions as thoroughly as differences between word conditions. Before concluding that the effects of words on duration judgments are fundamentally different than the effects of other emotional stimuli, it is important that future studies using other emotional stimuli carefully control for differences in features like familiarity and complexity.
CHAPTER 3: NOVELTY AND DURATION JUDGMENTS

A possible alternate explanation for the effects of taboo words on duration judgments that were revealed in Chapter 2 is that they were driven by the relative novelty of taboo words in experimental contexts, rather than their emotional content. Although half of the words displayed in Experiments 1 and 2 were taboo, it is relatively unusual to see taboo words in writing, particularly in university-sponsored experiments. In that sense, their appearance in the task could be considered novel, and this novelty could potentially impact duration judgments. Experiment 4, 5, and 6 investigated this possibility by manipulating novelty independently from semantic content. These experiments introduced a novel condition in which words were displayed in various combinations of fonts, colors, and sizes. These specific combinations were never repeated for more than one word. If the effects of taboo words on duration judgments can be replicated with this novel text condition, then it would leave open the possibility that the results shown in Experiments 1, 2, and 3 may have been at least partially driven by the effects of novelty. An alternate possibility is that emotion and novelty may impact attention in similar ways, independently leading to the same effects on duration judgments.

Novel stimuli tend to attract attention (for review, see Ranganath & Rainer, 2003). As described by Ranganath and Rainer (2003), there are two different types of novelty that are frequently manipulated experimentally: stimulus novelty, which is defined as the first encounter with a specific stimulus, and contextual novelty, which is when the occurrence of a stimulus is surprising given its current context. Studies investigating stimulus novelty tend to focus on repetition. As stimuli are repeated and novelty wears off, repeated items tend to be processed more fluently and efficiently, which is sometimes accompanied by reduced neural activity (see, e.g., Henson & Rugg, 2003). Therefore, compared to non-novel repeated stimuli, novel stimuli seem to require more neural activity to process. When the appearance of a stimulus is surprising given its context, it provokes an orienting response,
which causes attentional resources to be directed toward the surprising stimulus (see, e.g., Corbetta & Shulman, 2002). Both types of novel stimuli require extra attentional resources to process. Although taboo words may be an example of contextual novelty, and the novel text manipulation used in Experiments 4, 5, and 6 is an example of stimulus novelty, both have the potential to distract attentional resources away from concurrent timing tasks.

Much of the research on the effects of novelty on the perception of time has made use of the oddball paradigm. In the oddball task, participants view a series of repeated items with a few different, or “oddball”, items interspersed throughout. The repeated, or standard, items are usually shown for identical durations, while the oddball items are displayed for slightly different durations. In this task, participants are asked to decide whether the oddball items are displayed for a longer or shorter duration than the standard items. Nearly all of the previous studies using the oddball task have used simple visual stimuli, and none have used words. In one of the original studies using the oddball paradigm, Tse et al. (2004) showed participants a series of black circles, with static circles as the standard condition, and circles that were expanding in size as the oddball condition. Participants consistently overestimated the durations of the oddball stimuli, responding that the oddballs were longer than the standards even when they were considerably shorter. Tse et al. explained these findings by suggesting an attentional basis for this effect, claiming that the rate of processing increased when attention was engaged by the oddball stimuli. The offered explanation resembled the arousal-based explanations for timing effects reviewed in Chapter 1, which claim that the speed of the internal clock is increased by high-arousal stimuli, resulting in longer duration judgments. In fact, other researchers have suggested arousal as an explanation for the effects of novelty on duration judgments (see, e.g., Ulrich, Nitschke, & Rammsayer, 2006). In a similar experiment, Wittmann, van Wassenhove, Craig, and Paulus (2010) tested the effects of expanding, contracting, and static circles as oddball stimuli. They found overestimated durations for the expanding oddballs, but not for the contracting or static oddball stimuli.

Other studies have attempted to explain the oddball effect based on increased coding
efficiency for repeated items, rather than attentional or arousal mechanisms (for reviews, see Eagleman, 2010; Eagleman & Pariyadath, 2009). The explanation offered by Eagleman and Pariyadath (2009) is that repeated presentations of the same stimulus create a more efficiently coded neural representation of that stimulus. This is achieved through repetition suppression, in which the neural response to repeated stimuli decreases with repetition. Eagleman and Pariyadath proposed an energy readout model in which durations are judged based on the amount of neural energy used when representing a stimulus, with more neural energy correlating with longer perceived durations. Therefore, the higher coding efficiency that occurs with repetition is associated with shorter duration judgments. When an oddball stimulus appears after a series of repeated standards, it requires relatively more neural energy to represent, so its duration is overestimated compared to the standards. This is supported by the findings from the oddball effect, and also Rose and Summers (1995), which found that a repeated stimulus seems subjectively longer the first time it is displayed than it does on subsequent repetitions. In another study, Pariyadath and Eagleman (2007) attempted to rule out attentional explanations for the oddball effect by including emotional pictures as oddballs. These pictures, selected from the IAPS database, included pictures of spiders, sharks, and weapons. Pariyadath and Eagleman did not find any evidence for an increased size of the oddball effect in this emotional condition, and they interpreted this to mean that the oddball effect is not caused by increased attention to oddball stimuli.

Experiments 4 and 5 used the temporal bisection task with words displayed in standard and novel text styles. To facilitate direct comparison with Experiment 1, half of the words in Experiment 4 were displayed in novel text styles, while the remaining half were displayed in the standard text style. Next, in order to make a more direct comparison with the studies looking at novelty in the oddball effect literature, the number of words displayed in novel text styles was reduced in Experiment 5. In this “oddball” condition, the words displayed in novel text styles appeared relatively rarely in the experiment. Finally, Experiment 6 used the oddball task. In this task, participants provided duration judgments for
the oddball stimuli only, rather than for all stimuli as in the temporal bisection task. Because previous studies in the oddball literature have used visual, rather than verbal, stimuli, Experiment 6 compared the effects of oddball words and simple visual shapes.

