Pushing the limit: A further investigation into the exceptional ability to break Miller's processing capacity

Babette Patricia Rae
B.Psych.(Hons I)(Newcastle), B.Mus. (Newcastle)

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Science (Psychology)

April, 2018

This research was supported by an Australian Postgraduate Award (APA)
Statement of Originality

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision.

The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University’s Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.
Statement of Collaboration

I hereby certify that some of the work embodied in this thesis has been done in collaboration with other researchers. I have included below, as part of the thesis, a statement clearly outlining the extent of collaboration, with whom and under what auspices.

Some of the information in Chapter 1 of this thesis was similar to the work included in a chapter of which I am a co-author:


The content of Chapter 2 was published in a paper of which I am a co-author:


The initial inspiration for the work in Chapter 3 was discussed with Professor Roger Dean from the MARCS Institute for Brain, Behaviour, and Development (University of Western Sydney). Professor Dean created the tones used in this experiment and also collaborated with Professors Brown, Professor Heathcote, and I in the experimental design of this project.

The work in Chapter 4 was made possible by the help of Associate Professor Shuguang Wang as he was the key liaison between The University of Newcastle and the Sichuan University (Chengdu, China). Appreciation also goes to Head Professor Ziao Yan Wang and Head Professor Jiang Wi Zhao for their co-operation and assistance in recruiting participants from Sichuan University, China. Gratitude is also extended to Charlene Quinchun Gong for her assistance in translating documents and translating during the visit to China. Finally, this experiment was made possible by funding from the Keats Endowment Research Fund No. G1101146/KERF11.
The work in all Chapters, 1 through 6, was made possible by the assistance and guidance by my primary PhD supervisor, Professor Scott Brown, and my secondary PhD supervisor, Professor Andrew Heathcote, who were involved with the conception, design, analysis, and interpretive stages.
Publications during Candidature

During my candidature, I contributed to published work, listed below, some of which were not directly related to the content in this thesis.


* These publications are directly related to the work in this thesis.

** This publication is cited in Chapter 6 of this thesis in the section *Future Research Directions*
Acknowledgements

I would like to express sincere gratitude to my truly excellent primary supervisor, Professor Scott Brown, for his amazing support, guidance, patience and motivation. He has given me so many opportunities to expand my skill base throughout, and beyond the content of, the PhD. His incredible knowledge and expertise has been amazing, and I could not have hoped for a better supervisor throughout my Honours year and PhD candidature.

I would also like to thank my secondary supervisor, Professor Andrew Heathcote, who encouraged me to pursue research within the Newcastle Cognition Lab, took me on as Summer Scholarship student, and supported and guided me throughout my PhD candidature, giving me a variety of opportunities along the way. In particular, I am grateful for his facilitation of the research trip I conducted in China.

I would like to extend gratitude to the members of Newcastle Cognition Lab, past and present, staff and students, who have been a part of supporting me as I write my thesis. Particularly, I’d like to thank Dr. Pennie Dodds who gave me support throughout my Honours year and my early PhD. Pennie is now a friend and continued to provide encouragement as I finished my thesis. I’d also like to thank Dr. Ami Eidels who gave feedback on some of my projects, and Dr. Guy Hawkins who showed me support in the lab and allowed me to contribute to one of his papers.

I am also grateful to those who co-authored published material with me during my candidature: Lee Averell, Professor Scott Brown, Dr. Max Bushmakin, Dr. Pennie Dodds, Dr. Chris Donkin, Dr. Nathan Evans, Dr. Guy Hawkins, Professor Andrew Heathcote, Dr. Keith Nesbitt, and Associate Professor Mark Rubin.
I also very much appreciate Lynne Slater taking the time to proof-read my thesis, and for meeting with me to discuss her suggestions.

I would also like to thank my fabulous and supportive friends for encouraging me to finish this PhD. Bree Slater, you not only encourage me to bring my A-game to Candy Crush, but you have been my biggest cheerleader in making sure I finish my PhD. Thank you for the beautiful bracelet you gave me as encouragement to finish: *She believed she could so she did*. Thank you for believing me in when I didn’t believe in myself. Dr. Emily Freeman, thank you for your support and being available to answer questions I’ve thrown at you throughout the writing process, and for setting an amazing example for me with your balance of academic work and creative pursuits. Dr. Janet Truong, thank you for encouraging me to finish my PhD so I can set a good example for my son, Hamish, and show him that education is important. Emma Bence, thank you for the beautiful gift of a coffee plunger, and especially for your support and encouragement, especially as our gorgeous sons play together. Deb Bilby, thank you for bringing me coffee and for visiting when I needed it most. Jo Dyer, you have supported me from the start and your encouragement over many pots of tea, along with existential discussion, has been immeasurable. Thank you to Dr. Heather Douglas for being a great friend throughout our undergraduate studies and postgraduate studies and encouraging me to do my best in reaching for my academic goals. Thank you to all the wonderfully supportive women in Book Club who encouraged me the entire way, and for the beautiful bunch of flowers you surprised me with the meeting after I submitted. And to my other friends who have encouraged me along the way: Thank you. I am privileged to have such an amazing group of female friends, and as one of my favourite quotes says: *Empowered women empower women.*
I am so grateful to my supportive husband, Luke Swetnam, who has been with me through every stage of this journey. I love that you support my educational pursuits. Thank you for making me tea and bringing me chocolate. And thank you for taking Hamish on extra father-son play-dates so I had time to write.

Thank you to my family who have supported me through my PhD candidature. Particularly, I appreciate my parents allowing me to write from their house in the last stages of writing. I must extend an extra special thank you to my Dad, Richard Rae, who, at 65 still works full-time shift work, and yet has given me many of his days off over the past year to spend time with my son, Hamish, so that I can write. I appreciate it so much; Hamish has loved his adventures with Grandpa. Thank you also to my sister, Linda Baker, who has welcomed Hamish into her pack so that I can spend uninterrupted time writing at the library.

Thank you to my beautiful dogs, both here and remembered, Reuben, Mr. Homer Bear, Oscar, Ella, Billie Bean, and my cat, Bruce Willis II. You make my heart happy.

And finally, to my son, Hamish, I love you so much for just being you. You have made me want to finish my PhD to show you that determination, perseverance, and hard work are worth it. The words to describe my love for you haven’t been invented yet. If I weren’t a Muggle, my Patronus would be conjured by remembering the first time you smiled at me.
# Table of Contents

## PREFACE
- Statement of Originality: 1
- Statement of Collaboration: 2
- Publications during candidature: 4
- Acknowledgements: 5
- Table of Contents: 8
- Abstract: 10

## CHAPTER 1: Introduction: A background to absolute identification and absolute pitch
- Introduction and background: 12
- Overview of Thesis Chapters: 37

## CHAPTER 2: Using a Structural Forms Algorithm to investigate the psychological representation of uni-dimensional stimuli
- Introduction: 47
- Method: 54
- Results: 56
- Discussion: 62

## CHAPTER 3: Do overtones and tones of the Western musical scale allow for better absolute pitch identification?
- Introduction: 67
- Method: 72
- Results: 74
- Discussion: 76
**Abstract**

Absolute identification (AI) is a paradigm in which participants identify stimuli varying on one physical dimension, such as line length. Stimuli are presented during a practice phase with a unique label, and during a test phase the participant is required to report labels based on the stimulus alone. Miller’s (1956) processing capacity limit of 7±2 was long thought to be the ceiling for the number of unidimensional stimuli participants could learn to identify in an AI task. Recent research, however, has proven this limit to be breakable by some exceptional participants. This thesis contains four experimental chapters that aimed to answer three central questions that relate to the ability to exceed Miller’s processing capacity limit. Firstly, to what extent is this ability learned? Secondly, what quality do people have who can learn beyond this limit? And finally, how far beyond this limit is it possible to go? Following the introduction in Chapter 1, Chapter 2 describes a Structural Forms algorithm (Kemp & Tenenbaum, 2008) to investigate participants’ psychological representation of stimuli across several modalities in various AI learning tasks. The examination of whether the multi-dimensional nature of musical tones contribute to an increased accuracy, and whether the frequencies aligning with the current Western musical scale are better identified, particularly for musicians, is presented in Chapter 3. Chapter 4 describes an experiment conducted in China investigating whether native tonal-language speakers have an AI advantage with pitch compared to Western populations. Chapter 5 reports about a long term learning experiment with tones of varying frequency and lines of varying length. Together, these results confirm findings that breaking Miller’s limit in identifying unidimensional stimuli is possible through learning, and that it is correlated with having a complex psychological representation of the stimuli. It is suggested that future
research directions focus on disentangling these two factors, as well as collaborating with other psychological research areas such as neuro- and bio-psychology.
Chapter 1

Introduction: A Background to Absolute Identification and Absolute Pitch

People have a practically unbounded capacity for learning complex items such as names, letters of the alphabet, and faces (Dodds, Donkin, Brown & Heathcote, 2011). In comparison, a task that has long provided an exception to this rule is **absolute identification**. Absolute identification is the fundamental experimental task of identifying stimuli that vary on only one physical dimension (**uni-dimensional**), such as tone frequency (pitch; e.g. Hartman, 1954; Pollack, 1952), tone intensity (loudness; e.g. Garner, 1953), or line length (e.g. Lacouture, 1997). In a typical absolute identification task, stimuli are presented to the participant one at a time during a practice phase, each with a unique label (usually 1 through \( n \)). For example, in an experiment using tone of varying pitch as the stimulus, a participant might be presented with a set of \( n \) tones of varying frequency, labelled 1 through \( n \), from lowest to highest tone. Rae (2010) used 36 tones of varying frequency in an absolute identification task. In this particular case, each tone matched a note on the piano ranging from A3 (220Hz) to G#6 (1,661.22Hz). During the practice phase of this experiment, participants were played each of these 36 tones over headphones, one at a time and in ascending order. After each tone was played participants saw the following on screen:

"This is tone x. When you see this tone, press the x button."

In the test phase of an absolute identification experiment, randomly selected stimuli from the set are presented to the participant, who is then asked to recall the previously associated label by selecting the appropriate button on screen. After hearing a tone through a set of headphones, participants select the button on screen
that they think matches the tone they previously heard during the practice phase.

Figure 1.1 shows the array of selection buttons the participants were presented with in Rae (2010) during the pitch identification task.

<table>
<thead>
<tr>
<th></th>
<th>A3</th>
<th>A#3</th>
<th>B3</th>
<th>C4</th>
<th>C#4</th>
<th>D4</th>
<th>D#4</th>
<th>E4</th>
<th>F4</th>
<th>F#4</th>
<th>G4</th>
<th>G#4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>A4</td>
<td>A#4</td>
<td>B4</td>
<td>C5</td>
<td>C#5</td>
<td>D5</td>
<td>D#5</td>
<td>E5</td>
<td>F5</td>
<td>F#5</td>
<td>G5</td>
<td>G#5</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>A5</td>
<td>A#5</td>
<td>B5</td>
<td>C6</td>
<td>C#6</td>
<td>D6</td>
<td>D#6</td>
<td>E6</td>
<td>F6</td>
<td>F#6</td>
<td>G6</td>
<td>G#6</td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.1. The layout of the screen in the pitch identification experiments in Rae (2010). Each row represents an octave (12 semitones). Frequencies ranged from A3 (1) to G#6 (36). Participants selected the button on screen they believed to have matched the tone played during the previous trial.*

After participants have made a selection, they are typically given trial by trial feedback about whether their response was correct or not. For example, after the participant selects the button on screen they are either given the response “Correct” if their response matched the tone, or a correction such as “Incorrect. That was tone 34 (F#6)” if their response was incorrect. Some absolute identification experiments, such as Rouder et al., (2004) and Dodds et al., (2011) give participants two chances to correctly identify the same stimulus.
The length of absolute identification experiment varies greatly from one to many hours. Typically, each session is split into several blocks across a one-hour period, with enforced short breaks between each block. The length of the blocks will vary depending on how many uni-dimensional stimuli are in the set so that each of the stimuli can be presented an equal number of times during each block. For example, in Rae (2010) where there were 36 stimuli, in a one-hour session there were four blocks, with 108 trials in each block. During each block, each of the 36 stimuli was repeated three times in random order.

*Miller’s (1956) limit of 7±2*

In his seminal paper, Miller (1956) investigated limits in both short term memory and in absolute identification. He found that 7±2 was not only the number of chunks that can be held in short-term memory, but was also the number of items people could learn to perfectly identify in a uni-dimensional stimulus set. It is possible that when Miller presented this range, he did not mean for it to be a hard limit. Instead, it is possible that he meant it to be a confidence interval for people’s ability to retain chunks in memory and in absolute identification, so that 95% of participants fell within these bounds. Perhaps he would not have been surprised if people were occasionally able to retrieve 10 chunks from memory, for example.

While this is possibly the case, the operational definition of Miller’s (1956) 7±2 limit in this thesis is to use the upper limit of nine as a strict criterion boundary when applied to absolute identification tasks. The justification for this is two-fold. Firstly, the literature that documents the previous absolute identification experiments suggests that this had been the case for the overwhelming majority. In practice, the
inability to identify beyond this limit suggests that an upper limit of nine is, in fact, a more robust limit than the upper bound of a confidence interval.

This upper limit of Miller’s range (nine stimuli) is particularly surprising because it was shown to be resistant to many experimental manipulations, including the number of stimuli in the set (e.g. Garner, 1953; Pollack, 1952). This relatively low upper limit of nine stimuli holds even when the range of stimuli is greatly increased. Braida and Durlach (1972) showed that once adjacent stimuli are perceptually discriminable, further increases to the range do not influence performance. This limit appeared to be a fundamental aspect of human information processing rather than a sensory limitation, because the same limit applied to a wide range of stimulus modalities (including electric shocks, saltiness, line length, hue, and loudness: e.g., Lacouture, Li & Marley, 1998; Pollack, 1952; Garner, 1953; Miller, 1956).

A further key finding is that this capacity limit was even shown to be highly resistant to practice. For example, Garner’s (1953) participants engaged in up to 12,000 judgements of tones of varying intensity in a single condition yet were unable to identify beyond 7±2 stimuli at the end of the experiment. Similarly, Weber, Green, and Luce (1977) had their participants complete 12,000 trials identifying six white noise signals of varying intensity and found a low level of improvement of just 6 per cent. The final performance for these participants was well below ceiling level, despite the large amount of practice, monetary incentives, and the apparently easy task of identifying just six separate levels of loudness. Furthermore, Hartman’s (1954) participants practiced over an 8-week period, and although they were able to demonstrate substantial improvement, their best performance level was still well within Miller’s limit of 7±2: the equivalent of the identification of only five stimuli.
Such results established this truism about AI: It is a severe limitation in human ability to identify uni-dimensional stimuli, that is largely unaffected by practice.

The second justification in defining Miller’s (1956) upper limit of nine stimuli to be a hard limit rather than the upper bound of a confidence interval is that this thesis integrates the practice of absolute identification and absolute pitch. Absolute pitch is a similar task to absolute identification but limits stimuli to tones of varying frequency. Absolute pitch is shown to have its own limits with a very small number of people having an exceptional ability to accurately identify more than 50 tones. Miller acknowledged this in his seminal paper but was unable to explain why, nor did he make this the focus of his paper. Absolute pitch is described in more detail throughout this thesis.

As well as Miller’s (1956) limit of 7±2, another benchmark phenomenon associated with absolute identification is the bow effect (Kent & Lamberts, 2005; Pollack, 1953; Weber et al., 1977): When response accuracy is plotted as a function of stimulus position, the plot reveals a bow shape. This reflects that, in a stimulus set, both the smaller and larger stimuli are better identified than the items in the middle of the stimulus range. The same stimulus can have different rates of response accuracy depending on whether it is in the middle or at the edges of a stimulus set (Lacouture & Marley, 1995; Lacouture, 1997). The bow effect is shown regardless of the size of the stimulus set.

A further consistent finding in absolute identification literature is that of sequential effects, that is, the influence of recent stimuli and responses on the current decision (Gilden, Thornton, & Mallon, 1995; Holland & Lockhead, 1968; Van Orden, Holden, & Turvey, 2003; Wagenmakers, Farrell & Ratcliff, 2004, 2005; Ward & Lockhead, 1970, 1971). These sequential effects come in two varieties,
assimilation and contrast. Assimilation is seen where the response made on trial \( N \) is similar to the stimulus presented on trial \( N-1 \). Contrast is seen where stimuli further back in the sequence of trials have an opposite influence on responses, where the response on trial \( N \) tends to be dissimilar to the stimuli on trials \( N-2, N-3 \), and further back in the sequence. The contrast effect is usually smaller than the assimilation effect, and although it tends to decrease over trials, contrast can persist until about trial \( N-5 \) (Holland & Lockhead, 1968; Lacouture, 1997; Ward & Lockhead, 1971).

Assimilation and contrast effects are quite complicated and have proven difficult for theories of absolute identification to account for. A successful theoretical account of absolute identification must predict that responses are biased toward the previous stimulus (assimilation) but that the bias changes direction (contrast) as the stimulus recedes in memory. Some researchers believe sequential effects to be the defining feature of absolute identification (e.g. Laming, 1984; Stewart, Brown & Chater, 2005).

Theories of absolute identification

Theories of absolute identification aim to account for the benchmark phenomena associated with absolute identification. Donkin, Rae, Heathcote and Brown (2015) cover several different conceptual classes of theories (Thurstonian models, exemplar models, relative judgement models, and restricted capacity) which will be briefly outlined here. (For a more comprehensive coverage of theories of absolute identification, see Stewart et al., 2005.)

Thurstonian models of absolute identification assume that any given stimulus evokes a noisy representation of absolute magnitude on an internal scale. The scale is divided by decision criteria into a set of response categories. The criteria that divide
the response categories are allowed to vary between trials because of imperfect memory or shifting responses bias. The noisy response categories can account for information limits when the number of stimuli increase for a fixed range, as the bounds become closer and so the noisy stimulus representation and the noisy category bounds lead to more errors in responding. Durlach and Braida (1969) extended the basic Thurstonian model to account for the invariance of performance with changes in the range of stimuli, suggesting that only recently presented stimuli are used to define the context within which the current stimulus is presented.

Thurstonian models can also account for bow effects, but not entirely satisfactorily. They predict greater accuracy for edge stimuli because Thurstonian models that predict erroneous responses are mostly adjacent to the correct response, and edge stimuli have only one adjacent response, whereas internal stimuli have two. Braida, Lim, Berliner, Durlach, Rabinowitz, & Purks (1984) elaborated the basic Thurstonian model’s account of bow effects to allow for greater variability for stimuli near the centre of the range than the ends. They explained this by proposing that stimuli are judged against two “anchors” at the extreme ends of the stimulus range, and the distances between these anchors and the presented stimulus are counted using a noisy measurement unit. The further a stimulus is from those anchors, the noisier the distance measurement becomes. This produces a gradual bow effect with increasingly frequency identification errors toward the middle of the measurement range. Luce, Green, and Weber (1976) proposed a related model, suggesting that bow effects may be due to more attention being paid to the edges of the stimulus range, thus reducing the variability of stimulus representations in those regions.
Exemplar models have proven very successful in accounting for categorisation behaviour, and this makes them promising candidates for theories of absolute identification because of the close similarity between absolute identification and categorisation. Exemplar models assume absolute identification is accomplished by determining the similarity between a to-be-identified stimulus, and the memory representations for previous stimuli. Each stimulus is assumed to be represented in memory along with its associated label. The probability of response $I_i$ is proportional to the similarity between the current stimulus $j$ and all exemplars for response $i$. In general, exemplar models face the problem of not being able to account for the fundamental information limit of absolute identification: When the range of the stimulus set is increased but the number of stimuli remains fixed, these models predict that the memory representations should be more easily discriminable, and so performance should improve. However, Braida and Durlach (1972) showed this does not happen.

An example of an extension of an exemplar model is shown by Nosofksy and Palmeri’s (1997) exemplar-based random walk model. Nosofsky (1997) extended exemplar models to also make predictions about the time taken to make decisions, rather than just the decision that is made. The exemplar-based random walk model (Nosofsky & Palmeri, 1997) assumes that the representation of stimuli in memory are normally distributed across the stimulus dimension. Upon presentation of a stimulus, the exemplars race to be retrieved from memory with a speed that is an increasing function of the similarity between the current stimulus and each exemplar. Each time an exemplar is retrieved, a counter for the associated response is incremented, whereas the counters associated with other responses are decremented.
Relative judgement models of absolute identification defining feature is that decisions are based on the difference between current and previous stimuli (or responses), rather than being based directly on representation of the absolute magnitudes of stimuli. This leads to a natural focus on sequential effects. Relative versus absolute judgements is not just a theoretical distinction in absolute identification modelling but is also important in some applied areas. For example, musicians are frequently trained in the skill of relative judgement (judging musical “intervals”; discussed in depth later in this thesis) but almost never in absolute judgement, because relative judgement is useful in musical performance, whereas absolute judgement is not (and may even be maladaptive). This difference may help explain the rarity of “absolute pitch” in musicians (which is also discussed further below).

Holland and Lockhead (1968) proposed that responses are made by combining feedback from the previous trial and the perceived distance between the current and previous stimuli. This basic relative judgment mechanism accounts for assimilation and contrast by simply assuming that the judged difference between the current and previous stimuli is biased toward the previous stimulus and away from earlier stimuli. For example, if a small stimulus was presented on the previous trial, it means that, on average, the stimuli presented on earlier trials were probably larger stimuli. The memories for these larger earlier stimuli interfere with the judgement of the distance between the previous and current stimulus in a way that causes the distance to be underestimated. Hence, a response based on this distance will assimilate to the previous trial.

Laming (1984) proposed a strict version of a relative judgement model, assuming that no absolute information is used, and that only the difference between
the previous and current stimulus is considered. In particular, decisions are assumed to be made in a relatively coarse manner, such that the current stimulus is judged in terms of only five categories: “much less than”, “less than”, “equal to”, “more than”, or “much more than”. Such limited categorical information provides a natural account of the fundamental capacity limits, within a relative judgement framework.

Stewart et al. (2005) proposed the most current and successful relative judgement model of absolute identification. In their model, only the series of differences between each stimulus and the next is represented internally. These differences, along with feedback from the previous trials, and used to produce a response. Stewart et al. (2005) have demonstrated that their relative judgement model is capable of producing all the classical response-choice related benchmark phenomena such as the Miller’s (1956) information limit and the bow effect.

Restricted capacity models attribute poor performance in absolute identification to limited capacity in memory or attention. For example, Marley and Cook (1984, 1986) assume that the full range of stimuli in an experiment are mapped onto an experimental context, which might be thought of as a fixed-capacity attention or memory store. When a stimulus is presented its relative position within the set of stimuli is located within a context by reference to “anchors” located near the ends of the stimulus range. The relative position of the stimulus is then used to judge its magnitude and subsequently assign a response label. Lacouture and Marley (1995, 2004) proposed an alternative means of explaining the information capacity limit and bow effect through their mapping model.

The SAMBA model (Brown, Marley, Donkin, & Heathcote, 2008) extends on the previous restricted capacity models mentioned and attempts to account for the benchmark sequential effects in absolute identification, by using as deterministic
version of the leaky, accumulator model (Usher & McClelland, 2001), Brown and Heathcote’s (2005) ballistic accumulator model. SAMBA is a comprehensive theory of absolute identification because it showed it was capable of accounting for all the aforementioned benchmark phenomena in absolute identification, for both response choices and response times.

Recent research suggests Miller’s (1956) limit is breakable

Despite the longstanding assumption that uni-dimensional stimuli are unable to be learned beyond the upper limit of nine stimuli (Miller, 1956), recent work has identified exceptions. Rouder, Morey, Cowan, and Pflatz (2004) showed that all three participants in their experiments using lines of varying length demonstrated some learning to a level beyond Miller’s upper limit of nine stimuli, with one of the participants learning to identify 20 stimuli.

Dodds et al. (2011) extended Rouder et al.’s (2004) findings not only for line lengths, but also for dot separation, the degree of angular rotation for a line, and tone frequency across a series of seven experiments (see Table 1.1 for a summary of experiments and findings). Interestingly, although learning was observed for these modalities, no such learning was demonstrated for tones varying in intensity. This is reminiscent of Nosofsky’s failed personal attempt to learn tones of varying intensity (Shiffrin & Nosofsky, 1994). Two of their six participants who practiced tones of varying frequency showed substantial improvement, beyond Miller’s (1956) upper limit of nine. One of those two participants showed a huge amount of learning, and by the end of 10 experimental sessions could identify over 14 items, well beyond Miller’s limit.
Table 1.1

A summary of the experiments and findings from Dodds, Donkin, Brown and Heathcote (2011)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Purpose of experiment</th>
<th>Stimuli</th>
<th>Set size</th>
<th>Participants</th>
<th>Max. average performance (equiv. stimuli)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>To investigate whether Rouder et al.’s (2004) response technique of giving participants two responses for each trial was responsible for the increase in learning.</td>
<td>Lines (length)</td>
<td>30</td>
<td>6</td>
<td>9</td>
<td>Learning was not due to the two-response method.</td>
</tr>
<tr>
<td>2</td>
<td>To investigate whether relative strategies such as the edges of the computer screen were involved in the identification of stimuli.</td>
<td>Dots (separation)</td>
<td>2a. 30</td>
<td>2a. 5</td>
<td>2a. 10.3</td>
<td>External visual cues were not responsible for the learning effect in Experiment 1. Also, learning does not vary with set size.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2b. 15</td>
<td>2b. 5</td>
<td>2b. 8.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>To investigate whether substantial learning demonstrated in line lengths is due to the longer stimulus presentation than for tones.</td>
<td>Dots (separation)</td>
<td>30</td>
<td>6</td>
<td>8.3 (One participant slightly exceeded Miller’s (1956) upper limit of 9)</td>
<td>The extended stimulus presentation time associated with lines compared with auditory stimuli was not required for learning.</td>
</tr>
<tr>
<td>4</td>
<td>Identical to Experiment 3 except for the stimuli.</td>
<td>Lines (angle)</td>
<td>30</td>
<td>6</td>
<td>9.05 (Three participants exceeded Miller’s limit.)</td>
<td>Substantial learning also occurred for another visual stimulus—lines of varying inclination.</td>
</tr>
<tr>
<td>5</td>
<td>To investigate whether the learning effect could be seen beyond visual stimuli.</td>
<td>5a. Lines (length)</td>
<td>16</td>
<td>2a. 5</td>
<td>5a. 8.86 (Two participants exceeded Miller’s limit)</td>
<td>The findings for loudness (intensity) were consistent with previous findings of a low channel limit and little improvement with substantial practice.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5b. Tones (intensity)</td>
<td>16</td>
<td>2b. 5</td>
<td>5b. 3.86 stimuli.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Experiments 1-5 suggest the amount of improvement in the learning task is closely related to initial level of performance. This is tested using tones of varying frequency as they’ve previously been shown to have slightly higher pre-practice performance.</td>
<td>Tones (frequency)</td>
<td>36</td>
<td>6</td>
<td>6.02 (One participant learned to 14.4)</td>
<td>The amount of improvement was relative to initial performance. The participant who learned to 14.4 stimuli had several years of musical training beginning at an early age.</td>
</tr>
<tr>
<td>7</td>
<td>Experiment 7 attempted to decouple initial performance level from stimulus manipulations by manipulating participants’ motivation. Experiment 7 was identical to Experiment 6 but reimbursement was contingent on performance.</td>
<td>Tones (frequency)</td>
<td>36</td>
<td>6</td>
<td>6.87</td>
<td>The difference between initial performance in Experiments 6 &amp; 7 was not significantly different, although performance increase in Experiment 7 was almost double suggesting receiving motivational reimbursements improves performance.</td>
</tr>
</tbody>
</table>

Note: Almost all participants completed ten one-hour sessions.
Both Rouder et al.’s (2004) and Dodds et al.’s (2011) findings contradict Miller’s (1956) theory of a small upper limit for memory processing capacity. These results are an important finding, given that the memory processing capacity limitation has been a key element of many theoretical accounts of absolute identification.

A further key point that Dodds et al.’s (2011) study showed was a considerable variability in individual difference in the ability to learn in absolute identification tasks. The best and most consistent predictor of such differences seemed to be initial performance. Participants who show a higher level of performance at the start of the study were also those who showed the greatest improvement with practice.

Although these newer findings have shown that in isolated cases some participants can learn to identify beyond Miller’s (1956) upper limit of nine uni-dimensional stimuli (Rouder et al, 2004; Dodds et al, 2011), most people cannot exceed this limit. Given this seemingly dichotomous ability, two questions were now posed. Firstly, to what extent is learning possible in AI? Secondly, what quality do participants who are able to learn beyond Miller’s limit have that enables them to identify beyond nine uni-dimensional stimuli?

\textit{Absolute pitch}

A paradigm that may yield some answers is another area of study very similar to absolute identification called \textit{absolute pitch}. Absolute pitch is remarkably similar
to absolute identification in that both require the participant to label stimuli that vary
on a physical dimension with previously assigned labels. Absolute pitch, however,
restricts itself to tones of varying frequency (pitch). The tones used in absolute pitch
tasks are usually the discrete frequencies that correspond with the Western musical
scale; that is, the notes tuned on a standard concert piano (Takeuchi & Hulse, 1993;
Yost, 2009). Each of these notes is referred to by both tone chroma, such as C, C-
sharp (C#), or D, and octave number (for example: C#3), with the twelve chroma
repeating cyclically in octaves throughout the musical range. Therefore, absolute
pitch is a specific case of absolute identification that uses tones of varying frequency,
where participants usually label the stimuli using note names instead of arbitrary
numerical labels.

Absolute pitch is generally considered to be a very rare ability with only 1 in
10,000 of the general population reported as having the ability (Takeuchi & Hulse,
1993). However, rates rise to 1 in 1,500 amongst amateur musicians (Profit &
Bidder, 1988), and up to 1 in 7 in highly accomplished musicians (Baharloo,
Johnston, Service, Gitscheir, & Freimer, 1998). In a similar pattern to absolute
identification, there is a seemingly dichotomous ability where most people are
unable to label musical tones. In contrast, there is a well-known phenomenon where
some people are able to easily identify over 50 tones (Miller, 1956), which is
obviously well above Miller’s limit.

Being able to identify this number of tones is well beyond the findings shown
in recent absolute identification research (Rouder et al., 2004; Dodds et al., 2011), so
it is therefore possible that absolute pitch may hold some answers to investigate the
questions posed above: To what extent is learning possible in absolute identification?
And: What quality do participants who are able to learn beyond Miller’s (1956) limit
have that enables them to identify beyond nine uni-dimensional stimuli? The work I completed in my Honours year (Rae, 2010) began to address these questions. I conducted three experiments using the modality shown to have capacity limitation exceptions: Absolute pitch (tones of varying frequency). The results of Rae (2010) are outlined below, as it is important background information for the PhD rationale and experimental chapters in this thesis. That work left open many questions and identified new questions about learning in absolute identification tasks, which was the motivation for my PhD thesis.

Previous thinking considered absolute pitch an innate quality that was unable to be learned, with two distinct populations: those who possessed absolute pitch and those who did not (e.g. Revesz, 1953; Stumpf, 1883 in Takeuchi & Hulse, 1993). However, much of the evidence to support this has been anecdotal (Bahr, Christensen, & Bahr, 2005). Most recent researchers agree that there is a critical period, under the age of approximately seven or eight years, where children can easily learn absolute pitch if exposed to enough musical training (e.g. Levitin & Rogers, 2005; Takeuchi & Hulse, 1993). There are conflicting views, however, as to whether learning is possible beyond this critical period. Some researchers maintain that absolute pitch cannot be learned to any level of fluency after the critical period (e.g. Zatorre, 2003). Others argue, however, that with enough practice, learning is possible beyond this period (e.g. Levitin, 1994; Lundin, 1963). One of the aims of Rae (2010) was to investigate the level of learning that is possible to achieve as an adult.

Evidence that absolute pitch is, at least in part, a learned rather than purely innate, ability can be seen in studies where absolute pitch possessors are not uniform across identification of tones, more accurately identifying notes to which they have
had more exposure. Middle C and other white-key notes on the piano are often most accurately identified in tests of absolute pitch (Athos et al., 2007; Boggs, 1907; Lundin, 1963; Miyazaki, 1988; Takeuchi & Hulse, 1993) perhaps because these are the first notes learned in standard piano training such as in The Suzuki Method (Bahr et al., 2005; Bigler & Lloyd-Watts, 1979; Miyazaki, 2004). Similarly, flute players identify A4 (440Hz) most accurately as this is the note they tune to (Brammer, 1951; Ward, 1999), and in general musicians more accurately identify pitch on their own instrument (Pantev, Roberts, Schulz, Engelin & Ross, 2000; Takeuchi & Hulse, 1993).

Absolute pitch clearly bears considerable similarity to the absolute identification paradigm. Both tasks require participants to identify stimuli that vary on one physical dimension, although in absolute pitch tasks the labels are musical note names rather than arbitrary numerical labels. Both absolute pitch and absolute identification also yield similar results in performance. That is, for the most part, people are unable to identify beyond Miller’s (1956) upper limit of nine stimuli. The difference with absolute pitch, however, is that some participants are easily able to identify more than 50 musical tones, a feat which has not yet been observed in absolute identification experiments.

This apparent difference between the well-performing participants of absolute identification and absolute pitch might seem surprising as both tasks are fundamentally very similar. Rae (2010) aimed to reconcile the methodologies of absolute identification and absolute pitch practices, using the standardised experimental methods that have long been employed in tests of absolute identification. The main one of these differences was the stimuli used in each respective task. Piano tones have often used in absolute pitch tasks, and pure sine
tones are commonly used in absolute identification tasks. Lockhead and Byrd (1981) showed that piano tones were more accurately identified than pure sine tones, although the experimental procedures for assessing both types of tones were markedly different. Piano tones and pure sine tones have different physical qualities. Piano notes are complex waveforms consisting of a fundamental frequency and a series of harmonic overtones, with marked variations in timbre, volume, resonance and decay characteristics across the registers of the piano (Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983). As such, piano tones are multi-dimensional stimuli, so they are not bound by Miller’s (1956) upper limit of nine. Alternatively, pure sine tones are uni-dimensional, as they are simply the fundamental frequency itself so vary on the dimension of frequency alone (Yost, 2009).

The first of Rae’s (2010) experiments was a pitch identification task using musicians of varying expertise to identify three octaves (36 tones; chromatically separated from A3 [200Hz] to G#6 [1,161Hz]) of both piano and sine tones. These were the same stimuli that Dodds et al. (2011) used in Experiments 6 and 7. Participants were recruited from the Conservatorium of Music (University of Newcastle) and community music groups and choirs. Musically trained participants were specifically targeted in order to increase the chance that they would be able to break Miller’s (1956) upper limit of nine stimuli when identifying tones of varying frequency, in part because their initial starting accuracy was presumed to be higher than that of the general population. Participants were given a standard absolute identification task on a computer, where at the start of each section, there was a brief study phase to familiarise the participants with the stimuli and method of responding. Participants then completed a test phase of four blocks, where the first and third
blocks had feedback provided after each trial.

Feedback is typically given after each response in absolute identification tasks to aid with motivation and morale (e.g. Dodds et al., 2011; Rouder et al., 2004). However, providing feedback in tasks of identifying tones of varying frequency, especially to trained musicians, provides the option for participants to make use of relative pitch instead of, or as well as, absolute pitch. Relative pitch is the ability to identify differences between notes, rather than individual notes in isolation (Miyazaki, 1995). Musicians are explicitly trained to recognise these intervals, especially more common intervals such as unisons (when two successive tones are identical, e.g. A4 followed by A4), semitones (when a tone is followed by a tone that is chromatically separated by one step; the smallest non-unison interval possible, e.g. A4 followed by A#4) and octaves (when a tone is followed by another tone that has the same chroma name but is 12 semitones apart, e.g. A4 followed by A3; Levitin & Rogers, 2005).

In an absolute identification or absolute pitch experiment where feedback on a trial by trial basis is given, relative pitch abilities enable participants to identify stimuli purely based on coupling their ability to hear and identify an interval created by the previous and current stimuli, along with the stimulus label of the previous label. For example, if a musically trained participant was presented with the stimulus A4, even if they did not have absolute pitch abilities, the feedback telling them it was A4 would give them an anchor to be able to use that stimuli as the first note in the interval (i.e. two note sequence) so that when they heard the following (current) stimulus they could firstly identify the interval based on their training and then label it correctly by working out the interval (that is, the relationship or distance between the previous and current stimuli). Removing the trial by trial feedback would not
stop these trained participants being able to hear a musical interval, but without the knowledge of the first label in the pair, they would not be able to assign the correct label to the second note in the pair (the current stimulus).