**Experiment 4: Temporal Bisection Task with Novel Text Styles**

Experiment 4 was identical to Experiment 1, except that the taboo words were replaced with neutral words appearing in novel combinations of fonts, sizes, and colors. In the “novel” condition, the specific combinations of fonts, sizes, and colors were never repeated for more than one word. Participants completed a temporal bisection task comparing words displayed in the novel condition, as well as a standard condition in which all of the words were displayed in black 60-point Arial font. The novel words were designed to require more processing resources due to the novel combinations of text styles in which they were displayed. The comparison between novel and standard font conditions was used to test the hypothesis that the additional attentional requirements of the novel condition would cause underestimation of the durations of words displayed in novel compared to standard text styles. If the results for the novel font condition resemble the results for the taboo words in Experiment 1, then novelty could be a possible explanation for the shorter duration judgments for taboo than neutral words.

If the effects of taboo words on duration judgments are caused by their ability to attract attention, then it seems plausible that other manipulations that cause words to attract attention could also result in the same underestimation of durations. If this is the case, then the various combinations of fonts, colors, and sizes should cause the durations of words in the novel condition to be underestimated compared to the words in the standard condition. This should result in a lower proportion of “long” responses for novel words in the temporal bisection task, as well as a later bisection point for novel than standard words. If the effect of novelty on duration judgments is driven by attentional mechanisms, and not changes in the rate of the internal clock, it should be larger for shorter durations than for
longer durations.

The effects of the font manipulation on recall are more difficult to predict. If the novel words stand out more, then they might also be more likely to be recalled, as in the classic von Restorff effect (von Restorff, 1933). However, because half of the words were displayed in the novel font condition, individual words might not stand out enough to improve recall. The purpose of including them in the task was to attract attention, causing participants to direct more resources toward processing the word rather than attending to the timing task. However, in this case those attentional resources will not necessarily be directed toward encoding the words, but rather the novel combination of the font, color, and size. In that case, the words in the novel condition may not be recalled at a higher rate than the words in the standard condition.

Method

Participants. Forty-two UCLA undergraduates (18–24 years, $M = 19.7$ years, $SD = 1.44$; 13 male, 29 female) received partial course credit for participating. All participants spoke English as their native language.

Materials. The materials consisted of 182 neutral words from Janschewitz (2008). Half of these words were assigned to the standard condition and half were assigned to the novel condition. In the standard condition, all words were displayed in black 60-point Arial font. In the novel condition, the appearance of the text was manipulated by varying the font, color, and size of the text. No combination of these three features was repeated. The colors used were red, blue, green, brown, pink, teal, and purple. The fonts used were Baskerville, Comic Sans, Courier, Lucida Blackletter, and Lucida Calligraphy. The sizes were 20-, 40-, 80-, and 100-point font. The text styles used in the standard and novel conditions are displayed in Figure 11.
Standard condition: chair

Novel condition: chair chair chair chair chair
chair chair chair chair chair
chair chair chair chair chair

Figure 11. Text styles for the standard and novel conditions in Experiment 4. In the standard condition, words were shown in black 60-point Arial font. In the novel condition, words were shown in various combinations of the fonts, colors, and sizes depicted above. Note that the sizes shown here are relative, and all were large enough to be easily visible when displayed on the computer screen.

Procedure. The procedures for the temporal bisection task and the surprise free recall test were identical to Experiment 1.

Results

Figure 12 shows the mean proportion of “long” responses for words in the standard and novel font conditions for each of the seven durations. A 2 (condition: standard vs. novel) × 7 (duration) repeated measures ANOVA revealed a significant main effect of condition, indicating more “long” responses for words in the standard condition than in the novel condition, \( F(1,41) = 6.226, \eta_p^2 = .132, p = .017 \). There was also a significant main effect of duration, with more “long” responses as durations increased, \( F(6,246) = 473.839, \eta_p^2 = .920, p < .001 \). There was no significant interaction, \( F(6,246) = 1.414, \eta_p^2 = .033, p = .210 \).

This analysis was repeated without the practiced “short” and “long” durations, re-
The 2 (condition: standard vs. novel) \times 5 (duration) repeated measures ANOVA revealed a significant main effect of condition indicating more “long” responses for standard than novel words, $F(1,41) = 7.712, \eta_p^2 = .158, p = .008$. There was a significant main effect of duration, $F(4,164) = 355.274, \eta_p^2 = .897, p < .001$, but no significant interaction, $F(4,164) = 1.239, \eta_p^2 = .029, p = .296$.

To test internal clock speed predictions, the effect of condition was compared for the shorter half of durations used in the experiment (i.e., 400, 600, and 800 ms) versus the longer half of durations used in the experiment (i.e., 1200, 1400, and 1600 ms). When averaging across the shorter half of durations, the mean proportion of “long” responses was significantly higher in the standard condition ($M = 0.14, SD = .10$) than in the novel condition ($M = 0.11, SD = .10$), $t(41) = 2.574, p = .014$. However, when averaging
across the longer half of duration, the mean proportion of “long” responses did not differ significantly for the standard condition ($M = 0.88$, $SD = .13$) and the novel condition ($M = 0.88$, $SD = .13$), $t(41) = 0.467$, $p = .643$. This suggests that the effect of condition was driven by the differences at shorter durations.

The 1000 ms duration was of particular interest because it is equally close to the “short” and “long” response durations. Overall, participants showed a bias toward responding “long” to this duration, with a mean proportion of 0.58 ($SD = 0.22$) “long” responses, which was significantly greater than the chance proportion of 0.50, $t(41) = 2.480$, $p = .001$. However, they were not significantly more likely to respond “long” for standard words ($M = 0.60$, $SD = 0.24$) than for novel words ($M = 0.57$, $SD = 0.24$), $t(41) = 1.164$, $p = .251$.

Summing across all duration conditions, the mean proportion of “long” responses was 0.52 ($SD = .11$) for words in the standard condition and 0.51 ($SD = .11$) for words in the novel condition. One-sample $t$-tests were used to compare these values to the expected proportion of 0.50. The overall proportion of “long” responses did not differ significantly from 0.50 for standard, $t(41) = 1.417$, $p = .164$, or novel words, $t(41) = 0.450$, $p = .655$.