A pilot experiment conducted by Rae (2010) showed that musicians were indeed making use of feedback to use relative pitch strategies instead of, or in addition to, absolute pitch strategies. Figure 1.2 shows the results of the pilot study where the musically trained participants were clearly able to make use of feedback for the more easily identifiable intervals such as unisons, semitones, tones, and octaves. For this reason, feedback was withheld every second block in Experiment 1 to restrict the use of relative pitch techniques but was provided in alternating blocks to stay true to absolute identification practices.

![Figure 1.2](image)

*Figure 1.2. Average accuracy calculated as a function of the musical interval created by the tones on current and previous trial for both piano and sine tones (Pilot study). The centre of the graph (0 – Unison) represents repeated notes (e.g. C4, C4). Dotted grey vertical lines represent octave intervals. For example, an ‘Ascending, 1’ interval represents an octave interval where the stimuli were presented in ascending frequency order (e.g. C4, C5). A ‘Descending, 2’ interval represents a two octave interval where the stimuli were presented in descending frequency order (e.g. C5, C3). (Figure from Rae, 2010)*
Four main findings were revealed from Rae’s (2010) first experiment that inspired the work in this PhD thesis. Firstly, piano tones were more accurately identified than sine tones, with performance showing a strong positive correlation between participants’ recognition of both tones. The difference in identification of the piano and pure sine tones was possibly due to the extra physical dimensions present in the piano tones. The experiment discussed in Chapter 3 of this thesis directly investigates this possibility.

The second finding from Rae’s (2010) first experiment was that the range of performance was not dichotomous. Participants’ ability to identify tones ranged across a continuum from chance levels to very good performance. This was the case for both piano and sine tones across feedback and no-feedback conditions (see Figure 1.3). The distribution of performance across the sample suggested a mixture of two underlying distributions: a narrow, approximately bell-shaped distribution centred just above chance level, and a wide, approximately uniform distribution extending across the entire range. The first group consisted of the majority of participants, whose accuracy fell below Miller’s (1956) limit of nine stimuli. The remaining six participants were fairly evenly distributed beyond Miller’s limit, from 9 to 36 stimuli (25 - 100% accuracy).

The third finding from Rae’s (2010) first experiment was that, as expected, feedback significantly improved performance. For this reason, all absolute identification experiments conducted as a part of this PhD thesis contained both blocks with and without feedback. Finally, it was shown that participants with more advanced musical proficiency and participants who started musical training in the critical period (under eight years of age) were more accurate in identifying both piano and sine tones, although often these features overlapped.
Rae’s (2010) second experiment investigated whether the differences in the physical dimensionality of piano and sine tones were also perceived differently psychologically. Multi-dimensional piano tones have been suggested to have a helical representation with notes an octave apart being in close proximity (Bachem, 1950; Deutsch, 1999; see Figure 1.4). In contrast, the uni-dimensional nature of pure sine tones has been said to be responsible for their linear psychological representation (Patterson, 1990). The participants who took part in the first experiment also completed similarity ratings of the piano and sine tones used in the first experiment. Multi-dimensional scaling techniques (Cox & Cox, 1994; 2001; Shepard, 1974) were used to examine the relationships between items in a stimulus set, thus constructing a model of spatial arrangements using proximity plots (Dodds et al., 2010).
Figure 1.3. Histograms of the distribution of accuracy in the pitch identification task (Experiment 1). The panels are split into piano tones without feedback (top left), sine tones without feedback (top right), piano tones with feedback (bottom left), and sine tones with feedback (bottom right). Miller’s (1956) range of $7 \pm 2$ is marked by the bracket on the x-axis (Accuracy), which is equivalent to a range of 13.9% to 25%. (Figure from Rae, 2010)
Figure 1.4. The Psychological Representation of Pitch: The Bi-dimensional Pitch Helix (Deutsch, 1999). The vertical dimension represents pitch height and the circular dimension represents the chroma names (from A to G-sharp [#]). For each revolution of the circular dimension the pitch rises one octave. The distance between notes of the same chroma name (i.e. octaves) is smaller than of some other notes that are physically closer. For example, A3 (220Hz) is perceived as more similar to A4 (440Hz) than it is to D#4 (311.13Hz). A close distance between notes in this spatial representation corresponds to similarity between stimuli.

The results of the multi-dimensional scaling analysis showed that the distinction of a helical versus a linear representation was not dependent on the type of tones, but rather participants’ accuracy on the pitch identification task. Participants who exceeded Miller’s (1956) limit for both piano and sine tones in the pitch identification experiment were shown to have a helical representation for both piano and sine tones, while those participants who did not exceed Miller’s limit had no such representation for either tone (see Figure 1.5). The work in Chapter 2 of this thesis further investigated the psychological representation of stimuli, across a range of modalities used in absolute identification tasks.
Figure 1.5. Proximity plots modelled using non-metric multi-dimensional scaling of the similarity rating data (Experiment 1; Rae, 2010) for participants who scored higher than Miller’s (1956) upper limit of 9 (25%) for (a) piano tones and (b) sine tones, and for participants who scored lower than Miller’s upper limit of 9 for (c) piano tones and (d) sine tones. Axis magnitudes are arbitrarily scaled dimensionless units, so labels were not included. (Figure from Rae, 2010)

Rae’s (2010) third experiment was a 10–session practice experiment investigating to what extent trained musicians could learn to identify tones of
varying frequency. Pure sine tones were used for this experiment so that any effect of learning could be likely attributed to learning the frequency of the tones, rather than the extra dimensions present in piano tones. Six participants who completed the first two experiments participated in this study. Two of the participants started the learning experiment with an initial accuracy of above Miller’s (1956) upper limit of nine stimuli. These two participants increased in accuracy by an average of more than seven stimuli for tones, both with and without feedback. In comparison, the participants who started with an accuracy below Miller’s limit were only able to increase identification slightly, increasing by two stimuli for the feedback condition and one stimulus for the no-feedback condition. These findings replicated and extended Dodds et al.’s (2011) findings that participants who have a higher starting accuracy tend to learn more during the experimental task. The best performer in Dodds et al.’s experiment was able to identify 14.4 stimuli at the end of the sessions, while the two better performers in Rae (2010) were able to identify an average of 29.0 stimuli (with feedback) by the end of the 10 sessions. Figure 1.6 shows the learning of all six participants in Experiment 3 across the 10 sessions. The stronger effect of learning was likely attributable to the high level of musical training in the current sample, and further supports Dodds et al.’s proposition that a large increase in the identification of stimuli can be observed for participants whose initial accuracy is good.
Figure 1.6. Average accuracy (and corresponding equivalent number of correct stimuli) for each participant, separately for feedback (black circles) and no-feedback blocks (white circles), across the ten sessions of the learning experiment (Experiment 3; Rae, 2010). The grey region on each graph is Miller’s (1956) range of $7 \pm 2$ stimuli. (Figure from Rae, 2010)

Overview of Thesis Chapters

The findings of Rouder et al. (2004), Dodds et al. (2011), and Rae (2010) left some questions open for further investigation: To what extent is the ability to exceed
Miller’s (1956) limit learned? What quality do people have who can learn beyond nine stimuli? How far beyond Miller’s limit is it possible to go? The following four experimental chapters, outlined briefly here in the Introduction, aimed to investigate these questions.

**Chapter 2: Using a Structural Forms Algorithm to investigate the psychological representation of uni-dimensional stimuli**

The work in Chapter 2 aimed to investigate how participants who are able to identify beyond Miller’s (1956) upper limit of nine perceived physically uni-dimensional stimuli. It is clear that the number of stimuli that can be reliably identified increases exponentially as the number of physical dimensions increase (Eriksen & Hake, 1955; Miller, 1956; Rouder, 2001) when those dimensions are perceived as independent (Nosofsky & Palmeri, 1996). For example, people are able to identify well beyond Miller’s upper limit of nine for faces, names, and letters. Making this example more similar to an absolute identification task, if a participant was able to perfectly identify seven line lengths, and seven angles, they could potentially identify 49 different stimuli with these combined features.

The stimuli used in absolute identification always vary on only one physical dimension, but this does not guarantee that the corresponding psychological representations of this stimuli are also uni-dimensional. In some cases multi-dimensional scaling techniques (Cox & Cox, 1994, 2001; Shepard, 1974) are able to reliably infer multi-dimensional psychological representations of physically uni-dimensional stimuli. For example, hue, the attribute of colour that allows the distinction between red, green, and so on, is physically uni-dimensional, varying only on the dimension of wavelength. However, psychologically hue is perceived as
a circle (Shepard, 1962; MacLeod, 2003), often represented as a colour wheel instead of a linear representation of increasing wavelength. It is possible that the participants in Dodds et al.’s (2011) and Rouder et al.’s (2004) studies who exceeded Miller’s limit (1956) might have had a more complex psychological representation of the uni-dimensional stimuli, thereby, giving them an advantage.

Dodds, Donkin, Brown, and Heathcote (2010) used multi-dimensional scaling analyses based on similarity ratings to investigate whether participants who were able to identify lines of varying length beyond Miller’s (1956) upper limit of nine stimuli used a multi-dimensional psychological representation. They found that multi-dimensional scaling was not reliably able to distinguish between one- and two-dimensional representations, and concluded that it did not seem likely that participants were using a multi-dimensional psychological representation. However, as already discussed above, Rae (2010) revealed that those participants who exceeded Miller’s (1956) limit in the pitch identification task not only had a helical psychological representation for physically multi-dimensional piano tones, but also for the uni-dimensional sine tones. This might suggest that the complex psychological representation is modality specific.

However, although multi-dimensional scaling techniques are useful for inferring the latent structure from similarity ratings, and have certainly been used in association with absolute identification for some time (e.g. Shepard, 1974), a downfall is that it lacks a framework of inference about these arrangements. In order to investigate the number of dimensions that best represent a relationship between objects, conclusions are based on eyeballing the data, and making subjective judgements rather than statistically assessing the number of dimensions of the psychological representation of the stimuli. Lee (2001) investigated this problem in
detail and found that, for one- or two-dimensional representations, multi-dimensional scaling correctly identified the number of dimensions only 14 per cent of the time.

Chapter 2 reports on analyses (Dodds, Rae, & Brown, 2012) using an algorithm to investigate the psychological representation of stimuli, in order to address whether participants who are able to exceed Miller’s (1956) limit have a more complex representation than participants bound by Miller’s limit. The analyses use Kemp and Tenenbaum’s (2008) algorithm to infer the structure of psychological representations based on relational data. Similarly to multi-dimensional scaling, the algorithm allows the representation of structures such as trees, lines, and hierarchies. However, Kemp and Tenenbaum’s algorithm improves on multi-dimensional scaling techniques, allowing a coherent framework for inference and the comparison of different structural forms based on a likelihood.

The Structural Forms (Kemp & Tenenbaum, 2008) analysis in Chapter 2 is based on a combination of data from several absolute identification learning experiments: Line length, dot separation, and tone intensity from Dodds et al., (2011) and also the tones of varying frequency from Rae (2010). This analysis was primarily aimed at investigating whether it was able to reveal more about the psychological representation of the uni-dimensional stimuli than multi-dimensional scaling could. Furthermore, it also examines whether having a complex psychological representation of uni-dimensional stimuli is modality dependent.

Chapter 3: Do overtones and tones of the Western musical scale allow for better absolute pitch identification?

The work in Chapter 3 aimed to extend the finding from Rae (2010) that multi-dimensional piano tones were more accurately identified than uni-dimensional
pure sine tones. The experiment described in Chapter 3 aimed to investigate two possible reasons for this discrepancy in performance. Firstly, it examined whether the multi-dimensional nature of the stimuli gave rise to superior identification. Secondly, it explored whether learning plays a role in the identification of tones of varying frequency.

Although Rae (2010) showed that there was a strong positive correlation between those participants who were able to identify piano tones and sine tones beyond Miller’s (1956) upper limit of nine stimuli, piano tones were still more accurately identified than sine tones. The multi-dimensionality of the piano tones was considered a reason for their superior identification as has been done in past research (e.g. Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983). Each physical dimension that is added to a uni-dimensional stimulus exponentially increases the ability to identify that stimulus (Eriksen & Hake, 1955; Miller, 1956; Rouder, 2001), when those dimensions can be perceived as separate from each other (Nosofsky & Palmeri, 1996). It has been suggested that piano tones have additional cues to frequency such as variations in timbre, volume, and resonance and decay (Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983). The experiment described in Chapter 3 aimed to investigate whether one additional dimension of tone frequency (i.e. overtone) improves performance on identifying stimuli.

The second aim of Chapter 3 was to explore whether learning plays a role in the identification of tones of varying frequency. Rouder et al. (2004) and Dodds et al. (2011) showed that learning of uni-dimensional stimuli is possible in the short term. Further, Rae (2010) revealed a correlation between musical proficiency and superior pitch identification. However, it was not ascertained whether the musicians
had an innate ability to identify tones as such, and were thereby drawn to a musical career, or whether their long exposure to the musical tones of the Western scale had allowed their better identification of them.

For a long while, absolute pitch was considered an innate ability that was unable to be learned (e.g. Revesz, 1953; Stumpf, 1883 in Takeuchi & Hulse, 1993), although more recent research has suggested that some kind of learning is possible (e.g. Dodds, et al., 2011; Levitin, 1994; Levitin & Rogers, 2005; Takeuchi & Hulse, 1993). To explore whether learning plays a role in the identification of tones varying in frequency, the experiment described in Chapter 3 compared the identification of sine tones based on the Western musical scale with sine tones not found on the Western musical scale. The non-Western tones were based on a fundamental frequency of prime numbers, so participants were not likely to have heard them before, or certainly not have heard them enough times to develop the association between the tone and the corresponding label. Levitin (1994) suggests this association is the essential aspect to developing absolute pitch ability. The expectation in this research was that if learning plays a role in the superior identification of tones of varying frequency, then Western tones would be better identified than non-Western tones, and this would be particularly apparent in the performance of musicians.

Chapter 4: Do tonal language speakers have an advantage in absolute pitch identification?

The work in Chapter 4 aimed to further the investigation about the role of learning in the identification of tones of varying frequency. Several researchers have noted a correlation between absolute pitch ability and musical training during the
critical period of under approximately seven or eight years of age (Miyazaki, 1988; Levitin & Rogers, 2005; Takeuchi & Hulse, 1993). However, the exposure to musical training during this critical period is not sufficient to develop absolute pitch, as most people who are musically trained during this period do not develop the ability (Baharloo et al., 1998). The essential aspect of early learning seems to be the paired association of the pitch with the note name (Levitin, 1994).

While rates of absolute pitch are reported as being very low in the general population (1 in 10,000; Bachem, 1955; Profita & Bidder, 1988) the rate increases up to 1 in 7 for highly-trained musicians (Revesz, 1953; Baharloo et al., 1998). Further, the incidence of absolute pitch is reported to be even higher in Asia, particularly China, than it is in North America and other English speaking nations (Schellenberg & Trehub, 2008; Deutsch, Henthorn, Marvin & Xu, 2006; Deutsch, Henthorn & Dolson, 2004; Gregersen, Kowalsky, & Li., 2007; Gregersen, Kowalsky, Kohn, & Marvin, 1999; Schellenberg & Trehub, 2008). Tonal-language exposure, where different tones distinguish different lexical meaning for the same word (Maddieson, 2013), has been repeatedly offered as an explanation for these higher rates (Deutsch, 2002, 2006; Deutsch et al., 2004; Levitin & Rogers, 2005; Pfordresher & Brown, 2009). The rate of absolute pitch is suggested as being between three and five times higher in Asian musicians than non-Asian musicians (Deutsch et al., 2006; Gregerson, et al., 1999; Gregerson et al., 2007).

Chapter 4 is based on the premise that native tonal speakers might have an advantage in absolute pitch. Firstly, exposure to their everyday language may provide intensive training during the critical period, as the participants with high accuracy in absolute pitch tasks are often those who began their musical training in the critical period (Bachem, 1955; Bermudez & Zatorre, 2009; Deutsch et al., 2004;
Levitin & Rogers, 2005; Miyazaki, 1988; Profita & Bidder, 1988; Russo, Windell & Cuddy, 2003; Sergeant, 1969; Takeuchi & Hulse, 1993). Secondly, native tonal language speakers have had extended practice with pitch identification in labelling their language throughout their lives, just as higher rates of absolute pitch are demonstrated in participants who have had prolonged musical training (Bachem, 1955; Baharloo et al., 1998; Profita & Bidder, 1988; Revesz, 1953).

In order to investigate whether native tonal language speakers have an advantage in pitch identification, I conducted an experiment in Sichuan, China made possible by a Keats Endowment Research Grant. The experimental task was a standard absolute identification procedure using tones of varying frequency. Hence, not only could a direct feasible comparison be drawn between the results from previous absolute identification experiments, but the methodological limitations of many absolute pitch experiments would be avoided because of the rigorous experimental procedures of the absolute identification paradigm. Chapter 4 describes this experiment, in which sixty university-enrolled native tonal-language (Mandarin) speakers, both musicians and non-musicians, completed this task. It was expected that there would be a higher proportion of participants exceeding the identification of Miller’s (1956) upper limit of nine stimuli, compared with the previous findings in absolute identification tasks completed in English-speaking countries. It was also predicted that, due to the prolonged exposure to labelling tones of varying frequency, the musicians would have a further advantage.
Chapter 5: Do explicit instructions about the pairing of stimuli and labels increase novices’ learning ability in absolute identification practice tasks?

Chapter 5 describes the further investigation of how far learning is possible in absolute identification tasks. Although for a long while, learning beyond Miller’s (1956) upper limit of nine stimuli was not thought possible in absolute identification tasks, even with experimental manipulations (e.g. Garner, 1953; Pollack, 1952; Braida & Durlach, 1972; Lacouture et al., 1998; Weber et al., 1977; Hartman, 1954), recent research has shown that learning beyond this limit is possible in exceptional cases (Rouder et al., 2004; Dodds et al., 2011; Rae, 2010). The commonality among the participants who have shown to exceed Miller’s limit through practice is that they have all started above average in their ability to identify uni-dimensional stimuli. They have also been more likely to have a complex psychological representation of the physically uni-dimensional stimuli.