In the novel condition, separate repeated measures ANOVAs were performed to test for the effects of font, color, and size on the proportion of “long” responses. Because this experiment did not include all combinations of these conditions, they could not be combined in a single analysis. There were no significant effects of font, $F(4, 164) = 0.401$, $\eta_p^2 = .010$, $p = .807$, color, $F(6, 246) = 0.559$, $\eta_p^2 = .013$, $p = .762$, or size, $F(3, 123) = 0.557$, $\eta_p^2 = .013$, $p = .644$.

The bisection points were calculated for both standard and novel words for each participant. The mean bisection point was 999 ms ($SD = 166$) for novel words and 973 ms ($SD = 167$) for standard words. A paired $t$-test revealed that the bisection point was reached significantly later for novel than for standard words, $t(41) = 2.515$, $p = .016$. This suggests that, on average, participants perceived the durations of novel words to be approximately
26 ms shorter than the standard words.

The mean proportion of “long” responses was calculated by item and compared with the ratings given in Janschewitz (2008). This revealed a significant negative correlation between the mean proportion of “long” responses and number of letters. The mean proportion of “long” responses was not significantly correlated with number of syllables, or ratings of personal use, familiarity, or imageability. Correlation coefficients and $p$-values are displayed in Table 6.

Table 6

<table>
<thead>
<tr>
<th></th>
<th>Mean Proportion of “Long” Responses</th>
<th>Probability of Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation coefficient</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td></td>
<td>$r(180)$ $p$-value</td>
<td>$r(180)$ $p$-value</td>
</tr>
<tr>
<td>Number of Letters</td>
<td>-.164 $.027^*$</td>
<td>-.049 .513</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>.071 .343</td>
<td>-.013 .858</td>
</tr>
<tr>
<td>Personal Use</td>
<td>.093 .214</td>
<td>.180 .015*</td>
</tr>
<tr>
<td>Familiarity</td>
<td>.117 .115</td>
<td>.170 .022*</td>
</tr>
<tr>
<td>Imageability</td>
<td>.143 .054</td>
<td>.237 .001*</td>
</tr>
</tbody>
</table>

Figure 13 shows the mean numbers of novel and standard words recalled in the surprise free recall test. There was a trend suggesting that participants recalled more standard ($M = 3.0, SD = 1.7$) than novel words ($M = 2.5, SD = 1.8$), $t(41) = 1.961, p = .057$. Rate of recall was not influenced by display duration, $F(6,246) = 1.522, \eta^2_p = .036, p = .172$. For each word, probability of recall was compared to the ratings from Janschewitz (2008), with correlation coefficients and $p$-values displayed in Table 6. Recall was positively correlated with ratings of personal use, familiarity, and imageability. There were no correlations between recall and number of letters or syllables. Separate repeated measures ANOVAs were
performed to test for the effects of font, color, and size on recall. There were no significant
effects of font, $F(4, 164) = 2.048, \eta^2_p = .048, p = .090$, color, $F(6, 246) = 1.439, \eta^2_p = .034,$
$p = .200$, or size, $F(3, 123) = 0.753, \eta^2_p = .018, p = .523$.

**Figure 13.** Mean number of words recalled in the standard and novel conditions in the
surprise free recall test in Experiment 4. Error bars represent standard error of the mean.

**Discussion**

The results from Experiment 4 supported the prediction that durations for words in
the novel condition would be underestimated compared to words in the standard condition.
The results for the novel words were comparable to the results from the taboo words in
Experiment 1. In the temporal bisection task, participants responded “long” significantly
less frequently for novel than for standard words. The overall proportions of “long” responses
for both conditions did not differ significantly from 0.50, however, so there was no evidence
that timing performance was more accurate for novel than standard words. The bisection
point occurred significantly later for novel than standard words, indicating that durations
for words in the novel condition were underestimated compared to words in the standard
condition by approximately 26 ms on average. The bisection point was slightly lower for the
novel words in Experiment 4 (999 ms) than for the taboo words in Experiment 1 (1010 ms), but the bisection points for standard words in Experiment 4 (973 ms) and the neutral words in Experiment 1 (971 ms) were comparable between experiments. This suggests that the effect of novel text styles on duration judgments may be slightly smaller than the effect of taboo words.

As in Experiment 1, the effect of condition on duration judgments was significant for the shorter half of displayed durations, but not the longer half. This is inconsistent with arousal-based explanations that suggest that timing effects are caused by changes in the rate of the internal clock. Instead, this result is consistent with attention-based explanation. Viewing words displayed in novel text combinations distracts from the timing task and slightly reduces the amount of time counted by the accumulator. If the time is reduced by a constant amount across display durations, the effect will primarily be visible at shorter display durations, where the lost time makes up a larger proportion of the total.

In Experiment 4, there was a negative correlation between the number of letters in a word and its mean proportion of “long” responses. Although this is the opposite of what was found in Experiment 1, it supports the findings in other literature (Warm & McCray, 1969). It makes sense that durations for words with a greater number of letters would be more likely to be underestimated because longer words take more time and processing resources to read. This reduces resources available for the timing task, resulting in underestimation. The correlational analysis also compared the probability of recall for each word with the ratings from Janschewitz (2008), showing that recall was positively correlated with ratings of personal use, familiarity, and imageability. These correlations are as expected, since more familiar and high frequency words are typically easier to remember, as are highly imageable words. The effect of novelty on free recall was not significant, but there was a trend suggesting that more words were recalled from the standard condition than from the novel condition. It is not entirely clear why this occurred, but one possibility is that the novel text combinations made it more difficult for participants to encode the words into memory.
Experiment 5: Temporal Bisection Task with Oddball Condition

The results from Experiment 4 were a bit surprising given the previous literature on the oddball effect, which typically shows that durations for stimuli that stand out tend to be overestimated compared to standard stimuli. Experiment 5 was designed to be very similar to Experiment 4, except that the novel words appeared much less frequently. This was intended to create an experiment that more closely resembled other studies looking at the oddball effect. In Experiment 4, half of the words appeared in novel text styles, while in Experiment 5, only one out of every 12 words appeared in a novel text style. In Experiment 5, these “oddball” words were novel in the sense that they appeared in unique combinations of font, color, and sizes, and also in the sense that they were relatively rare in the context of the experiment. Just as in Experiment 4, the standard words always appeared in black 60-point Arial font.

Based on the same explanations given for Experiment 4, it was predicted that the durations for the oddball words would be underestimated compared to the standard words. Because they were more rare in Experiment 5, it seems plausible that they could be even more effective at attracting attention, leading to underestimation of durations. An alternate prediction would be that durations of the oddball words would be overestimated, just like the simple, typically abstract, visual stimuli used in the standard oddball paradigm. However, this effect would be difficult to explain using the same attentional mechanisms suggested for Experiment 4. Again, the free recall results are more difficult to predict. Because the oddball words were relatively rare, they stood out much more than the novel words in Experiment 4. For this reason, it is more plausible that the von Restorff effect would occur, causing a larger percentage of oddball words to be recalled than standard words (von Restorff, 1933).
Method

Participants. Thirty-nine UCLA undergraduates (18–23 years, $M = 20.0$ years, $SD = 1.2$; 11 male, 28 female) received partial course credit for participating. All participants spoke English as their native language.

Materials. The materials consisted of 182 neutral words from Janschewitz (2008). Fourteen words were assigned to the novel “oddball” condition and all of the remaining words were assigned to the standard condition. All words in the standard condition were displayed in black 60-point Arial font. The appearance of the words in the oddball condition was manipulated by varying the font, color, and size of the text as described in Experiment 4. Assignment of words to the oddball condition was counterbalanced between subjects.

Procedure. The procedure was very similar to Experiment 4. In order to make the appearance of words in the oddball condition less predictable, they were separated by sequences of 9–15 standard words in a row, with a mean of 12 standards occurring between oddballs. The procedure was otherwise identical to the temporal bisection task and surprise free recall test described in Experiment 1.

Results

Figure 14 shows the mean proportion of “long” responses for words in the standard and oddball font conditions for each of the seven durations. A 2 (condition: standard vs. oddball) $\times$ 7 (duration) repeated measures ANOVA revealed a weak trend in the direction of more “long” responses for words in the standard condition than in the oddball condition, $F(1, 38) = 2.816$, $\eta^2_p = .069$, $p = .102$. There was a significant main effect of duration, with more “long” responses as durations increased, $F(6, 228) = 172.890$, $\eta^2_p = .820$, $p < .001$, and a significant interaction between condition and duration, $F(6, 228) = 3.141$, $\eta^2_p = .076$, $p = .006$. In order to explain this interaction, paired $t$-tests compared the mean proportions
of “long” responses for the standard and oddball conditions in each of the seven duration conditions. This revealed that there were significantly more “long” responses in the standard than the oddball condition at 600 ms, $t(38) = 3.875, p < .001$, and 800 ms, $t(38) = 3.533, p = .001$, while there were significantly more “long” responses in the oddball than the standard condition at 1600 ms, $t(38) = 3.853, p < .001$. There were no significant differences at 400 ms, $t(38) = 1.172, p = .248$, 1000 ms $t(38) = 0.195, p = .847$, 1200 ms, $t(38) = 0.196, p = .846$, or 1400 ms, $t(38) = 0.337, p = .738$.

![Figure 14](image)

*Figure 14.* Mean proportion of “long” responses by duration for the standard and oddball conditions in Experiment 5. Error bars represent standard error of the mean.

This analysis was repeated without the practiced “short” and “long” durations, revealing a similar pattern of results. The 2 (condition: standard vs. oddball) $\times$ 5 (duration) repeated measures ANOVA revealed a significant main effect of condition indicating more “long” responses for standard than oddball words, $F(1, 38) = 4.243, \eta_p^2 = .100, p = .046$. 
There was a significant main effect of duration, $F(4, 152) = 105.069$, $\eta^2_p = .734$, $p < .001$, as well as a significant interaction, $F(4, 152) = 2.986$, $\eta^2_p = .073$, $p = .021$, with a larger proportion of “long” responses for the standard than oddball condition at 600 and 800 ms, but not at 1000, 1200, or 1400 ms, as described above.

To test internal clock speed predictions, the effect of condition was compared for the shorter half of durations used in the experiment (i.e., 400, 600, and 800 ms) versus the longer half of durations used in the experiment (i.e., 1200, 1400, and 1600 ms). When averaging across the shorter half of durations, the mean proportion of “long” responses was significantly higher in the standard condition ($M = 0.19$, $SD = .09$) than in the oddball condition ($M = 0.09$, $SD = .14$), $t(38) = 5.089$, $p < .001$. However, when averaging across the longer half of duration, the mean proportion of “long” responses did not differ significantly for the standard condition ($M = 0.79$, $SD = .11$) and the oddball condition ($M = 0.81$, $SD = .19$), $t(38) = 0.784$, $p = .434$. This suggests that the effect of condition was driven by the differences at shorter durations.

Summing across all duration conditions, the mean proportion of “long” responses was 0.49 ($SD = .08$) in the standard condition and 0.45 ($SD = .14$) in the oddball condition. One-sample $t$-tests were used to compare these values to the expected proportion of 0.50. Participants chose “long” for the oddball words at a rate significantly lower than 0.50, $t(38) = 2.109$, $p = .042$, while the proportion of “long” responses for the standard words did not differ significantly from 0.50, $t(38) = 1.122$, $p = .269$. This suggests that durations for the oddball words were underestimated, but durations of standard words did not seem to be over- or underestimated.