The work in Chapter 5 was motivated by three factors. Firstly, Rae (2010) revealed that the participants who were able to exceed Miller’s (1956) limit in identifying notes of varying frequency were also musically proficient, having had years of musical training. However, the relationship between musical proficiency and the ability to identify tones of varying frequency is not clear. That is, it is not clear which of these causes the other. In absolute pitch literature, Levitin (1994) suggests that the essential aspect of learning to identify tones of varying pitch is the paired association with the note name. This chapter assumes that if this is the case for at least some of the other modalities in absolute identification, most people might be poor at identifying uni-dimensional stimuli because they have never needed to do so. In other words, they lack the training, not the fundamental ability.
The second motivation for Chapter 5 was the need for replication of exceptional cases in a paradigm where very few people have demonstrated the ability to identify uni-dimensional stimuli beyond Miller’s (1956) upper limit of nine (Rouder et al., 2004; Dodds et al., 2011; Rae, 2010). The third, and final, motivation for this chapter was to incorporate the use of no-feedback blocks into the absolute identification learning task. Although feedback has typically been used in absolute identification tasks, it introduces the possibility of allowing the participants to use relative judgement strategies in addition to, or instead of, absolute judgement. Rae (2010) demonstrated that blocks with feedback are more accurately identified in tones of varying frequency, but the work described in Chapter 5 also extended this possibility to lines of varying length. Chapter 5 describes an experiment that incorporated the above three aims, where nine highly motivated participants were recruited to attempt to learn tones of varying frequency and lines of varying length.

Chapter 6: General Discussion

The final chapter of this thesis, Chapter 6, discusses the findings of the previous chapters to attempt to answer the questions posed in this Introduction: To what extent is the ability to exceed Miller’s (1956) limit learned? How far beyond Miller’s limit is it possible to go? What quality do people have who can learn beyond nine stimuli?
Chapter 2
Using a Structural Forms Algorithm to investigate the psychological representation of unidimensional stimuli

As presented in Chapter 1, Miller’s (1956) famous limit of 7±2 stimuli was long thought to be the limit of the number of items that people could learn to perfectly identify in a uni-dimensional stimulus set in an absolute identification task. The fact that most people are still subject to this upper limit of nine is particularly surprising because it is resistant to many experimental manipulations, including extensive practice (e.g. Weber et al., 1977), the number of stimuli in the set (e.g. Garner, 1953) and stimulus spacing (e.g. Braida & Durlach, 1972). Most importantly, this limit appeared to be a fundamental aspect of human information processing rather than a sensory limitation, because the same limit applied to a wide range of stimulus modalities from electric shocks to saltiness (e.g., Lacouture et al., 1998; Pollack, 1952; Garner, 1953).

Despite this longstanding assumption that uni-dimensional stimuli are unable to be learned beyond an upper limit, recent work has identified exceptions. One of Rouder et al.’s (2004) participants was able to learn to perfectly identify 20 line lengths. Dodds et al.’s (2011) reported related learning effects not only for line lengths, but also for dot separation, line angle, and tone frequency. Rae (2010) reported two participants who were able to identify an average of 29.0 tones of varying frequency after 10 practice sessions. These findings contradict Miller’s theory of a small upper limit to memory processing capacity. This could represent an important finding because a small, or null, effect of learning has been included as a central element of many theoretical accounts of absolute identification and memory
(including: Stewart et al., 2005; Petrov & Anderson, 2004; Marley & Cook, 1984; and Brown et al., 2008). If Dodds et al.’s (2011), Rouder et al.’s (2004), and Rae’s (2010) results are taken at face value, they might imply that supposedly fundamental capacity constraints can be altered by practice.

There is, however, an alternative explanation. The number of stimuli that can be reliably identified increases exponentially as the number of dimensions increase (Eriksen & Hake, 1955; Miller, 1956; Rouder, 2001), at least when those dimensions can be perceived independently as separable dimensions (Nosofsky & Palmeri, 1996). For example, people are able to identify hundreds of faces, names and letters, all of which vary on multiple dimensions. Or, if an observer could perfectly identify say, seven line lengths and also seven angles, they might be able to identify 49 different stimuli with these combined features, such as circle sectors. With an additional assumption, this line of reasoning might reconcile the learning effects observed by Rouder et al. (2004), Dodds et al. (2011), and Rae (2010) with the long-standing results of Miller (1956). The extra assumption that is required is that some stimulus sets which vary on just one physical dimension might nevertheless invoke a more complex psychological representation. As with physically multi-dimensional stimuli, more complex psychological representations support richer percepts, perhaps allowing multiple ways to estimate the magnitude of a stimulus and hence better identification.

The stimuli used in absolute identification always vary on just one physical dimension, but this does not guarantee that the corresponding psychological representations are uni-dimensional continua. For example, perceived hue is represented either on a circle or a disc (Shepard, 1962; MacLeod, 2003) and the psychological representation of pitch is a helix (Bachem, 1950) even though the
corresponding physical stimuli vary on only one dimension (wavelength, in both cases). In Dodds et al.’s (2011), Rouder et al.’s (2004), and Rae’s (2010) studies, it might have been that those exceptional observers who learned to identify stimuli beyond Miller’s (1956) limit managed this feat by constructing more complex psychological representations for the uni-dimensional stimuli. If these observers had access to percepts on dimensions that are even partially independent, this could explain their improved performance without challenging Miller's long-standing hypothesis that performance on any single dimension is severely limited.

**Examining Psychological Representation**

In the absence of additional evidence, there is an unsatisfying circularity to this argument. The only evidence that suggests that these physically uni-dimensional stimuli have more complex psychological representations, is that those same stimuli can be learned. The only tested prediction from the hypothesised complex representation is that those same stimuli can be learned well. One method of independently probing psychological representation is to use multi-dimensional scaling (Cox & Cox, 1994; 2001). Multi-dimensional scaling determines relationships between objects by examining estimates of the perceived similarity of pairs of the objects. In some cases, such as with colour, multi-dimensional scaling techniques are able to reliably infer the complex psychological representation extracted from apparently uni-dimensional stimuli. This success presumably depends on the clear and consistent form of the representation across different people – allowing data to be averaged across subjects. In turn, the consistency of the psychological representation across subjects is probably an upshot of the basic physiology of the retina. In less clear-cut cases multi-dimensional scaling is not
always sensitive to subtle or inconsistent changes in the form of psychological representations.

Dodds, Donkin, Brown and Heathcote (2010) collected similarity ratings for line lengths, which was one of the stimulus types that Rouder et al. (2004) and Dodds et al. (2011) identified as an exception to Miller's (1956) limit. Dodds et al. (2010) found that multi-dimensional scaling was not reliably able to distinguish between one- and two-dimensional representations. In contrast, Rae (2010) used multi-dimensional scaling and found that the participants who exceeded Miller’s limit in a pitch identification task were shown to have a helical psychological representation of tones varying in frequency. The problem, however, is that multi-dimensional scaling lacks a framework for inference about these arrangements. This means that, if one wishes to recover the number of dimensions that best represent a relationship between objects, the conclusions are based on subjective judgements. Lee (2001) investigated this problem in detail and found that, for one- or two-dimensional representations, multi-dimensional scaling correctly identified the number of dimensions only 14 per cent of the time.

A recent advance in estimating the structure of psychological representations provides an alternative to multi-dimensional scaling. Kemp and Tenenbaum (2008) developed an algorithm to infer the structure of psychological representations based on relational data. Their method is based on a universal grammar for generating graphs, and the generality of those graphs allows the algorithm to represent structures as varied as trees, hierarchies, and points in vector spaces (as in multi-dimensional scaling; see Figure 2.1). An important benefit of Kemp and Tenenbaum's algorithm, that is not available with multi-dimensional scaling, is that it includes a coherent framework for inference, allowing probabilistic comparison of
different structural forms based on penalized likelihood, where the penalty term
depends on structural complexity.

Figure 2.1. (Figure 2 from Kemp & Tenenbaum, 2008) A hypothesis space of
structural forms. (A) Eight structural forms and the generative processes that produce
them. Open nodes represent clusters of objects: A hierarchy has objects located
internally, but a tree may only have objects at its leaves. The first six processes are
node-replacement graph grammars. Each grammar uses a single production, and
each production specifies how to replace a parent node with two child nodes. The
seed for each grammar is a graph with a single node (in the case of the ring, this
node has a self-link). (B–D) Growing chains, orders, and trees. At each step in each
derivation, the parent and child nodes are shown in grey. The graph generated at
each step is often rearranged before the next step. In B, for instance, the right side of
the first step and the left side of the second step are identical graphs. The red arrows
in each production represent all edges that enter or leave a parent node. When
applying the order production, all nodes that previously sent a link to the parent node
now send links to both children.
The work reported in this chapter used Kemp and Tenenbaum’s (2008) algorithm to investigate the psychological representation of the stimuli used in absolute identification experiments. The analysis was limited to undirected graph structures only, on the assumption that the similarity of two stimuli should not depend on the order of comparison (or, if it did, that this dependence was not of primary interest). The search was also limited to just two of Kemp and Tenenbaum’s forms – the chain and ring (see Figure 2.2). Chain structures are the standard assumptions for absolute identification stimuli: one-dimensional continua, where the psychological distance between stimuli is found by summing the distance from one neighbour to the next, and the next again, and so on. Ring structures represent just a small increase in complexity from chains, capturing the additional property that stimuli near one end of the set might be perceived to have something in common with stimuli at the extreme other end. This kind of relationship is found in both of the well-known cases of physically uni-dimensional stimuli having multi-dimensional psychological representations: Long wavelength light has a perceived hue (red) which is similar to the hue perceived for short wavelength light (violet; Kemp & Tenenbaum, 2008); similarly, the lowest frequency note in an octave (A) is perceived as similar to the highest (G#; Bachem, 1950).
Figure 2.2. An illustration of a (a) chain structure and (b) ring structure for lines of varying length. Note that in the chain structure stimuli 1 and 7 are far apart, while in the ring structure they are much closer. Based on the chain and ring structures shown in Figure 2 in Kemp and Tenenbaum (2008).

Data

A direct way to investigate psychological structure relies on similarity estimates obtained by direct interrogation where participants are presented with two stimuli and asked to rate their similarity on some scale, as was done in Rae (2010). Such ratings have many problems. Firstly, there is a severe limit on sample size because participants find it difficult to give many repetitions of these responses. Secondly, the numerical similarity ratings provided by participants depend on the experimenter's choices. For example, different ratings would be provided if the observers are asked to rate similarity from 1-10 or from 0-100, or on a Likert scale, and the precise nature of this dependence is unclear. Even more troubling, it is unclear whether similarity ratings obtained by this method are based on the particular psychological representation of interest: that is, the one underlying absolute identification performance. In this case, to circumvent all three problems, similarity judgments were replaced with confusion matrices calculated from many thousands of
absolute identification trials. These confusion matrices encode how often each pair of stimuli are confused with each other (e.g., when stimulus A is presented, what is the probability that it is identified as stimulus B?). Figure 2.3 shows a simple hypothetical confusion matrix. The assumption is thus that the probability of confusing two stimuli is monotonically related to their similarity.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>67</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>19</td>
<td>58</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>15</td>
<td>57</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>7</td>
<td>11</td>
<td>26</td>
<td>13</td>
<td>11</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>27</td>
<td>23</td>
<td>15</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>13</td>
<td>13</td>
<td>17</td>
<td>14</td>
<td>12</td>
<td>16</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>43</td>
<td>12</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>57</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>J</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>63</td>
<td>9</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>11</td>
<td>71</td>
</tr>
</tbody>
</table>

Stimulus

*Figure 2.3.* An example of a hypothetical confusion matrix. The responses to each of the stimuli are plotted on this matrix so that the correct responses are seen from the top left to the bottom right diagonally. For example, the 3 in the top row and fifth column represents the number of times stimulus E was presented and a participant responded by labelling that stimulus as A.

**Method**

Confusion matrices were calculated using the data from four absolute identification experiments reported by Dodds et al. (2011; see Table 2.1). In all four experiments, participants were given extensive practice over a series of 10 sessions.
leading to around 5,000 observations per participant. Each experiment included five or six participants. Three of the experiments used a smaller number of stimuli (15 or 16) allowing for unconfounded comparison between different stimulus types. These three experiments included the only one in which participants did not exceed Miller's (1956) limit of 7±2 stimuli (tone intensity) and two in which they did (line length and dot separation). The other experiment used 30 line lengths. This experiment was included because it showed some of the greatest improvement in performance with practice. However, a direct comparison between smaller and larger set size experiments must be looked upon with caution because of the varying statistical reliability of the data sets. The larger set sizes resulted in one quarter as many observations contributing to each element of the confusion matrix – as few as three observations per matrix element. The penalized complexity used in Kemp and Tenenbaum's (2008) algorithm means that noisier data lead to a preference for simpler structures – a bias towards identifying chain structures, in this case. This might extend to the smaller set size experiments, because even with 5,000 observations in the smaller set sizes, the average number of observations contributing to each confusion matrix element was between 16 and 20. The Discussion section of this chapter returns to this point.
Table 2.1

Data sets used from Dodds et al. (2011)

<table>
<thead>
<tr>
<th>Experiment*</th>
<th>Stimuli</th>
<th>Set Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Line Length</td>
<td>30</td>
</tr>
<tr>
<td>2b</td>
<td>Dot Separation</td>
<td>15</td>
</tr>
<tr>
<td>5a</td>
<td>Line Length</td>
<td>16</td>
</tr>
<tr>
<td>5b</td>
<td>Tone Intensity</td>
<td>16</td>
</tr>
</tbody>
</table>

*Note: Experiment refers to the experiment number as listed in Dodds et al. (2011)

The data from Rae (2010, Experiment 3) was also analysed. To recap from Chapter 1, six musically trained participants practiced with a set of 36 pure sine tones varying in frequency – their frequencies matched the fundamental frequency of the standard piano notes from A3 to G#6. The procedure for this experiment was similar to the procedure outlined for Experiment 6 in Dodds et al. (2011), except that responses were labelled not only with a number (1-36) but also the corresponding piano note name. There were 10 learning sessions providing 4,860 identifications per participant.

Results

Confusion matrices were constructed for individual participants, for a) their entire 10 hours of practice, b) the first 5 sessions of practice and c) the last 5 sessions. Data for individual participants were used as opposed to averaged data because of the small number of participants and large individual variation (see Table 2.2 for variation in accuracy). As described by Kemp and Tenenbaum (2008), feature data were simulated from the confusion matrices.
Version 1.0 (July 2008) of the Matlab implementation of the structural forms algorithm was used. For each confusion matrix, the best chain and best ring structure was identified, and their penalised likelihoods were recorded. In all cases, default values for the algorithm’s parameters were used. One modification was made to the algorithm, for numerical stability, restricting the search over edge lengths to disallow lengths that were extremely close to zero (smaller than $10^{-10}$). Note that this restriction still permits edge lengths of precisely zero, because adjacent stimuli can be collapsed into single nodes using the rules of the graph grammar. This restriction only disallows extremely small, but non-zero, separation between stimuli. In Table 2.2, the differences are reported in penalized log-likelihood between the chain and ring structure fits. To put the likelihood results in statistical perspective, differences in log-likelihood can be used to approximate the posterior probability that one model out of the pair (chain or ring) was the data generating model. This approximation should be interpreted with some care, as it relies upon some strong assumptions - for example, that the data generating model was one of the pair under consideration (e.g., Raftery, 1995). Nevertheless, in using this interpretation a difference in log-likelihood of two units corresponds to about three-to-one odds in favour of one model over the other, and a difference of six units in log-likelihood to better than twenty-to-one odds.

**Whole Data Sets**

The analyses are reported for smaller set sizes (15 or 16 stimuli) separately from the larger set sizes (30 or 36 stimuli). This allows cleaner comparison within each group because the number of data per entry in the confusion matrices are
comparable: about 18 observations per entry for the small set sizes, and about 4 for the large set sizes.

**Small Set Sizes**

Small set size experiments included those that used 16 tones of increasing loudness, 16 line lengths and 15 dots varying in separation. Participants who practiced tone loudness did not improve their performance much with practice, and their confusion matrices were also unanimously better described by chain structures than ring structures (see Table 2.2, where positive log-likelihood differences imply support for chain structures over ring structures). These results are consistent with Miller's (1956) original hypothesis that absolute identification is subject to a severe capacity limit when the stimuli really are uni-dimensional.

In comparison, the confusion matrices exhibited for some of the participants who practiced 16 line lengths and 15 dot separations were better described by the ring structure than the chain. In these two experiments, the ring structure was deemed more likely for only about half of the participants (5 of 11; see Table 2.2). The support for a ring structure is even more surprising when considering that that data used in these experiments were confusion matrices rather than similarity ratings. For example, if asked for a similarity rating, a participant might rate the extreme edge stimuli as very similar, while being unlikely to confuse those stimuli in an identification experiment. This presumably biases the results towards the chain structure, and yet several participants were still better described by ring structures.

Those five participants for whom the ring structure provided a better description in these experiments also demonstrated higher initial identification performance, and more improvement with practice. At the beginning of practice (first session), their mean accuracy was 54%, compared with 44% for the
participants better described by chain structures. Further, over the course of practice, those subjects identified as having ring-like representations improved their identification performance by 32% compared with 29% for the chain-like participants. Inferential tests were not calculated on these differences due to the very small number of participants (five in one group, six in the other).