The bisection points were calculated for both standard and oddball words for each participant. The mean bisection point was 1068 ms ($SD = 227$) for novel words and 1033 ms ($SD = 131$) for standard words. A paired $t$-test revealed no significant effect of condition on bisection points, $t(38) = 1.119$, $p = .270$. The mean proportion of “long” responses was calculated by item and compared with the ratings given in Janschewitz (2008).
were no significant correlations between the mean proportion of “long” responses and num-
ber of letters, number of syllables, or ratings of personal use, familiarity, or imageability. Correlation coefficients and $p$-values are displayed in Table 7.

Table 7
*Correlations Between Experiment 5 Timing and Recall Results and Word Ratings from Janschewitz (2008)*

<table>
<thead>
<tr>
<th></th>
<th>Mean Proportion of “Long” Responses</th>
<th>Probability of Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation coefficient</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td></td>
<td>$r$ (180)</td>
<td>$p$-value</td>
</tr>
<tr>
<td>Number of Letters</td>
<td>.013</td>
<td>.859</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>.014</td>
<td>.848</td>
</tr>
<tr>
<td>Personal Use</td>
<td>-.051</td>
<td>.398</td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.108</td>
<td>.495</td>
</tr>
<tr>
<td>Imageability</td>
<td>-.103</td>
<td>.081</td>
</tr>
</tbody>
</table>

Because there were only 14 words in the oddball condition versus 168 in the standard condition, the free recall results were analyzed using the percentage of words recalled in each condition, rather than the raw numbers. Figure 15 shows the mean percentage of words recalled in the standard and oddball conditions. A significantly larger percentage of words were recalled in the oddball condition than in the standard condition, $t(38) = 4.478$, $p < .001$. Rate of recall was not influenced by display duration, $F(6, 228) = 0.726$, $\eta^2_p = .019$, $p = .629$. When comparing the raw data, participants recalled significantly more words in the standard condition ($M = 2.64$, $SD = 2.15$) than in the oddball condition ($M = 1.03$, $SD = 1.14$) words in the oddball condition, $t(38) = 4.428$, $p < .001$. For each word, probability of recall was compared to the ratings from Janschewitz (2008). Correlation coefficients and $p$-values are displayed in Table 7. Recall was not correlated with number of letters, number of syllables, or ratings of personal use, familiarity, or imageability.
Discussion

Unlike in Experiment 4, there was not a strong overall effect of condition on the mean proportion of “long” responses in Experiment 5. If anything, there was only a weak trend toward fewer “long” responses for words in the oddball condition than in the standard condition. When the learned time intervals, 400 and 1600 ms, were removed from the analysis, however, there was a significant effect of condition, with oddball stimuli being relatively underestimated compared to standard stimuli. There was no significant difference between the bisection points for oddball and standard words, perhaps due to increased variance in the oddball condition, since it was calculated based on responses from a smaller number of trials.

As in Experiments 1 and 4, the effect of condition on duration judgments in Experiment 5 was significant for the shorter displayed durations, but not the longer durations. This was apparent in the analysis comparing the shorter and longer halves of the range of durations used in the experiment, and also in the significant interaction between condition and duration. The difference in the proportion of “long” responses was fairly large for the
lower durations, but disappeared almost entirely at the longer durations, in a pattern that appeared much more extreme than in Experiments 1 and 4. Compared to the results of the other experiments, the shape of the function for responses in the standard condition was rather unique (see Figure 14). The function appeared relatively linear compared to the sigmoidal shapes of all of the other conditions. This pattern suggests that responses in the standard may have been generally less accurate than in other conditions. The most correct pattern of responses would not be linear at all, but would be “short” for all durations from 400-800 ms, half “short” and half “long” at 1000 ms, and “long” for all durations from 1200–1600 ms. It is not entirely clear why the standard condition in this experiment would show different results than the others, but it may have to do with the larger numbers of consecutive trials from the same condition, which ranged from 9–15 with a mean of 12. In Experiments 1 and 4, the neutral and standard conditions were randomly interspersed with the potentially more interesting taboo and novel conditions, respectively. It is possible that timing accuracy worsened for the standard condition in Experiment 5 due to lack of interest and fatigue during the trials between the more interesting and, by design, attention-grabbing oddball words.

Experiment 5 was also the only experiment in which durations were actually underestimated compared to the absolute durations of the displayed stimuli. Participants chose “long” in response to words in the oddball condition at a rate significantly lower than the expected 0.50, while the proportion of “long” responses for words in the standard condition did not differ significantly from 0.50. So, not only were the durations for the oddball words underestimated compared to the standard condition, but they were underestimated in absolute terms as well. In Experiment 4, the proportion of “long” responses for words in both the novel and standard conditions did not significantly differ from 0.50. In Experiment 1, proportion of “long” responses for neutral words was actually higher than 0.50, but it did not differ from 0.50 for taboo words. The durations for words in the novel condition in Experiment 4 and the taboo words in Experiment 1 were underestimated compared to the
control conditions in their respective experiments, but were not underestimated in absolute
terms. It is possible that the durations for words in the oddball condition were underesti-
mated even more due to the combined effects of the novel text styles and relative rarity in
the experiment, but more research would be necessary to fully separate out the effects.

Participants recalled a significantly larger percentage of words from the oddball con-
dition than from the standard condition. This suggests that the words that were distinct
from the others were easier to recall, as in the typical von Restorff effect (von Restorff, 1933).
This finding contrasts with the results from Experiment 4, which showed a trend suggesting
that the words displayed in novel text styles were less likely to be remembered than the stan-
dard words. The only difference between the two experiments was the relatively frequency
of words in the novel text condition compared to the standard condition: Half of the words
were in the novel condition in Experiment 4, while only one out of every 12 were in the
oddball condition in Experiment 5. So why would this difference reverse the pattern of free
recall results? One possibility is that the relative rarity of oddball words in Experiment 5
made them stand out more than the novel words in Experiment 4, while the more common
novel words in Experiment 4 did not have this benefit.

Probability of recall did not correlate with number of letters or syllables, or with any
of the ratings from Janschewitz (2008), unlike in Experiment 4, which showed significant
positive correlations with familiarity, frequency of use, and imageability. This lack of cor-
relations was likely due to the strong effect of condition on recall in Experiment 5, which
improved recall for all words in the oddball condition, regardless of their length or ratings.

Despite the differences discussed above, the overall patterns of results from the tem-
poral bisection tasks in Experiments 4 and 5 were similar: Durations for words displayed in
novel text styles were more likely to be underestimated, especially when displayed for shorter
durations. This contrasts with results from other studies using the oddball paradigm, which
typically show that durations for stimuli that perceptually stand out tend to be overes-
timated, rather than underestimated. However, none of those other studies used words as
stimuli. It is possible that words and other visual stimuli are processed in different ways that could account for these opposite effects. Words displayed in novel text styles require more resources to encode, which directs attention away from time, resulting in underestimation of durations. This appears to be the case both when the novel text styles are common and rare within the context of the task. It is possible that an oddball shape, such as a red circle after a series of black circles, would not require more resources to process, and therefore would not result in the same duration effects found with words. Experiment 6 explored this possibility in more detail, comparing the effects of words and simple shapes using the oddball task.

**Experiment 6: Oddball Task Comparing Words and Shapes**

Experiment 6 used the oddball task instead of the temporal bisection task. In the oddball task, participants view several standard items that are shown for the same duration, followed by an oddball stimulus that is visually different in some way and shown for a slightly different duration. The participants are then asked to respond by deciding whether the oddball was displayed for a shorter or a longer duration than the standards. Unlike in the temporal bisection task, they only make duration judgments for the oddball stimuli, rather than for each item that is displayed. Typically, previous studies have shown that durations of oddball stimuli are overestimated. However, most of those studies have used simple visual stimuli (e.g., shapes in different sizes and colors), and none have used words. In an attempt to understand the differences between previous studies using the oddball paradigm and the results from Experiments 4 and 5, Experiment 6 compared the effects of words and shapes in the oddball task. Participants viewed sets of 9–15 standard words or shapes, followed by a visually different word or shape, which was the “oddball”. The standard words were displayed in black 60-point Arial font, and the standard shapes were medium-sized black circles. The oddball words varied in font, color, and size, while the oddball shapes were displayed in varying sizes and colors.

Based on previous literature, it was predicted that durations of the shapes in the odd-
ball condition would be overestimated. However, based on the results of Experiments 4 and 5, it was predicted that durations of the words in the oddball condition would be underestimated. If the results reveal no significant difference between the word and shape conditions, it will suggest that the results from Experiments 4 and 5 differed from other oddball experiments due to the use of the temporal bisection task, rather than the oddball task. In contrast, if significant differences between the word and shape conditions are discovered, it will suggest that the different results are caused by the stimuli, with word oddballs processed fundamentally differently than shape oddballs.

Method

Participants. Sixty UCLA undergraduates (18–23 years, $M = 20.0$ years, $SD = 1.5$; 6 male, 53 female) received partial course credit for participating. All participants spoke English as their native language.

Materials. The materials consisted of 112 neutral words and 112 circles. Seven words and seven circles were assigned to be the oddballs. The oddball words were displayed in novel combinations of different fonts, sizes, and colors, while the oddball circles were displayed in varying colors and sizes. The standard words were shown in black 60-point Arial font, and the standard circles were black and approximately 7 cm in diameter. The text styles for the oddball words varied as described in Experiment 4. The oddball circles appeared in red, blue, green, brown, pink, teal, and purple, and ranged in size from 4–11 cm in diameter.

Procedure. Standard words and shapes were displayed for 1000 ms each, with interstimulus intervals varying in range of 1–3 seconds. One oddball word and one oddball shape was assigned to each of the following durations: 400, 600, 800, 1000, 1200, 1400, and 1600 ms. In order to make the appearance of the oddballs unpredictable, they were
separated by 9–15 consecutive standards, with a mean of 12 standards occurring between oddballs. There were 14 sequences of standards that were each followed by an oddball, with seven sequences made up of words and seven sequences made up of shapes. The order of sequences was randomized. After each oddball appeared, participants were asked to decide whether it was shown for a longer or shorter duration than the standards. They entered their response by pressing either a key labeled “longer” or a key labeled “shorter”. After the key press was recorded, the next sequence of items began immediately.

Results

The mean proportions of “longer” responses for the word and shape oddball stimuli are shown in Figure 16. A 2 (condition: shape vs. word) × 7 (duration) repeated measures ANOVA revealed a significant main effect of condition, with a larger proportion of “longer” responses being given in response to shape oddballs than to word oddballs, $F(1, 59) = 6.057$, $\eta_p^2 = .093$, $p = .017$. There was a significant main effect of duration, with more “longer” responses given in response to longer durations, $F(6, 354) = 201.464$, $\eta_p^2 = .773$, $p < .001$. There was also a significant interaction between condition and duration, $F(6, 354) = 5.008$, $\eta_p^2 = .078$, $p < .001$. In order to investigate this interaction, paired $t$-tests compared the mean proportion of “longer” responses for shapes versus words in each of the seven duration conditions. This revealed that the interaction was driven by significantly more “longer” responses in the shape than the word condition at 800 ms, $t(59) = 3.639$, $p = .001$, and at 1400 ms, $t(59) = 3.291$, $p = .002$. There were no significant differences at 400 ms (every participant responded “shorter” for every trial in both conditions), 600 ms, $t(59) = 1.000$, $p = .321$, 1000 ms, $t(59) = 1.529$, $p = .132$, 1200 ms, $t(59) = 1.524$, $p = .133$, or 1600 ms, $t(59) = 0.574$, $p = .568$.

Summing across all duration conditions, the mean proportion of “longer” responses was 0.53 ($SD = .11$) in the shape condition and 0.49 ($SD = .11$) in the word condition. One-sample $t$-tests were used to compare these values to the expected proportion of 0.50.
Participants chose “longer” for the oddball shapes at a rate significantly greater than 0.50, \( t(59) = 2.143, p = .036 \), while the proportion of “longer” responses for oddball words did not differ significantly from 0.50, \( t(59) = 0.778, p = .440 \). This suggests that durations for oddball shapes were overestimated, but that durations of oddball words were not over- or underestimated.