**Large Set Sizes**

Table 2.2 shows accuracy and log-likelihood differences for experiments with 30 line lengths and 36 tone frequencies. Only four of the twelve participants demonstrated greater likelihood for a ring structure compared to the chain structure. As with the smaller set size experiments, those who demonstrated a ring structure demonstrated greater improvement in performance ($M_{ring} = 0.36$) compared to those who demonstrated a chain structure ($M_{chain} = 0.22$) as well as greater pre-practice performance.
Table 2.2

Accuracy and log-likelihood values for each participant in each of the five experiments

<table>
<thead>
<tr>
<th>Experiment (Stimuli)</th>
<th>Participant</th>
<th>Initial Accuracy</th>
<th>Improvement in Accuracy</th>
<th>Overall Log-Likelihood Difference*</th>
<th>Early Log-Likelihood Difference*</th>
<th>Late Log-Likelihood Difference*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tone intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(16)</td>
<td>1</td>
<td>0.34</td>
<td>0.10</td>
<td>8.3570</td>
<td>19.326</td>
<td>12.250</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.30</td>
<td>0.19</td>
<td>20.092</td>
<td>18.705</td>
<td>25.209</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.33</td>
<td>0.12</td>
<td>12.508</td>
<td>26.442</td>
<td>17.106</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.27</td>
<td>0.09</td>
<td>24.007</td>
<td>27.554</td>
<td>20.384</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.34</td>
<td>0.08</td>
<td>26.205</td>
<td>23.19</td>
<td>22.165</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.31</td>
<td>0.13</td>
<td>19.078</td>
<td>19.527</td>
<td>17.521</td>
</tr>
<tr>
<td><strong>Line lengths</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(16)</td>
<td>1</td>
<td>0.58</td>
<td>0.22</td>
<td>−2.542</td>
<td>0.120</td>
<td>−5.116</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.51</td>
<td>0.38</td>
<td>−3.423</td>
<td>0.171</td>
<td>−4.711</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.41</td>
<td>0.31</td>
<td>7.041</td>
<td>12.733</td>
<td>4.816</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.40</td>
<td>0.28</td>
<td>3.344</td>
<td>9.770</td>
<td>−0.223</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.46</td>
<td>0.30</td>
<td>5.143</td>
<td>13.69</td>
<td>−2.046</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.57</td>
<td>0.24</td>
<td>−3.365</td>
<td>−3.874</td>
<td>−4.798</td>
</tr>
<tr>
<td><strong>Dot separation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15)</td>
<td>1</td>
<td>0.44</td>
<td>0.20</td>
<td>7.852</td>
<td>10.378</td>
<td>8.621</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.51</td>
<td>0.37</td>
<td>−0.877</td>
<td>4.088</td>
<td>−5.481</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.53</td>
<td>0.27</td>
<td>4.472</td>
<td>7.229</td>
<td>2.468</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.39</td>
<td>0.37</td>
<td>5.703</td>
<td>11.38</td>
<td>−12.56</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.53</td>
<td>0.41</td>
<td>−2.658</td>
<td>2.360</td>
<td>−1.536</td>
</tr>
<tr>
<td><strong>Line length</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(30)</td>
<td>1</td>
<td>0.21</td>
<td>0.26</td>
<td>24.556</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.18</td>
<td>0.10</td>
<td>13.405</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.29</td>
<td>0.47</td>
<td>−1.426</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.20</td>
<td>0.10</td>
<td>24.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.31</td>
<td>0.41</td>
<td>57.852</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.17</td>
<td>0.24</td>
<td>−22.265</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tone frequency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(36)</td>
<td>1</td>
<td>0.40</td>
<td>0.29</td>
<td>−4.273</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.59</td>
<td>0.31</td>
<td>−0.394</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.20</td>
<td>0.13</td>
<td>27.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.21</td>
<td>0.09</td>
<td>24.934</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.19</td>
<td>0.08</td>
<td>22.549</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.22</td>
<td>0.14</td>
<td>0.634</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note that difference values are calculated by subtracting the likelihood values for ring structures from the likelihood values for chain structures, so that positive log-likelihood differences imply support for chain structures over ring structures.
Effect of Practice

Dodds et al. (2011) noted that participants improved their performance markedly given practice at absolute identification for all stimulus sets except for tones varying in intensity. In order to examine whether the improvement in performance was associated with a change in psychological structure, the confusion matrices were examined for each participant in the small set size experiments separately for early (1:5) and late (6:10) practice sessions (See Table 2.2). This split in the data was not examined from the large set size experiments because the sample size was too small – an average of fewer than two observations per entry.

For those who practiced tone loudness (Table 2.2) there was no difference in the estimated structure between early and late sessions: The data from every participant, for both early and late sessions, were always better described by chain structures than rings. For those who practiced line length or dot separation (Table 2.2), the chain structure was also dominant for early sessions (10 out of 11 participants). For six participants however, the most likely structure changed from a chain to a ring from early to late sessions. Three participants demonstrated a chain structure both in the early sessions and in the late sessions, and one other demonstrated a ring structure in both early and late sessions. No participant demonstrated the reverse switch – from ring to chain structure. Consistent with the hypothesis that high performance in absolute identification is only possible through more complex psychological representations, the single participant who demonstrated a ring structure during early practice also had very high performance in early practice. As well, the three participants who demonstrated a chain structure even late in practice were amongst the poorest performers late in practice.
A repeated theme in the above findings is that more complex (ring) structures are associated with better identification and with more improvement with practice. To investigate this more formally, the correlation between both improvement in performance and initial accuracy was calculated, and log-likelihood differences between ring and chain structures. Both improvement in accuracy and initial accuracy demonstrated a strong negative relationship with log-likelihood difference values, where smaller log likelihood differences (representing a preference for a ring-structure) was associated with greater overall improvement in accuracy ($r = -.70, p < .001$) and greater initial performance ($r = .65, p < .001$; see Figure 2.4).

![Figure 2.4](image_url)

*Figure 2.4.* Accuracy and improvements in accuracy as a function of difference in Log-Likelihood values for ring and chain structures (where a negative Log-Likelihood value indicates a preference for a ring structure). Note: Two outliers were removed from this analysis (where log-likelihood difference was < -10 and >30).

**Discussion**

For more than fifty years, absolute identification with uni-dimensional stimulus sets has been assumed to be subject to a strict performance limit, Miller’s (1956) magical number 7±2. More recently, Rouder et al. (2004) and Dodds et al.
(2011) have shown that some stimulus sets support much greater performance than this limit (including line length or angle, and tone frequency) while at least one does not (tone intensity). One way to reconcile these newer findings with previous literature is to hypothesize that some stimulus sets, while physically varying on only one dimension, give rise to a more complex psychological representation. The data from Dodds et al. (2011) and Rae (2010), and the structural forms algorithm developed by Kemp and Tenenbaum (2008) provide a method for investigating this hypothesis in a way that was not previously possible because of limitations in analytic tools such as multi-dimensional scaling.

The results provide consistent support for the previously untested hypothesis that improved identification performance is only possible with more complex psychological representations. When data was examined from the identification of tones varying in intensity (for which identification performance was severely limited) a uniformly strong support for the simplest uni-dimensional psychological representation was found – a chain, as assumed in all theoretical accounts of identification. This result was observed for all participants, and was also confirmed as the most likely structure in both the early and late practice data. Data from those stimulus sets for which Dodds et al. (2011) and Rae (2010) found significantly improved performance with practice yielded different results. The psychological representations of the stimuli for the more than one third of these participants (9 of 23) were better described by the ring structure than the chain structure. This figure rose to 8 of 11 participants when only data from the second half of practice were considered, in stark contrast to the 1 of 11 participants identified as using a ring structure in the first half of practice. The hypothesized relationship between identification performance and structure was further supported by strong correlations
between performance in practice and the goodness-of-fit of the ring and chain structures.

To check these results with data from another laboratory, data from two of Rouder et al.’s (2004) participants was also analysed. Those participants practiced line length stimuli in a similar procedure to that described above, using set sizes of 13 line lengths in one experiment and then 20 line lengths in another. Both participants in the first experiment, and one out of two of the participants in the second experiment were better described by the Kemp and Tenenbaum’s (2008) ring structure. The data from Rouder et al’s 30-length experiment was not analysed, as the sample sizes became prohibitively small. The participants that demonstrated a more complex structure were also those that demonstrated higher initial accuracy ($M_{ring} = .85; M_{chain} = .68$).

A natural question arising from these analyses is why uniform results were not observed. That is, if improved performance in the identification task really is supported by more complex psychological representations of the stimuli, why these representations were not observed for every participant. Two explanations seem plausible. First of all, in all experiments there was considerable variability amongst the participants in identification performance. About half of the participants did not learn to improve their performance beyond Miller's (1956) limit of 7±2 stimuli. This is consistent with the hypothesis that those participants should maintain the simplest (chain) psychological representations. Secondly, there is an inherent bias favouring the chain structure over the ring structure in noisy data. This bias arises because general noise (such as non-task-related responses, and random error) bias the confusion matrices towards uniformity, and uniform confusion matrices are –
according to the structural forms algorithm – better described by chain than ring structures due to the higher complexity penalty attracted by ring structures.

The results indicate that better performance through practice in identification is associated with more complex psychological representations of stimuli. However, the results do not provide insight into exactly how those representations arise, nor what extra stimulus information is being represented. For example, it is easy to speculate that participants might learn to judge line lengths using information from several sources – perhaps the extent of the retinal image, or the magnitude of the saccade needed to traverse the line, or even cues gained by comparing the line to external objects such as the display monitor. Magnitude estimates obtained from these sources would presumably be highly, but not perfectly, correlated, which could result in psychological representations more complex than chains. Further study might examine such hypotheses by attempting to limit the information available from such cues, for example by presenting visual stimuli using virtual reality goggles.

In summary, it seems that tone loudness was the only stimulus modality that showed consistent evidence for only a single underlying psychological dimension. Line length, dot separation and tone frequencies showed evidence for more complex psychological representations than simple chain structures - particularly for highly-performing participants and post-practice data. The implications of these results are remarkable for the study of memory in terms of absolute identification – if these stimuli are truly represented on multiple dimensions, uni-dimensional absolute identification does not apply to these stimuli. In the extreme, it might be that the long history of study of uni-dimensional absolute identification should have been limited to the study of tones varying in loudness. Or in the very least, that the identification
of other stimulus types only qualifies as uni-dimensional as long as participants are not well practiced.
Chapter 3

Do overtones and tones of the Western musical scale allow for better absolute pitch identification?

Chapter 1 of this thesis outlined the background of absolute pitch identification within the framework of the paradigm of absolute identification. Like absolute identification, absolute pitch requires participants to label stimuli that vary on a physical dimension with previously assigned labels, in this case tones of varying frequency. The discrete frequencies used in absolute pitch experiments usually correspond with the current Western musical scale (Takeuchi & Hulse, 1993; Yost, 2009). In absolute pitch experiments, tones are usually labelled with their musical label (chroma and octave; e.g. C#3) while in absolute identification experiments they are labelled with an arbitrarily assigned numeral label.

Chapter 1 also summarised the findings from Rae (2010) that used the absolute identification paradigm to investigate learning capacity and the relationship between psychological representation of stimuli and the ability to identify tones of varying frequency. Rae (2010) found two key points that inspired the current chapter. Firstly, although there was a strong positive correlation in performance, the results showed that piano tones were more accurately identified than pure sine tones. The second finding was that participants with more advanced musical proficiency were more accurate in identifying both piano and sine tones. This chapter investigates two questions that arose from those findings. Firstly, it examines whether learning plays a role in the identification of tones of varying pitch, and secondly, it examines whether the multi-dimensional nature of the piano tones gave rise to superior identification compared with pure sine tones.
Does learning play a role in the identification of tones of varying pitch?

Research has shown that apart from a few very rare individuals (Dodds, et al., 2011; Takeuchi & Hulse, 1993), most people are unable to identify more than Miller's (1956) upper limit of nine stimuli of tones of varying frequency (pitch; e.g. Hartman, 1954; Pollack, 1952). And while it was previously thought absolute pitch was an ability unable to be learned (e.g. Revesz, 1953; Stumpf, 1883 in Takeuchi & Hulse, 1993), recent research discussed in previous chapters has suggested that some kind of learning is possible (e.g. Dodds, et al., 2011; Levitin, 1994; Levitin & Rogers, 2005; Takeuchi & Hulse, 1993). Evidence of this comes from studies showing absolute pitch possessors are not always uniform across the identification of tones, instead more accurately identifying notes they have had more exposure to (Athos et al., 2007; Boggs, 1907; Brammer, 1951; Lundin, 1963; Miyazaki, 1988; Pantev et al., 2000; Takeuchi & Hulse, 1993).

Rae (2010) revealed a correlation between musical proficiency and a superior ability to identify tones of varying frequency. However, what was not clear was whether the musical proficiency caused the better identification of tones, or if an ability to identify the tones contributed to an increased musical proficiency. One way this could be investigated is to compare the identification of the tones familiar to musicians with tones that are not. Musicians are greatly exposed to the discrete tone frequencies that correspond with the Western musical scale through many years of practice. To a lesser extent, non-musicians are also exposed to these tones through popular culture media such as music and film.

If it is true that learning plays a role in the ability to identify tones of varying frequency, then comparing the identification of tones from the Western musical scale with tones not found in the Western musical scale (non-Western) might help to
identify the effects of training. If the ability to identify tones of varying frequency is, at least in part, a learned skill then it would stand to reason that Western tones would be more accurately identified than non-Western tones, particularly for musicians.

**Are piano tones better identified than sine tones because of their multi-dimensional nature?**

The second motivation for the work in this chapter was the finding from Rae (2010) that even in a strict experimental absolute identification setting, piano tones were more accurately identified than pure sine tones. Previous studies had suggested this also, although the methodologies used were not as scientifically rigorous as the paradigm of absolute identification (Takeuchi & Hulse, 1993). Rae (2010) showed that while there was a strong positive correlation between the identification of piano tones and sine tones, piano tones were still identified more accurately.

While learning may be a reason for the superior identification of tones, as described above, the physical multi-dimensionality of the piano tones may also be considered as a reason for their superior identification as has been done in past research (e.g. Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983). Piano tones have additional cues to frequency such as variations in timbre, volume, resonance and decay (Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983). Each physical dimension that is added to a uni-dimensional stimulus exponentially increases the ability to identify that stimulus (Eriksen & Hake, 1955; Miller, 1956; Rouder, 2001), as long as those dimensions can be perceived as separate from each other (Nosofsky & Palmeri, 1996).

As well as additional independent physical dimensions assisting in the identification of uni-dimensional stimuli, the psychological representation of the
stimuli has also been shown to give an advantage. Chapter 2 of this thesis discussed how in using a Structural Forms algorithm (Kemp & Tenenbaum, 2008) it was apparent that better pitch identification, amongst some other modalities in absolute identification, correlated with a more complex psychological representation of the physical uni-dimensional stimuli. Also, Rae (2010) reported that the participants who were able to exceed Miller’s (1956) limit in the pitch identification task for both piano and sine tones were shown to have a helical psychological representation of both tone types using multi-dimensional scaling techniques (Cox & Cox, 1994; 2001; Shepard, 1974) based on similarity ratings, although the limitation of such techniques have been discussed in the previous chapter.

While a more complex psychological representation was shown to correspond with superior pitch identification in general (Rae, 2010), the question still remains as to whether the additional physical dimensions in the piano tones were responsible for their superior identification. Each musical tone consists of a fundamental frequency and a series of overtones. For example, Middle A on the piano (the note that many orchestral instruments tune to) consists of a fundamental frequency of 440Hz, with overtones at the multiples of the fundamental frequency so that the first overtone is 880Hz, the second is 1,320Hz, the third is 1,760Hz and so on (Yost, 2009).

Miller (1956) and others stated that additional physical dimensions were able to exponentially increase the identification of the stimuli only when those dimensions were independent (Eriksen & Hake, 1955; Rouder, 2001; Nosofsky & Palmeri, 1996). However, given that the perception of a musical note with its many overtones is the fundamental frequency alone, it may not be the case that the additional overtones add independent dimensions that assist in the identification of the tone. Nevertheless, Rae (2010) showed that piano tones were still better
identified than sine tones, which perhaps suggests that although the additional overtones were not physically perceivable, they might be psychologically separable. A way to test this under laboratory conditions with careful stimulus control is to compare the identification of pure sine tones (uni-dimensional with a base frequency alone) with pure sine tones with an additional overtone added.

The motivation for the work described in this chapter was to investigate the two issues discussed above. Firstly, whether learning plays a role in the identification of tones of varying frequency; that is, whether musical proficiency correlates with superior pitch identification only for tones found in the current Western musical scale. Secondly, whether piano tones are better identified because of the additional information provided in overtones. In order to examine these points, computer generated sine tones were used, both corresponding to the Western musical scale as well as to tones not found in the Western musical scale. Further, both sets of tones, Western and non-Western, were split into either having a base frequency alone (uni-dimensional) or with one additional overtone (multi-dimensional; additional information in the Method section of this chapter).

It was predicted that if learning plays a significant role in pitch identification ability, then Western tones should be better identified than non-Western tones, particularly for musicians. It was also predicted that if an overtone in addition to the fundamental frequency of a tone provides an extra psychological cue, this should allow for the better identification of the tones with overtones, compared with the uni-dimensional pure sine tones.
Method

Participants

One hundred and thirty five participants (47 musicians) were recruited from an undergraduate psychology pool and from the general university population. Participants were reimbursed with a $25 EFTPOS voucher per hour or given course credit. Participants were classified as musicians if they had attained a level of at least 4th grade Australian Music Examination Board (AMEB) or equivalent. All other participants were classified as non-musicians. No participants reported having hearing difficulties.

Stimuli

Stimuli consisted of four sets of twenty computer generated sine tones. The first set of tones were based on Western tones, spaced a minor third (three semitones) apart ranging from 49Hz (G1) to 3,951Hz (B7). This spacing was chosen so that the frequency range in the Western scale was as similar as possible to the range in the non-Western tones. The second set of tones used the first set as fundamental frequencies and also had an overtone of one octave (twelve semitones; that is, double the fundamental frequency) - the first harmonic overtone that naturally occurs in piano tones. The third set of tones were spaced based on prime number fundamental frequencies, ranging from 50Hz to 3,970Hz, none of which are found in the current Western musical scale (see Table 3.1). The fourth set of tones used the third set as fundamental frequencies and also had an inharmonic overtone 2.7 times the frequency of the fundamental tone.

All tones were created by Professor Roger Dean from MARCS Institute (University of Western Sydney) and were presented via Sony Stereo Headphones.
The experiment was conducted using Matlab R2010b. Participants used a mouse to respond. Participants also completed a questionnaire about their musical experience, language skills, hearing problems, and absolute pitch skills (Appendix 1).

Table 3.1

<table>
<thead>
<tr>
<th>The frequencies (Hz) of the non-Western tones based on prime number fundamental frequencies. All tones were created by Professor Roger Dean from MARCS Institute (University of Western Sydney)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>290</td>
</tr>
<tr>
<td>1,130</td>
</tr>
</tbody>
</table>

**Procedure**

Participants completed a short questionnaire (Appendix 1) about their musical history before completing the experimental task. Each participant was randomly assigned to one of the four experimental conditions. At the start of the experimental task, there was a brief study phase to familiarise participants with the stimuli and the method of responding, which is standard in absolute identification tasks. Participants were visually presented with two rows of ten buttons on screen labelled with a unique number (1 through 20) in order of ascending frequency (1 through 10 on the top row of buttons, and 11 through 20 on the bottom row). Musical note names (e.g. G#) were not used at all during this experiment, so that no advantage was given for the Western tones, as the non-Western tones could not have had such labels as they did not correspond to notes in the Western musical scale. Participants were then
presented with each of the stimuli via headphones, one at a time in ascending frequency order. To move onto the following tone the participant had to select the correct button onscreen that corresponded to the stimulus label, after reading the following instructions, “This is pitch x. When you hear this tone, press the x button.”

Following this, the test phase consisted of seven blocks of 100 trials for each participant. During each block, each of the 20 stimuli was repeated five times in random order. On each trial a fixation cross was presented for 500ms, before one of the 20 tones was randomly presented for one second. The participant was asked to respond with the label that was attached to the stimulus in the study phase. If participants were incorrect, the correct answer was displayed for 500 milliseconds. If they were correct, “Correct” was displayed instead. Feedback was provided after each trial during the first six blocks, whereas block seven gave no feedback.