The bisection points were calculated for both words and shapes for each participant. The mean bisection point was 1033 ms \((SD = 225)\) for words and 954 ms \((SD = 170)\) for shapes. A paired \( t \)-test revealed that the bisection point was reached significantly later for words than for shapes, \( t(59) = 2.594, p = .012 \). This suggests that, on average, participants perceived the durations of oddball shapes to be approximately 78 ms longer than oddball words. The bisection point for shapes was significantly lower than the duration of
the standards, \( t(59) = 2.070, p = .043 \), suggesting that durations for oddball shapes were overestimated. The bisection point for words did not differ significantly from the duration of the standards, \( t(59) = 1.119, p = .268 \).

**Discussion**

The results from Experiment 6 indicated that durations of shape oddballs were judged to be longer than durations of word oddballs. This difference suggests that the results of Experiments 4 and 5, which were different from the results of typical oddball studies, may have been caused by the use of words as stimuli, rather than simple shapes. The differences probably were not driven by the use of the temporal bisection task in place of the oddball task, but rather by the use of words. Using shapes, Experiment 6 replicated the oddball effect by demonstrating that the durations of oddballs were overestimated. However, when using words as stimuli, the durations of oddballs were neither overestimated nor underestimated. The conclusions would have been strengthened if the results had replicated the underestimated durations of oddball words, as in Experiment 5, but perhaps some of the difference was due to the use of the oddball task instead of the temporal bisection task.

The attentional mechanisms that cause durations for words with novel appearances to be underestimated may not affect duration estimates for shapes in the same way. Novel words attract attention, and when attention is diverted away from time, durations tend to be underestimated. The extra time and effort it takes to read and process a novel word reduces the cognitive resources available for completing the timing task. Shapes, however, are far less complex. An oddball shape might grab attention, or even increase arousal as suggested by some theories. Because shapes, and circles in particular, are so much simpler than words, novel shapes may not take any longer to process than standard shapes. In other words, it might take longer to read a word in an unusual font than in a standard font, but it may not take any longer to perceive a red circle than a black circle. This could potentially explain why durations for oddball words are underestimated compared to oddball shapes.
A potential limitation of Experiment 6 comes from the operationalization of the word and shape conditions. In the word condition, participants viewed lists of different words, with the oddball being defined as the one word at the end of the list that was displayed in a non-standard text style. In the shape condition, participants viewed the same circle repeated several times, with the oddball being defined as a circle displayed in a different size and color than the standard circle. This design was chosen in order to directly compare the results of Experiments 4 and 5 with the results of previous studies using the oddball task. Words were never repeated in Experiments 4 and 5, while studies using the oddball task have typically shown the same stimulus multiple times. The goal of Experiment 6 was to directly compare these conditions, as defined by previously existing paradigms, using the oddball task. However, this caused the word and shape conditions to be confounded with repetition, with the word condition consisting of lists of different words, and the shape condition consisting of sequences of the same shape repeated multiple times. Therefore, it is possible that the differences between the results for the word and shape conditions were not necessarily due to stimulus type, but rather to repetition effects. In an attempt to rule out this potential problem, data for another variation of this study are currently being collected. In the new version, the word condition consists of sequences of the same word repeated multiple times in a standard text style, followed by a final oddball repetition of that word displayed in a novel text style. In the shape condition, each list consists of multiple repetitions of the same shape, followed by an oddball displaying that shape in a different size and color. Each of the seven word lists are made up of repetitions of different words, while each of the seven shape lists are made up of repetition of different shapes. If the new experiment replicates the results of Experiment 6, it will be reasonable to conclude that the word and shape oddballs are processed differently. However, if the new experiment does not replicate Experiment 6, then it will support the alternate possibility that repetition effects caused the differences between conditions in Experiment 6.
General Discussion

The results of Experiments 4 and 5 showed that the effects of novelty on duration judgments are comparable to the effects of taboo words, with both leading to the underestimation of durations compared to control conditions. This leaves open the possibility that the effects of taboo words on duration judgments are driven, at least in part, by novelty. Even though taboo words occurred frequently in the task in the Experiment 1, and therefore were low in stimulus novelty (as defined by Ranganath & Rainer, 2003), they are still unusual in university-sponsored experiments, so they were high in contextual novelty. This response to contextual novelty could have directed attention away from the timing task, leading to the shorter duration judgments for taboo than neutral words that were shown in Experiment 1. However, it is also possible that the effects of taboo words on duration judgments are not due to novelty. Emotion and novelty could influence attentional processing in similar ways, which would result in similar influences on duration judgments. If emotional and novel stimuli are both inherently attention-grabbing, then both could be effective at distracting attention away from a concurrent timing task. This would cause the durations for both types of stimuli to be underestimated. Further research is needed to distinguish between these possibilities empirically.

In Chapter 2, the results suggested that the underestimation of durations of taboo words relative to neutral words could have been caused by emotion-linked memory encoding processes. However, the results of Experiment 4 showed that, although the durations of words in the novel text condition were underestimated compared to standard words, the novel words were not recalled more than the standard words. If anything, there was a trend in the opposite direction. This suggests that memory encoding may not be the only process that distracts attention away from time. Also, the free recall test only measured recall of the words, not their fonts, colors, or sizes. Recall of the words relies on successful encoding of each word’s occurrence in the task, but other aspects of each word’s context may have been
encoded as well. Since the text styles were the source of novelty, it is plausible that some
attentional resources may have been directed toward encoding font, color, and size, although
further research to test this possibility is still required.