**Results**

As is typical in tests of absolute identification, feedback was provided for anchoring and motivation (e.g. Rouder, et al., 2004). However, as discussed previously, providing feedback allows for the potential use of relative pitch strategies instead of, or in addition to, absolute pitch strategies. An analysis confirmed that feedback was providing participants with an advantage to identify the tones, as blocks with feedback were identified more accurately than blocks without feedback (34.3% accuracy and 27.5% accuracy respectively, $t(134) = 241.02, p < .01$). There were no statistically significant interactions with feedback for musicality (musicians vs. non-musicians, $p = .203$), tonality (Western vs. non-Western tones, $p = .158$), or dimensionality (fundamental frequency only vs. with additional overtone, $p = .098$). The remainder of the analyses that follow have been conducted only using the blocks
without feedback as this gives a more true account of absolute pitch identification by removing the possibility of correctly responding by relative tone judgement.

Musicians outperformed non-musicians in identifying tones (30.7% accuracy versus 25.9% accuracy respectively; \( t(134) = 13.81, p < .001 \)). Figure 3.1 shows the accuracy of pitch identification for the four conditions split by musicality (that is, whether participants were musicians or non-musicians). However, while musicians outperformed non-musicians overall, this was only seen for the Western tones, with a significant interaction \( (F(1,131) = 4.05, p = .046) \). Musicians and non-musicians performed at very similar levels of accuracy in identifying the non-Western tones. Overall there was not a significant difference between the absolute identification of Western and non-Western tones \( (t(134) = .165, p = .20) \).

Contrary to predictions, uni-dimensional tones were identified more accurately than tones with an overtone (28.85% accuracy and 26.25% accuracy respectively; although this different was not shown to be significant, \( t(134) = .481, p = .49 \)). Musicians were particularly better at identifying uni-dimensional Western tones, as Figure 3.1 shows, although this interaction was only shown to be approaching significance, \( F(1,127) = 3.78, p = .054 \).
Figure 3.1. The accuracy of musicians and non-musicians in identifying tones split by the conditions (Western uni-dimensional [1D], Western with harmonic [2D], non-Western uni-dimensional [1D], and non-Western with overtone [2D]). The parallel dotted lines shows Miller’s (1956) range of $7\pm2$ (in this case equating to .25 to .45 accuracy). No participant exceeded Miller’s limit without feedback in any of the four conditions (musician or non-musician). Error bars were calculated per condition using standard error calculations.

Discussion

The purpose of the work in this chapter was twofold. The first aim was to investigate whether learning plays a role in the identification of tones of varying pitch. This was in order to extend on Rae’s (2010) findings that participants with more advanced musical proficiency were more accurate in identifying tones of varying frequency. It was predicted that if absolute pitch is, at least in part, a learned skill, musicians would be better at identifying tones of varying frequency,
particularly those from the Western scale. The results replicated those of Rae and showed that overall, musicians outperformed non-musicians in identifying tones of varying frequency. However, this was only the case for the Western tones. These results suggest that musicians’ superior ability to identify tones of varying frequency might be, at least in part, learned from their long exposure to the Western tones during their many years of training. Chapters 4 and 5 in this thesis expand on this concept further.

The second aim of the work in this chapter was to investigate whether an additional overtone could increase participants’ ability to identify tones of varying frequency. This was in order to extend Rae’s (2010) finding that piano tones were more accurately identified than pure sine tones in an absolute identification framework. In this study it was considered that this could have been due to the physical multi-dimensionality of the piano tones compared with the uni-dimensionality of the pure sine tones (e.g. Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983), or perhaps to the psychological separability of the dimensions. However, the results showed that the additional overtone gave no additional information to the identification of the tones. Contrary to predictions, uni-dimensional tones were identified more accurately particularly by musicians, although these results did not show statistical significance.

In this study, the findings that uni-dimensional tones were identified more accurately, although not significantly, than multi-dimensional tones suggest that perhaps the rich cues in piano tones are more than simply additional overtones. As mentioned previously, piano tones are complex waveforms with marked variations in timbre, volume, resonance and decay characteristics across the registers of the piano (Lockhead & Byrd, 1981; Patterson, 1990; Terhardt & Seewann, 1983; Yost, 2009).
While research suggests that musicians are more accurate in identifying pitch on their own instrument (Pantev et al., 2000; Takeuchi & Hulse, 1993), perhaps the additional overtone in the multi-dimensional tones in this study did little to remind the participants of the musical instruments to which they are most familiar.

Miller’s (1956) statement that additional dimensions assist exponentially in the identification of stimuli only takes effect when those dimensions are perceived as independent (Nosofsky & Palmeri, 1996). The contrary to predicted results in this study, where tones with two dimensions were not identified more accurately than uni-dimensional tones, further support the results of Kemp and Tenenbaum (2008), Dodds, Rae, and Brown (2012; Chapter 2 of this thesis), and Rae (2010) that the psychological representation of the dimensions is likely more important than the physical dimensionality in the identification of the stimuli.

It is also possible that the inharmonic overtones of the non-Western tones, as well as providing no useful information, were actually providing confounding information. Given that the musicians in this condition were slightly less accurate in identifying these tones – even though not with statistical significance - it is possible that these tones were counterproductive to the musicians as they might have had to unlearn them to make use of the fundamental frequency. However, given that the accuracy in identifying the uni-dimensional non-Western tones was not much better, for both musicians and non-musicians, this seems unlikely.

To be consistent across the experiments in this thesis, blocks were presented both with and without feedback in this experiment. Blocks were given with feedback as they are typically used in absolute identification experiments, and were also given for anchoring and motivation (e.g. Rouder, et al., 2004). Blocks without feedback were given to minimise the potential use of relative pitch strategies instead of, or in
addition to, absolute pitch strategies. A preliminary analysis showed that feedback made identification of the tones of varying frequency significantly more accurate, and so blocks with feedback were not used in the final analysis. This allowed a more accurate representation of absolute pitch strategies, but also calls into question absolute identification experiments and their use of feedback. Both issues will be discussed in the final chapter (Chapter 6; General Discussion) of this thesis.

An issue that could have potentially confounded the results in this experiment was the spacing of the tones. Both the Western and non-Western tones had a similar frequency range, however the spacing was different for both tones. The non-Western tones were based on prime number fundamental frequencies and so there was not a regular interval spacing, whereas the Western tones were based on an even minor third spacing (three semitones; harmonic minor 7th arpeggio). While the use of relative pitch would have been difficult, the spacing of the Western tones was still regular and consistent compared with the more seemingly haphazard spacing of the non-Western tones. This still could have given a slight advantage to the identification of the Western tones, especially for the musicians, although this is considered unlikely.

Another possible issue that could have confounded the results is that the tones were not labelled with musical names. Non-musicians likely were not affected by this strategy but it could have made identification of musical notes more difficult for musicians. While it would not have made sense to label the non-Western tones with musical note names, this was an unavoidable issue as it was necessary to standardise the labels for Western and non-Western tones, and for musicians and non-musicians alike. However, given that musicians were better at identifying the Western tones, it seems that any potential confound caused by this was limited.
This study supported Rae’s (2010) findings that musicians outperformed non-musicians in identifying tones of varying frequency, and further extended this to show that musicians are better able to identify Western tones than non-Western tones. However, the results were contrary to the prediction that tones with overtones would be better identified than uni-dimensional tones. This suggests these dimensions were not psychologically separable, and were not responsible for the better identification of piano tones compared with pure sine tones.
Chapter 4
Do Tonal Language Speakers have an Advantage in Absolute Pitch identification?

One of the main questions this thesis has aimed to investigate is whether learning is at least in part responsible for the ability to identify uni-dimensional stimuli, such as tones of varying frequency, in an absolute identification framework. Specifically, the investigation has aimed to shed light on whether the participants who are able to exceed Miller’s (1956) upper limit of nine in identifying uni-dimensional stimuli have relied on learning as a strategy. Rae (2010) showed that participants who are musically proficient, that is people who had spent many years in musical training, were better able to identify tones of varying frequency. Further, the work in Chapter 3 of this thesis showed that musicians were better able to identify frequencies corresponding to the Western musical scale, both compared with non-musicians and with non-Western tones. Both of these findings suggest that learning has a role to play in the identification of tones of varying frequencies.

The next stage, described in this chapter, was to investigate whether learning plays a role in the identification of tones of varying frequency by examining a corresponding ability that has been reported as being advantageous to pitch identification: native tonal language fluency. Chapter 1 of this thesis presented the rates of absolute pitch as being very rare, with rates as low as 1 in 10,000 in the general population (Bachem, 1955; Profita & Bidder, 1988), while as high as 1 in 7 amongst the most highly-trained musicians (Revesz, 1953; Baharloo et al., 1998). These rates, however, are reported from studies that have focused on English-speaking populations. The incidence of absolute pitch is reported to be higher in Asia than it is in North America (Schellenberg & Trehub, 2008), and specifically amongst
Chinese musicians (Deutsch et al., 2006; Deutsch et al., 2004; Gregersen et al., 2007; Gregersen et al., 1999; Schellenberg & Trehub, 2008).

Tonal-language exposure has been repeatedly offered as an explanation for these higher rates (Deutsch, 2002, 2006; Deutsch et al., 2004; Levitin & Rogers, 2005; Pfordresher & Brown, 2009). In tonal languages, such as Mandarin, Cantonese, or Vietnamese, different tones distinguish different lexical meaning for the same word (Maddieson, 2013). For example, in Mandarin there are four different tones: high, mid-high and rising, initially low and descending and then rising, and high and falling. Deutsch et al. (2006) reported that the incidence of absolute pitch in Chinese music conservatories was approximately three times the rate of absolute pitch in music conservatories in the United States. Further, Gregerson et al. (1999) reported the rate of absolute pitch to be four times higher in Asian musicians, and Gregerson et al. (2007) reported the rate to be five times higher in Asian musicians, than non-Asian musicians.

A possible account of why native tonal language speakers have an advantage in absolute pitch is that exposure to their everyday language may provide two elements that have been shown to correlate with superior pitch identification. The first of these elements is early-life training of pitch identification, during a “critical period”, and secondly, extended practice at pitch identification. People with high accuracy in absolute pitch tasks are often those who began their training in the critical period (Bachem, 1955; Bermudez & Zatorre, 2009; Deutsch et al., 2004; Levitin & Rogers, 2005; Miyazaki, 1988; Profita & Bidder, 1988; Russo et al., 2003; Sergeant, 1969; Takeuchi & Hulse, 1993). Higher rates of absolute pitch are also shown amongst individuals who have had extended musical training (Bachem, 1955; Baharloo et al, 1998; Profita & Bidder, 1988; Revesz, 1953). However, exposure to
early musical training or extended training is not sufficient to develop absolute pitch, as most people who experience these conditions do not develop the ability (Baharloo et al., 1998).

However, while the above arguments provide a plausible reason to expect better absolute performance in speakers of tonal language, the link is not as clear as it seems at first. While tonal language speakers do use tones to differentiate meaning of the same syllables, the differences between these tones are defined by relative, not absolute, pitch. For example, in theory, what is a “high” tone for one speaker may be a “low” tone for another speaker. “High” tones are defined relative to surrounding phonemes, which is inconsistent with the idea of absolute pitch. Consistent with this idea, other studies have found no evidence for tonal-language speakers having an absolute pitch advantage. For example, Schellenberg and Trehub (2008) compared the performance of 9-12 year old Chinese participants with participants of European heritage of the same age. They required the participants to identify popular TV themes tuned up, tuned down, or at the original pitch. The results provided no support for tonal-language speakers having a pitch identification advantage.

The aim of the work in the current chapter was to test whether a large difference in the proportion of people with absolute pitch exists between native speakers of a tonal language compared with non-tonal language speakers. Previous research provides some conflicting reports on this matter. Deutsch and colleagues suggest that the potential to acquire absolute pitch is universally present at birth, and if the training to associate pitches with labels is sufficient during the critical period, then the person will develop absolute pitch (Deutsch, et al., 2004; Deutsch, et al., 2006). Deutsch et al. (2004) found that tonal language speakers (Mandarin and Vietnamese) displayed a stable form of absolute pitch in enunciating words
compared with English (non-tonal) speakers, and concluded that tonal language speakers have an absolute pitch template to produce and understand speech. Deutsch, Dooley, Henthorn, & Head (2009) also found similar results when testing Mandarin speaking participants from isolated villages in China.

Deutsch, et al. (2006) also reported evidence of the tonal language speakers having better pitch identification skills than non-tonal language speakers. As well as finding that earlier musical training was correlated with higher rates of absolute pitch, they found that Chinese music students had a higher incidence of absolute pitch than American music students. Extending these findings, Deutsch et al. (2009) found that the prevalence of absolute pitch among students in an American music conservatory was a function of age of onset of musical training, ethnicity, and fluency in speaking a tonal language. Similarly, Lee and Lee (2009) found that in an absolute pitch identification task with tonal Mandarin speakers there was a higher prevalence of absolute pitch compared with an earlier similar task with non-tonal language speaking participants (Lee & Hung, 2008). However, these studies did not use a standardised method for evaluating absolute pitch performance. As well, sampling constraints leave open the possibility that, for example, American students from families that speak tonal languages at home are not the same as American students from families that do not.

In the work presented in this chapter, an absolute identification test was administered to Mandarin-speaking students from a university in mainland China. Half the students were studying music, and half were not. It was expected that if native tonal language speakers have a pitch advantage due to their early exposure to pitch labelling, the proportion of people in the sample who could identify more than Miller’s (1956) upper limit of nine stimuli should be much larger than the proportion
reported for English-speaking populations. For the highly-trained musicians in this sample, this hypothesis corresponds to a rate of between 13% and 33% (i.e. between 4-5 times the rate in English speaking musicians, which has been reported as between 1 in 15 [Revesz, 1953] and 1 in 30 [Baharloo et al., 1998]. It should be noted that the rate of 1 in 15 is a conservative test, as Baharloo et al (1998) suggested that the rate might rise to 1 in 7 for highly accomplished musicians.

Method

Participants

Sixty native-tonal language speaking (Mandarin) university students from in Sichuan University in Chengdu, Sichuan, from The People’s Republic of China were recruited. Thirty of these were enrolled in a university music course, while thirty were enrolled in a social science course. Of the thirty music students, 17 began their musical training in the critical period. Of the social science students, 11 had completed some non-tertiary musical training, with six beginning their training in the critical period. Four participants reported having absolute pitch: three from the music group, and one from the social science group.

Stimuli & Apparatus

Stimuli consisted of 36 pure sine tones matching the fundamental frequencies of musical notes from A3 (220Hz) to G#6 (1,661Hz). Pure sine waveforms were generated using MatlabR2010b. These tones were identical to the ones used in Rae (2010). All tones were presented via Sony Stereo Headphones (MDR-XD100). The experiment was conducted using an LG Ultrabook with a 13” monitor and Matlab R2010b. Participants used an external mouse to respond. Participants also completed
a questionnaire that was presented in Mandarin Chinese, with questions related to musical experience, language skills, hearing problems, and absolute pitch skills (Appendix 2). Support from a Mandarin-English translator was available throughout the experimental process.

Procedure

Participants completed the questionnaire before completing the experimental task. At the start of the experimental task, there was a brief study phase to familiarise participants with the stimuli and the method of responding, which is standard in absolute identification tasks. Participants were visually presented with three rows of twelve buttons on screen labelled with the musical note name (A3, F#4, etc.) and a corresponding number from 1 (A3) to 36 (G#6) in order of ascending frequency. Both note names and numeric labels were used so participants could use whichever they were more comfortable with. Participants were then presented with each of the stimuli via headphones, one at a time in ascending frequency order. To move onto the following tone the participant had to select the correct button onscreen that corresponded to the stimulus label, after reading the following instructions, “This is pitch x. When you hear this tone, press the x button.”

Following this, the test phase consisted of six blocks of 108 trials for each participant. During each block, each of the 36 stimuli was repeated three times in random order. On each trial a fixation cross was presented for 500ms, before one of the 36 tones was randomly presented for one second. The participant was asked to respond with the label that was attached to the stimulus in the study phase. Feedback was given after each response, except during the last block of trials. If participants were incorrect, the correct answer was displayed for 500ms. If they were correct,
“Correct” were displayed. Feedback was provided to ensure that any poor performance on the task was not due to participants still learning the procedure, or other short-term deficits. Data from blocks with feedback were not used to measure absolute pitch, as the provision of feedback enables musically-trained participants to solve the task using relative pitch methods (Levitin & Rogers, 2005; Lockhead & Byrd, 1981). The task took approximately one hour to complete, including an enforced one minute break between blocks.

Results

The analyses of these data use both null hypothesis significance tests based on frequentist assumptions, and also Bayesian equivalents. The null-hypothesis significance testing (NHST) analyses, such as ANOVAs, are included for completeness and for comparison with previous research. The Bayesian analyses are better-suited to the research question at hand, as they allow for the direct quantification of evidence in favour of the null hypothesis, or against it. The Bayesian ANOVAs were calculated using the default settings from Morey and Rouder’s (2013) package for the R statistical language, with Bayes factors calculated by comparing a full model (that included all main effects and interactions) against models that drop each effect in turn. The Bayes factors are reported in the direction such that numbers greater than 1 indicate support for the presence of an effect, and numbers less than 1 indicate evidence in favour of the null (no effect).

An 2x2 mixed-design ANOVA showed that music students performed significantly better on the pitch identification task than social science students (F(1,56)=365.21, p < .001). However, this was only the case for blocks with feedback. Blocks with feedback had a significantly higher accuracy than blocks
without feedback ($F(1,14)=108.27, p < .001$), although this was only the case for the music students, which is confirmed by the significant interaction ($F(1,9)=70.32, p < .001$, see Figure 4.1).

The corresponding Bayesian ANOVA showed that there was a very strong effect of feedback, with blocks with feedback having higher accuracy than blocks without ($BF = 4.2 \times 10^{22}$; see Figure 4.1). Music students performed with higher accuracy than social science students, although there was only very strong evidence for this in blocks where feedback was given ($BF$ for school only $= 3.09$; $BF$ for school by feedback interaction $= 6.44 \times 10^{10}$).

Figure 4.1. A comparison of mean accuracy between music students and social science students, separated for blocks with and without feedback. Error bars were calculated per condition using standard error calculations. The parallel dotted grey lines represent Miller’s (1956) limit of $7\pm2$ stimuli.
Figure 4.2 shows individual performance for all participants, separately for blocks with and without feedback. No participant from the social science school (black symbols in Figure 4.2) exceeded Miller’s limit of 7±2, either with or without feedback. Six participants from the music school exceeded Miller’s limit in the blocks that included feedback. However, this performance benefit for music students disappeared when feedback was withheld, presumably because the music students were relying on their relative pitch skills. Without feedback, only one single participant exceeded Miller’s limit. This person was from the music school and performed with 54 per cent accuracy in the presence of feedback, and with 40 per cent accuracy in the absence of feedback. To compare this against Miller’s limit of 7±2 stimuli, this performance corresponds to the equivalent of about 19 perfectly-identified stimuli in the presence of feedback, and around 14 perfectly-identified stimuli without feedback.
Figure 4.2. A scatterplot showing the proportion of correct responses for each person, separately for blocks that included feedback (x-axis) and blocks without feedback (y-axis). Music students are represented by a red triangle and social science students are represented by black circles. The dotted grey lines represent Miller’s (1956) upper limit of 9 stimuli (equivalent to .25 accuracy). The dotted green lines represent chance performance of .028 (1/36).

The above analyses demonstrate that music students had a performance advantage relative to others, but that this advantage was mostly due to relative pitch, not absolute pitch. The next question that was addressed was whether the data were consistent with the hypothesis that there is a greatly elevated proportion of absolute pitch possessors amongst Mandarin speakers. For this, a mixture model was
estimated for the data. Each participant’s ability was modelled as measured by their equivalent number of correctly-identified stimuli (i.e., the percentage accuracy, multiplied by 36) either being drawn from a normal distribution with mean=7 and standard deviation=1 or from a uniform distribution between 9 and 36. These two distributions represent, respectively, people who do not “possess absolute pitch” and people who do. The particular choices of the distributions used to represent those two populations are not crucial to the results of the analysis, and were informed by previous research such as Miller’s (1956) famous 7+-2, and Rae (2010). The key quantity of interest, and the only estimated parameter of the model, was the rate at which these two components are mixed in the population; this is just the proportion of people with absolute pitch. The posterior distribution over this proportion was estimated using Markov chain Monte-Carlo sampling. To improve the stability of the sampling, the logit of the proportion parameter was estimated, and assumed a relatively wide prior based on previous research: a normal distribution in logit space, with mean -2 and standard deviation 2, which gives a peak of the prior distribution at a rate of 1 in 44, and spreads 95% of the prior distribution between rates of just over one in $10^9$ to one in 10 people. Ten thousand samples were drawn using MCMC with a starting sample near the mode of the prior, discarding the first 5,000 samples for burn-in.

Figure 4.3 shows the marginal posterior distribution over the mixture parameter, when using all participants’ data. The red line shows the prior distribution. These distributions show that the data are much more consistent with absolute pitch rates being equal in Western and Chinese populations, than being unequal as is often suggested. To quantify these results the Savage-Dickey method of calculating Bayes factors was used, comparing the evidence in favour of key prior
values (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). The data provide evidence against the rates of absolute pitch indicated by both of the vertical blue lines in Figure 4.3. These lines correspond to high rates of absolute pitch which might be expected under the hypothesis of an advantage for tonal language speakers. That is, .133 and .267 represent the expected rate of absolute pitch possessors based on the premise that Chinese speaking musicians would have four times the rate of absolute pitch than Western musicians (estimated at 1 in 15 [Revesz, 1953] to 1 in 30 [Baharloo et al., 1998] in English-speaking musicians, which are represented by the dotted black lines). A rate of four times higher was chosen as an average point based on suggestions that the incidence of absolute pitch in Chinese music conservatories or for Asian musicians is approximately three times (Deutsch et al., 2006), four times (Gregerson et al., 1999), or five times (Gregerson et al., 2007) the rate of Western musicians.

The highest rate (four times the 1-in-15 rate for musicians) has a Bayes factor of more than 800-to-1 against it, while the lower of the two rates (four times a 1-in-30 rate) still has a Bayes factor of about 3-to-1 against it. In contrast, the current data provide evidence for the rates of absolute pitch indicated by the dashed black lines in Figure 4.3; those lines correspond to rates of absolute pitch reported amongst English-speaking musicians. The two lines correspond to rates of 1-in-15 and 1-in-30, and the associated Bayes factors are about 3.5-to-1 and 7.5-to-1 in favour.
Figure 4.3. The posterior distribution over the rate of people with absolute pitch in the population. The vertical blue lines .133 and at .267 represent the expected rate of absolute pitch possessors based on the premise that Chinese speaking musicians would have four times the rate of absolute pitch than Western musicians (estimated at 1 in 15 to 1 in 30 in English-speaking musicians, which are represented by the dotted black lines).

Discussion

The work described in this chapter aimed to investigate whether the high rates reported in native Mandarin speaking musicians are observed when using the rigorous testing framework of absolute identification. While estimates of absolute pitch rates in Western populations range from 1 in 10,000 in the general population (Bachem, 1955; Profita & Bidder, 1988), rising to 1 in 30 (Revesz, 1953) to 1 in 15 (Baharloo et al., 1998) in professionally trained Western musicians, Chinese speaking musicians are often reported to have three to five times this rate (Deutsch et al., 2006; Gregerson et al., 1999; Gregerson et al., 2007).

To investigate this premise, two groups of native Mandarin speaking university students from The People’s Republic of China were tested: University students
enrolled in a music degree, and university students enrolled in a social science degree. The music students performed better on average than social science students, but this advantage was limited to blocks in which corrective feedback was provided. This suggests that the musically trained participants were using relative pitch strategies, rather than completely absolute pitch strategies, combining the knowledge of the label of the previous tone that was provided via feedback, and formal training in the identification of musical intervals.

Only one participant exceeded Miller’s (1956) upper limit of nine stimuli without feedback. This participant indicated they were exposed to musical education during the critical period, although they did not self-identify as an absolute pitch possessor. Identifying such a participant is reassuring as it confirms that the experimental procedure was able to measure identification performance in high-performing individuals. This individual did not perform even close to perfectly - they were wrong roughly about as often as they were right – but they were still classified as “possessing absolute pitch” because their performance was well above the general-population limit identified by Miller’s famous 7±2. This result of one in thirty of the musicians having absolute pitch ability corresponds to the rate Revesz (1953) suggested for English-speaking musicians, but does not fit with the hypothesis of a large advantage for native speakers of tonal languages.

The results of this study do not support the previous findings (see citations above) that the incidence of absolute pitch is higher amongst Chinese speaking musicians than Western musicians. There are four possible reasons for this apparent discrepancy in findings. The first relates to the use of absolute identification methodology. As is common practice in absolute identification tasks, this experiment used uni-dimensional sine tones (fundamental frequency only), which eliminates the
possibility of using extra cues such as familiarity with timbre, attack and decay characteristics, and harmonic overtones to identify the tones. Previous findings that tonal language speakers have superior absolute pitch ability have sometimes relied on participants recognising or producing vocal cues (Deutsch et al., 2006; Deutsch, et al., 2004). Perhaps the lack of extra cues, such as timbre, in this study deflated the number of participants that might have otherwise performed better in an absolute pitch task using tones or words. However, as shown in Rae (2010), and in the work in Chapter 3 of this thesis, these additional cues, or lack thereof, are not likely to have significantly altered the performance of the participants.

The second possibility is that speaking a tonal language does not rely on absolute pitch ability, but instead on relative pitch. As mentioned earlier in this chapter, tonal language speakers require the ability to detect the relative pitch height between lexical units, rather than the absolute pitch of the tones (Fromkin, Rodman, & Hyams, 2010; Hove, Sutherland, & Krumhansl, 2010; Shellenberg & Trehub, 2008; Ye & Connine, 1999). Whether or not tonal languages rely on an absolute pitch template is still debated.

The third possibility is that the sampling technique differed from what is often reported in other studies that test absolute pitch in native tonal-language speaking populations. Other studies have targeted recruitment at participants with self-reported absolute pitch, and this will have led to samples which are unrepresentative of the population. The current study conveniently sampled tertiary music students, with no mention of absolute pitch being a criterion for inclusion. Perhaps the difference of absolute pitch ability suggested in the literature reflects a more conservative self-estimate of absolute pitch ability in Asian musicians compared with Western musicians, rather than higher rates of absolute pitch.
The fourth possibility as to why the results of this study do not show a greater rate of absolute pitch for mandarin speakers than for Western populations is that the sample size was relatively small. However, given that Deutsch et al. (2006) and Gregerson et al. (1999, 2007) suggested that the rates of absolute pitch is three to five times higher in tonal language speaking participants, the sampling procedures should have at least show some effect size, even if not to the size that Deutsch, Gregerson, and colleagues suggest. The Bayesian analysis above directly tests this idea, and demonstrates that there is strong evidence against the rates of absolute pitch in tonal language speakers which have been suggested by previous authors.

In summary, the work in this chapter aimed to investigate whether Mandarin speaking musicians had increased rates of absolute pitch ability compared with reported Western rates, using an absolute identification framework. The data were more consistent with the hypothesis of identical rates of absolute pitch in English-speaking and Mandarin-speaking populations, than with different rates.
Chapter 5:

Do explicit instructions about the pairing of stimuli and labels increase novices’ learning ability in absolute identification practice tasks?

Chapter 1 of this thesis outlined the history of the paradigm of absolute identification. Miller’s (1956) seminal paper described how people were unable to identify more than 7±2 (an upper limit of nine) uni-dimensional stimuli in an absolute identification task. Following his finding, many other researchers reported the same, showing that the limit to identify uni-dimensional stimuli was resistant to many experimental manipulations (e.g. Garner, 1953; Pollack, 1952; Braida & Durlach, 1972; Lacouture et al., 1998; Weber, Green, & Luce, 1977; Hartman, 1954).

More recent research, however, has shown that in some exceptional cases Miller’s upper limit of nine stimuli is able to be surpassed. Rouder et al. (2004) reported on three participants who were able to learn line lengths beyond Miller’s upper limit; one considerably beyond it. Dodds et al. (2011) extended these findings not only for line lengths but also for dot separation, the degree of angular rotation for a line, and tone frequency. Rae (2010) also reported on six participants who exceeded Miller’s limit when identifying tones of varying frequency, with two of those six going on to increase their performance even further over a period of ten practice sessions. All of the participants who learned to identify uni-dimensional stimuli beyond Miller’s upper limit of nine had one thing in common: They all started above with above average accuracy.

While these exceptional cases have been documented, exceeding Miller’s (1956) limit in identifying uni-dimensional stimuli is still not the norm in AI tasks.
As research has shown for many years, the vast majority of participants can still not exceed Miller’s limit, even with practice and experimental manipulations (e.g. Garner, 1953; Pollack, 1952; Braida & Durlach, 1972; Lacouture et al., 1998; Weber et al., 1977; Hartman, 1954; Miller, 1956). There is evidence that the participants in these exceptional cases of breaking Miller’s limit have unique qualities. Firstly, people who are able to exceed Miller’s limit are much more likely to have a complex psychological representation of the physically uni-dimensional stimuli. Rae (2010) showed that the participants who were able to exceed Miller’s limit in a pitch identification task all had a helical representation of the tones, using multi-dimensional scaling techniques (Cox & Cox, 1993; 2001; Shepard, 1974), while the participants who were not able to exceed Miller’s limit did not. Further, Chapter 2 of this thesis outlined the use of a Structural Forms algorithm (Kemp & Tenenbaum, 2008; Dodds et al., 2012) that inferred the structure of psychological representations based on relational data. The analyses showed that the better performers in absolute identification tasks of lines of varying length, dot separations, and tones of varying frequency (Dodds et al., 2012; Rae, 2010) had more complex psychological representations of the stimuli.

However, while having a more complex psychological representation of the uni-dimensional stimuli is part of the story for well-performing participants, there is more to it. Rouder et al. (2004), Dodds et al. (2012), and Rae (2010) all showed that the participants who were able to learn to improve their identification of uni-dimensional stimuli beyond Miller’s (1956) limit were the participants who started with above average performance. Further, Rae (2010) revealed that the well performing participants in the pitch identification tasks were also musically proficient, having had years of musical training. Although this showed a relationship
between musical proficiency and pitch identification, it is still not clear whether the musical proficiency led to the superior ability to identify tones, or whether some innate or learned skill facilitated the participants’ path to musical proficiency.

It is possible that there is a mediating variable between musical proficiency, and being able to identify tones of varying frequency beyond Miller’s (1956) limit. As previously mentioned, while it was long thought absolute pitch was an innate ability that was unable to be learned (e.g. Revesz, 1953; Stumpf, 1883 in Takeuchi & Hulse, 1993), several researchers have noted a correlation between AP ability and musical training during the critical period (under approximately seven or eight years of age; e.g. Miyazaki, 1988). Further, some researchers such as Levitin (1994) and Lundin (1963) believe that learning is possible beyond this critical period. However, not all people who are musically trained, in the critical period or beyond, go on to develop absolute pitch. Levitin (1994) suggests that the essential aspect of learning to identify tones of varying pitch is the paired association with the note name.

If Levitin’s (1994) proposal that the key element of learning to identify musical tones is the explicit pairing of the notes with the note name, perhaps this might also extend further to all, or at least some, modalities in absolute identification tasks. It would stand to reason that the absolute identification of uni-dimensional stimuli, such as lines of varying lengths, is not a skill necessary in everyday life, so it is unlikely that most people have explicitly learned to pair stimuli with arbitrary labels. Perhaps the exceptional participants who are able to identify beyond Miller’s (1956) upper limit of nine stimuli are not special because they have a unique capacity to learn to identify and label uni-dimensional stimuli, but rather they have, as Levitin says is the key, either naturally or via some specific training simply made and rehearsed the connection between the stimuli and corresponding labels. Perhaps,
if given explicit training and long enough to practice, most people would be able to pair uni-dimensional stimuli with labels. In other words, breaking Miller’s limit would not be such a unique skill after all.

This proposition, that perhaps it is simply the explicit pairing and extended practice that is needed for participants to break Miller’s (1956) upper limit of nine stimuli in an absolute identification task, is the primary motivation for this chapter. If people with no previous specific training are given long enough to explicitly learn to pair the uni-dimensional stimuli and associated labels, perhaps they would be able to break Miller’s limit. The second motivation for the work in this chapter was that given that ability to exceed Miller’s limit in an absolute identification task is still the exception with very few cases reported (Rouder et al., 2004; Dodds et al., 2012; Rae, 2010) in a long history of people not being able to (e.g. Garner, 1953; Pollack, 1952; Braida & Durlach, 1972; Lacouture et al., 1998; Weber et al., 1977; Hartman, 1954), this experiment aimed to give highly motivated people the chance to replicate these relatively rare findings, especially in the light of the replication crisis (For more information on the replication crisis in psychology, see Maxwell, Lau, & Howard, 2015).

The third and final motivation for this chapter was to incorporate the use of no-feedback blocks into the absolute identification learning task. This element is a unique contribution compared with other long term absolute identification tasks that have previously been attempted (Rouder et al., 2004; Dodds et al., 2012). Typically, AI tasks have used feedback on a trial by trial basis as a way to motivate and orient the participant throughout the task (e.g. Rouder et al., 2004; Dodds et al., 2012). While feedback does more than likely help with motivation and anchoring, it also introduces the possibility of allowing the participants to use relative judgement
strategies in addition to, or instead of absolute judgement. Rae (2010) as well as the work in Chapters 3 and 4 have shown that there is a performance difference in identifying uni-dimensional stimuli depending on whether feedback is provided or not. In all cases feedback increased the accuracy of identifying the stimuli. While Rouder et al. (2004) and Dodds et al. (2012) showed that in some exceptional cases learning beyond Miller’s (1956) limit was possible, all experiments in both papers also provided feedback after each trial. Using blocks with no feedback in addition to blocks with feedback would allow for a truer measure of absolute judgement, where the chance for using relative judgement would be minimised.

For this experiment nine highly motivated participants, with no exceptional ability in identifying uni-dimensional stimuli, were chosen to attempt to learn tones of varying frequency and lines of varying length. Both of these modalities have shown the best previous successes in learning (Rouder et al., 2004; Dodds et al., 2012; Rae, 2010) and they were chosen to give the best possible chance for the participants to demonstrate learning in this experiment.

Method

Participants

Nine participants were recruited from an undergraduate psychology pool and from the general university population. Participants were reimbursed with a $25 EFTPOS voucher per hour of participation, except for one participant who refused payment. Participants were chosen on the basis that they had no expertise in the modality. That is, participants who took part in the tones of varying frequency experiment had no musical training, and participants who took part in the line lengths experiment were not overly familiar with line lengths through a profession or
training such as architecture, carpentry, graphic design, or the like. All participants reported having normal to corrected-normal vision, and all participants who completed the pitch experiment reported having good hearing. All participants reported having high motivation to learn to identify the stimuli. The nine participants produced eleven data sets, as two participants completed two sets of data as shown below in Table 5.1.

Table 5.1.

Participants A, B, C, and D completed the in-laboratory version of the lines of varying length experiment. Participants A, B, and E completed the in-laboratory version of the tones of varying frequency experiment. Participants F, G, H, and J completed the remote delivery version of the tones of varying frequency experiment. Note that participants A and B were the only two participants who completed two experiments across two modalities.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>In laboratory delivery</th>
<th>Remote delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lines of varying length</td>
<td>Tones of varying frequency</td>
</tr>
<tr>
<td>A</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stimuli

Two sets of 24 uni-dimensional stimuli were used for this experiment. The first set of stimuli consisted of computer generated pure tones of varying frequency spanning two Western octaves chromatically (by semitone, the smallest musical interval) from A3 (220Hz) to G#4 (415.40Hz). The second set of stimuli consisted of lines with length determined by a power function with an exponent of 1.05. The lengths, in pixels, of the lines are given in Table 5.2.

Table 5.2

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>44</td>
<td>46</td>
<td>49</td>
<td>51</td>
<td>54</td>
<td>56</td>
</tr>
<tr>
<td>59</td>
<td>62</td>
<td>65</td>
<td>68</td>
<td>72</td>
<td>75</td>
<td>79</td>
</tr>
<tr>
<td>83</td>
<td>87</td>
<td>92</td>
<td>96</td>
<td>101</td>
<td>106</td>
<td>111</td>
</tr>
<tr>
<td>117</td>
<td>123</td>
<td>129</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The laboratory versions of the experiment were conducted using Matlab R2010b, and the online version (tones of varying frequency only) was accessed via a link to a server: http://baal.newcastle.edu.au/LTAP/data (link no longer in operation). Tones were presented via Sony Stereo Headphones (MDR-XD100).