Experiment 6 helped to compare the current experiments investigating the effects
of words on duration judgments with previous literature that has primarily used simple
visual stimuli. As discussed throughout this paper, research using words has consistently
demonstrated underestimation of durations, while comparable research using visual stimuli
in similar situations has demonstrated overestimation of durations. This applies to studies
of the effects of emotion on the perception of time, as reviewed in Chapter 1, and to studies
of the effects of novelty, as in the oddball effect literature. Durations for taboo words
are underestimated compared to neutral words, while durations for other emotional stimuli
tend to be overestimated compared to neutral conditions. Durations for novel words are
underestimated compared to standard words, while durations for visual oddball stimuli are
overestimated. This suggests that words may affect duration judgments differently than
visual stimuli, perhaps through their effects on attention. By comparing words and shapes
in the same study, Experiment 6 provided some preliminary support for this possibility.
CHAPTER 4: SUMMARY AND CONCLUSIONS

The six experiments in this dissertation provide support for information-processing models that describe time perception of verbal stimuli. According to the attentional gate model, words that capture attention, either due to their emotional content or novel visual appearance, direct cognitive resources away from the processing of time. These resources may be used for various non-temporal functions, including processing of less fluent words and encoding emotional or novel words into memory. In prospective timing tasks, less time is counted and durations are underestimated when attention is directed away from time. In retrospective timing tasks, attention is less relevant because participants do not typically attend to time without being instructed to do so. Instead, according to the memory storage model, participants must form their retrospective duration judgments based on their memories for non-temporal information associated with the time interval they are judging. In this case, the same attentional processes that cause durations of emotional and novel words to be underestimated in prospective tasks cause these words to be remembered more easily. When there are more memories associated with a time interval, those intervals are judged to be longer in retrospective tasks. Thus, durations for emotional and novel words are overestimated retrospectively for the same reasons that they are underestimated prospectively.

Experiments 1, 2, and 3 explored the effects of taboo words, defined as low-valence, high-arousal words considered inappropriate in polite social settings, on prospective and retrospective duration judgments. In the one previous study that had tested the effects of taboo words on the perception of time, Tipples (2010) found results that differed from much of the rest of the literature on time perception and emotional stimuli, so this was a topic that required further replication and expansion.

Experiment 1 compared duration judgments for taboo and neutral words in a temporal bisection task. Replicating previous literature, the results revealed that durations for taboo words were underestimated compared to neutral words. This effect was larger for
shorter stimulus durations than for longer stimulus durations, which supported explanations of attention-based but not arousal-based mechanisms. In addition, more taboo than neutral words were remembered in a surprise free recall task that was given after the timing task. There was a trend suggesting that durations for the words that were recalled may have been underestimated compared to the words that were not recalled. This provides some support for the idea that memory-encoding processes help to redirect attention away from time, causing durations to be underestimated.

Experiment 2 compared duration judgments for taboo and neutral words in an ordinality comparison task. This simpler task replicated several of the results from Experiment 1. Durations of taboo words were underestimated compared to neutral words. When asked to compare the durations of taboo and neutral words displayed in pairs, participants were significantly less likely to correctly identify a taboo word as being displayed for longer than they were to correctly identify a neutral word as being displayed for longer. This effect was larger for shorter duration pairs, which supported attentional explanations. In the surprise free recall test, participants remembered more taboo words than neutral words, suggesting that memory encoding may have interfered with the timing task.

Experiment 3 compared duration judgments for taboo and neutral words in a retrospective duration reproduction task. Participants gave significantly longer reproduced durations for taboo than neutral words. Combined with the results of Experiments 1 and 2, this suggests that durations for taboo words are underestimated prospectively but overestimated retrospectively. This supports the idea that the enhanced memory encoding for emotional stimuli causes durations to be overestimated, as suggested by the memory storage model (Ornstein, 1969).

Experiments 4, 5, and 6 examined the effects of novel font styles on prospective duration judgments in the temporal bisection task and in the oddball paradigm. These studies attempted to replicate the results of Experiments 1 and 2 by using non-emotional words displayed in novel text styles. This revealed that novel text styles and taboo words
produced a similar effect on the perception of time. This similarity suggests that the effects of taboo words on the perception of time may not necessarily have been caused by emotion per se, but rather the attentional effects of those emotional reactions.

Experiment 4 compared duration judgments for words displayed in novel and standard text styles in the temporal bisection task. It was designed to be identical to Experiment 1, except that the taboo words were replaced with words shown in various novel combinations of fonts, colors, and sizes. The results revealed that durations for words in the novel condition were underestimated compared to words in the standard condition, with larger effects at shorter durations. This replicated the results of Experiment 1 and suggested that attentional mechanisms driven by novelty can cause the same pattern of timing effects seen with taboo words in the temporal bisection task.

Experiment 5 attempted to replicate the results of Experiment 4 with the words displayed in novel text styles appearing with a reduced frequency, which became the “oddball” condition. Although this study used the temporal bisection task, the relative rarity of the novel text styles was intended to make the results comparable to other studies looking at the oddball effect. Words in the oddball condition were underestimated compared to the standard condition, but only for the shorter stimulus durations used in the task. Memory was also better for the words in the oddball condition than the words in the standard condition.

Experiment 6 compared word and shape stimuli in the oddball task. Previous studies using simple visual stimuli have shown that the durations for oddballs are usually overestimated, so this study attempted to directly compare simple visual stimuli and words in order to reconcile the results of Experiment 5 with previous findings. The results revealed that durations for word oddballs were underestimated compared to shape oddballs. This suggested that oddball words and shapes may affect duration judgments differently, and that the differing results of the current studies and published papers on the oddball effect could potentially have been caused by the use of words instead of shapes.

Although the experimental evidence and the applications of theoretical models de-
scribed in this dissertation are new, some of the general ideas are not: William James observed over a century ago that durations seem longer prospectively when “we grow attentive to the passage of time itself”, but also that durations seem longer in retrospect depending on “the multitudinousness of the memories which the time affords” (James, 1890). Focusing attention on time seems to extend the subjective length of durations, while focusing attention on non-temporal information decreases subjective duration estimates. When making retrospective judgments from memory, the same processes that cause attention to be directed away from time can trigger the encoding of greater quantities of non-temporal information into memory, resulting in longer duration judgments.
REFERENCES


Bates, D., Maechler, M., & Bolker, B. (2012). *Lme4: Linear mixed-effects models using S4 classes* [Computer Software].


doi:10.1016/j.bandc.2013.06.008


doi:10.1111/j.1467-9280.2008.02112.x