The second set of stimuli (lines of varying length) were presented in black on a white background, using a 21 inch CRT monitor set at a resolution of 1152 x 864 pixels. Images were positioned in the centre of the screen, with 22 x 22 pixel variation in position from trial to trial to discourage participants from using the edge of the monitor as a size cue.
Procedure

There were two methods of delivery for the experiment. Firstly, there was a laboratory based version of the task, using lines of varying length and tones of varying stimuli delivered via Matlab R2010b, using specific equipment (described in Stimuli section above) and with a set number of sessions (20) and blocks (7 per session). The second method was an online version, using tones of varying stimuli only, where participants could choose the total amount of sessions, and between three and seven blocks per session, and were free to remotely use any compatible device. All participants involved in the pitch identification task across both delivery methods used Sony Stereo Headphones (MDR-XD100). An experiment using lines of varying length was not conducted in the second method of delivery as not being able to control for the size of monitor or screen would potentially have given confounding results.

Participants were given the choice as to the delivery method and stimulus modality before the experiment began to maximise the chance of motivation and learning. Five participants completed the experiment via delivery method one. Two completed both lines of varying length and tones of varying frequency, two completed lines of varying length only, and one completed tones of varying pitch only. Four participants completed delivery method two where tones of varying frequency was the only option.

All participants were given explicit instructions about the purpose of the experiment: To learn to correctly identify the stimulus set by the explicit pairing of the stimuli with the labels. All participants displayed a high level of motivation to learn the respective stimulus set/s and demonstrated a clear understanding of the objective.
At the start of the experimental task, there was a brief study phase to familiarise participants with the stimuli and the method of responding, which is standard in AI tasks. Participants were visually presented with two rows of twelve buttons on screen labelled with a unique number (1 through 24) in order of ascending frequency (1 through 12 on the top row of buttons, and 13 through 24 on the bottom row). Musical note names (e.g. G#) were not used at all during this experiment, to allow for a direct comparison between the stimulus modalities, and also because none of the participants were musicians. Participants were then presented with each of the stimuli, via headphones for tones or on screen for lines, one at a time in ascending order. To move onto the following stimulus the participant had to select the correct button onscreen that corresponded to the stimulus label, after reading the following instructions,

“This is pitch x. When you hear this tone, press the x button.”

Or

“This is line x. When you see this line, press the x button.”

Following this, the test phase consisted of seven blocks of 120 trials each for the laboratory based method of delivery, and for between three and seven blocks of 120 trials (at the participant’s choosing) for the remote method of delivery. During each block, each of the 24 stimuli was repeated five times in random order. On each trial a fixation cross was presented for 500 milliseconds, before one of the 24 tones or lines was randomly presented for one second. The participant was asked to respond with the label that was attached to the stimulus in the study phase. If participants were incorrect, the correct answer was displayed for 500ms. If they were correct, “Correct” was displayed instead. Feedback was provided after each trial
during all blocks except for the last one in the session, which was always block seven for the in-laboratory delivery. The last block in the session for the remote delivery was between blocks three and seven, depending on how many blocks the participant had chosen that session. Each seven blocks (the equivalent of one laboratory session) took approximately one hour to complete. Participants A and B, who chose to complete both lab-based experiments, were given the tasks in the opposite order.

Results

All participants’ first sessions were analysed prior to the continuation of subsequent sessions to make sure that their starting points were not above Miller’s (1956) upper limit of nine stimuli, as this would have violated the goal of training novices. All participants’ accuracy for the initial session(s) was below Miller’s upper limit of nine stimuli (1956) for both feedback and no feedback blocks, so participants were asked to complete the remainder of the sessions.

Lines of varying length (in laboratory with set blocks and sessions)

Figure 5.1 plots the accuracy of each participant across their 20 sessions. Due to the large number of figures, all have been placed at the end of the Results section in this chapter for ease of reading. Of the four participants, three participants showed a significant increase in learning for blocks with and without feedback across the 20 sessions, while the fourth participant only showed an increase for blocks with feedback (regression results are in Appendix 3.1, all p’s < .001). The regression coefficient values in Appendix 3.1 (plus Appendix 3.2 and Appendix 3.3) seem lower than they might be, due to having been calculated per block using a proportion
correct calculation for accuracy, in keeping with the rest of the analyses in this chapter. For example, Participant B shows a regression coefficient of .003 (Appendix 3.1) for learning of lines of varying length (with feedback). While this might seem like a low rate of learning, especially given their increase of 9.0 stimuli across the 20 experimental sessions, .003, when converted to an accuracy using percentage (instead of proportion correct) across sessions (instead of blocks), this becomes a regression coefficient of 2.1 percentage points per session.

The same three participants who showed learning across blocks with and without feedback also learned to exceed Miller’s (1956) upper limit of nine stimuli, both with and without feedback. One participant (Participant B) showed an outstanding increase of 9 stimuli for blocks with feedback and 5.7 stimuli for blocks without feedback, learning to identify 16.3 with feedback and 13.5 without\(^1\).

Figure 5.4 shows the improvement for participants, by comparing average accuracy for the first two blocks with their average accuracy for their last two blocks (separately for both feedback and non-feedback blocks). Appendix 3.1 reports the results of one-tailed one-sample t-tests ($\mu = .375$; equivalent to 9 stimuli) for each participant’s blocks in the final two sessions for blocks with feedback. Binomial analyses were conducted on the accuracy for blocks without feedback in the last two sessions (using .375 as the probability the outcome will occur; equivalent to 9 stimuli). This was because the degrees of freedom for a t-test would have been equal to 1 for the same analysis for blocks without feedback (that is, there was only one block without feedback in each session). The relevant $p$-values are shown in Appendix 3.1 (plus Appendices 4.2 and 4.3). One participant (Participant D) of the

---

\(^1\) Calculated by comparing the average of the first two and last two sessions, separated for blocks with and without feedback. All averages are presented in Table 5.2.
four failed to learn to exceed beyond Miller’s upper limit of nine stimuli for both feedback and no-feedback blocks. The average accuracy, and equivalent stimuli, for the first two and last two sessions are presented in Appendix 3.1.

**Tones of varying frequency (in laboratory with set blocks and sessions)**

Figure 5.2 shows the accuracy of each participant across their 20 sessions. Two participants (Participants A and B, who also completed the lines of varying length experiment) showed learning throughout the experiment, but only for blocks with feedback (regressions results are in Appendix 3.2, p’s < .001). These same two participants learned to identify a little beyond Miller’s (1956) upper limit of nine stimuli, again only for blocks with feedback. However, only Participant A’s final identification of 9.9 stimuli was statistically significant. Figure 5.4 shows the improvement for participants, by comparing average accuracy for the first two blocks with their average accuracy for their last two blocks (separately for both feedback and non-feedback blocks). Appendix 3.2 reports the results of t-tests and binomial tests as described for Appendix 3.1. One participant (Participant E) failed to show overall learning for blocks with or without feedback and did not learn to exceed nine stimuli. The average accuracy, and equivalent stimuli, for the first two and last two sessions are presented in Appendix 3.2.

**Tones of varying frequency (Remote delivery: Participant selected number of sessions and blocks per session)**

Figure 5.3 shows the accuracy of each participant across their sessions. While all participants showed at least a slight increase in identification across the sessions, no participant learned to identify beyond Miller’s (1956) upper limit of nine stimuli.
Participant G showed significant learning for blocks both with and without feedback (regressions results are in Appendix 3.3, p’s < .001). Participants H and J also showed an increase (p < .001 and p = .002 respectively). Although, no participant here was able to learn to identify beyond nine stimuli, the rate of learning shown in Figure 5.3 suggests that more sessions could have eventually allowed them to learn to an accuracy beyond Miller’s limit. Figure 5.4 shows the improvement for participants, by comparing average accuracy for the first two blocks with their average accuracy for their last two blocks (separately for both feedback and non-feedback blocks). Appendix 3.3 reports the results of t-tests and binomial tests as described for Appendix 3.2 The average accuracy, and equivalent stimuli, for the first two and last two sessions are presented in Appendix 3.3.
Figure 5.1. Proportion correct per block for each participant in the lines of varying length (laboratory based) learning experiment. Each circle on the graph represents the proportion correct per block. Open circles represent blocks with feedback given, and the black circles represent blocks with no feedback given which were always the last block of the session. The red line represents the slope for blocks with feedback, and the blue line represents the blocks with no feedback. The parallel dotted black lines represents Miller’s (1956) lower and upper limits of 5 and 9 stimuli (equivalent to .208 and .375).
Figure 5.2. Proportion correct per block for each participant in the tones of varying frequency (laboratory based) learning experiment. Each circle on the graph represents the proportion correct per block. Open circles represent blocks with feedback given, and the black circles represent blocks with no feedback given which were always the last block of the session. The red line represents the slope for blocks with feedback, and the blue line represents the blocks with no feedback. The parallel dotted black lines represents Miller’s (1956) lower and upper limits of 5 and 9 stimuli (equivalent to .208 and .375).
Figure 5.3. Proportion correct per block for each participant in the tones of varying pitch (remote delivery) learning experiment. Each circle on the graph represents the proportion correct per block. Open circles represent blocks with feedback given, and the black circles represent blocks with no feedback given which were always the last block of the session. The red line represents the slope for blocks with feedback, and the blue line represents the blocks with no feedback. The parallel dotted black lines represents Miller’s (1956) lower and upper limits of 5 and 9 stimuli (equivalent to .208 and .375).
Improvement across all three experiments reported in Chapter 5 (line lengths in-laboratory, tone frequency in-laboratory, and tone frequency remote-delivery) separately for each participant and feedback and no-feedback condition. Improvement is calculated by subtracting the average performance (in stimuli identification) across the first two sessions from the average performance across the last two sessions. Each participant is represented by two columns: The first column (black) is for improvement with feedback and the second column (grey) is for improvement without feedback. All participants showed improvement across both feedback and non-feedback conditions except for Participant D. Note that Participant E’s improvement for the no-feedback condition was 0.

Discussion

The main aim of this experiment was to investigate whether people with no previous specific training, given enough time to practice, were able to learn lines of varying length and/or tones of varying frequency to an accuracy beyond Miller’s upper limit of nine stimuli (1956) in an absolute identification task. In order to make
the intention clear, participants were told explicitly that the way to succeed in this task was pairing of each stimulus with its label, as Levitin (1994) suggested was the key to learning absolute pitch.

Of the 11 learning attempts, all initial accuracy for feedback and no-feedback blocks was below Miller’s (1956) limit. This allowed for an assumption that if learning was to take place then it was at least partly a direct result from the practice in the experiment. Of the two modalities, lines of varying length showed the best learning with three out of the four participants learning beyond Miller’s limit for both feedback and no-feedback blocks. One outstanding participant learned to identify an equivalent of 16.3 stimuli with feedback and 13.5 without feedback, both well above Miller’s upper limit of nine stimuli. However, this was still not to the level of accuracy that one participant showed in Rouder et al. (2004), although these researchers only used blocks with trial by trial feedback, and the size of the largest stimuli set was larger than the one used in this experiment (30 versus 24 stimuli).

The learning attempts using tones of varying pitch also showed some increase in accuracy, but not to the same extent as lines of varying length. Of the seven participants, across both laboratory based and remote delivery, no one learned to identify above ten stimuli, with only two participants learning to identify slightly beyond Miller’s upper limit only for blocks with feedback.

It is possible that the discrepancy in performance between the two modalities in this experiment might be attributable to a difference in perception or potential processing capacity. The greater levels of learning shown for lines of varying length compared with tones of varying frequency could be based in a fundamental distinction between processing of visual and auditory stimuli. This is definitely a possibility, and the idea of whether stimulus modalities are truly comparable in
absolute identification tasks is discussed further in *Future Research Directions* in Chapter 6.

It might also be considered that the visual modality in this experiment (lines of varying length) provided the participants with additional spatial cues such as distance from the edge of the computer screen. However, the lines presented on screen were varied in position from trial to trial to specifically discourage participants from using the edge of the monitor as a size cue. Further, Dodds et al. (2011) examined whether external cues were responsible for the learning found for lines of varying length, and concluded they were not. Given the literature supporting the ability of some people to identify well beyond 50 tones of varying frequency (e.g. Miller, 1956), it is highly probable that the small sample size used in this experiment happened by chance to be better at learning to identify line lengths compared with tone frequency. Perhaps with a larger sample size a more equal level of learning would be demonstrated for both modalities.

It is possible that had the number of sessions been increased in the current study, there would have been more participants who were able to exceed Miller’s (1956) upper limit of nine stimuli, both for lines of varying length and tones of varying frequency. Nine of the eleven participants across the three tasks were still showing an increase in accuracy up to their last block. Had circumstances allowed, it would have been of interest to see where the participants reached their ceiling.

A second motivation for this experiment was to attempt to replicate the rare findings of exceptional cases breaking Miller’s (1956) limit in an absolute identification task (Rouder et al., 2004; Dodds et al., 2017; Rae, 2010). The results have replicated the finding that learning is possible, especially for lines of varying length, with one person in particular showing very good rates of accuracy increase.
However, several of the participants’ performances were also reminiscent of older literature that suggests that learning is not possible beyond nine stimuli in absolute identification tasks (e.g. Garner, 1953; Pollack, 1952; Braida & Durlach, 1972; Lacouture et al., 1998; Weber, Green, & Luce, 1977; Hartman, 1954). A question of interest this distinction poses is one that has been posed in this thesis: What is it that differentiates those who can and cannot learn to identify stimuli in absolute identification tasks? Chapter 6 discusses this further.

The final motivation for this experiment was to incorporate the use of no-feedback blocks into the absolute identification learning task. This was to investigate whether the same rates of learning took place as when the participants were given feedback on a trial by trial basis as they were in Rouder et al. (2004) and Dodds et al. 2012). Rae (2010), along with Chapters 3 and 4 of this thesis, showed that feedback allowed for higher accuracy in identifying the tones of varying frequency. The work in the current chapter aimed to investigate whether this was also the case for lines of varying length. The results in this experiment also showed this to be the case with all 11 learning attempts across both modalities showing a clear advantage for accuracy in blocks where feedback was given. This suggests that while trial by trial feedback may motivate and anchor the participant throughout the task, it more than likely also allows for the use of relative judgement instead of, or in addition to, absolute judgement. These results suggest that the truest measure of absolute identification is when no feedback is given. This calls into question the results of previous absolute identification tasks that have given feedback on all trials.

Perhaps in the future, absolute identification should be measured where feedback is withheld. This would likely give a truer measure of absolute identification, where the possibility of using relative judgement is removed, or at the
very least, limited. Blocks with feedback should still be included for learning, motivation, and anchoring, although the emphasis should not be placed on these results. However, as shown in this experiment, and the experiments described in Chapters 3 and 4 of this thesis, there is a relationship between participants’ identification of stimuli with and without feedback. The results of this chapter showed that most people who demonstrated an increase in learning to identify stimuli with feedback also learned to identify without feedback. It is therefore likely that the rate of learning is similar, but perhaps the high limits reported in previous absolute identification literature would simply be more conservative. Chapter 6 expands further on this issue.

As discussed in the Introduction of this chapter, the performers who learned to identify beyond Miller’s (1956) upper limit of nine stimuli in previous experiments had an above average initial performance in common. This was also the case for the participants in Rouder et al. (2004), Dodds et al. (2012), and Rae (2010). The results of this chapter were partly reminiscent of this for the participants who completed the in-laboratory version of the experiment. The participant who showed the most learning for line lengths (16.3 stimuli with feedback and 13.5 stimuli without feedback) started above average in their identification of the stimuli (7.2 for stimuli with feedback and 7.8 stimuli for stimuli without feedback). Further, the two participants who learned to identify beyond Miller’s limit in the tones of varying pitch experiment for feedback blocks also started their initial sessions above average (8.7 and 7.0 stimuli with feedback). However, two of the participants in the experiment described in this chapter who showed learning beyond Miller’s upper limit for tones of varying length, for both feedback and no-feedback blocks, did not start with above average accuracy (5.3 and 6.4 for blocks with feedback, and 4.6 and
This suggests that having a good initial accuracy does not always predict, nor guarantee, high levels of learning in an absolute identification task. This is discussed further in Chapter 6.

There were two methods of delivery for the current experiment. The first method was an in-laboratory version, set with 20 sessions, and seven blocks per session. In order to try and be as flexible as possible to elicit the most amount of learning, a remote access version was also attempted for the tones of varying frequency. While there were no outstanding performers in the tones of varying frequency learning attempts, the performance of the participants who completed the experiment remotely showed less learning overall than the laboratory based participants. This could have been because participants chose not to complete as many blocks as the in-laboratory participants, thereby not giving themselves the chance to reach beyond Miller’s (1956) limit. In addition, perhaps the remote access allowed for distractions so that participants were not able to completely focus on the task. It is also possible that tones of varying frequency are inherently more difficult to learn than lines of varying length, and had the remote experiment used lines instead of tones learning beyond Miller’s limit may have been demonstrated. The concept of modality dependent learning, while briefly discussed earlier in this Discussion section, is further discussed in Chapter 6.
Chapter 6:
General Discussion

Absolute identification is a paradigm where participants identify stimuli that vary on only one physical dimension, such as tone frequency (pitch; e.g. Hartman, 1954; Pollack, 1952; Dodds et al., 2011), line length (e.g. Lacouture, 1997; Rouder et al., 2004), or tone intensity (loudness; e.g. Garner, 1953; Dodds et al., 2011). In a typical absolute identification task, stimuli are presented to the participant one at a time, each with a unique label. Participants are then presented with randomly selected stimuli from the set, and are asked to recall the previously associated label. While people have a seemingly infinite capacity for learning complex items such as names, and faces (Dodds et al., 2011), absolute identification has long provided an exception to this.

Absolute identification has been associated with certain benchmark phenomena such as the bow effect, where when response accuracy is plotted as a function of stimulus position, the plot reveals a bow shape regardless of the stimulus set size (Kent & Lamberts, 2005; Pollack, 1953; Weber et al., 1977; Lacouture & Marley, 1995; Lacouture, 1997). Sequential effects, that is, the influence of recent stimuli and responses on the current decision, have also been consistently shown (Gilden et al., 1995; Holland & Lockhead, 1968; Van Orden et al., 2003; Wagenmakers, Farrell & Ratcliff, 2004, 2005; Ward & Lockhead, 1970, 1971). Sequential effects come in two varieties, assimilation and contrast (for further information see Donkin et al., 2015).

Miller’s (1956) now famous processing capacity limit, 7±2, has also long been said to be the limit of the number of items people could learn to perfectly identify in a unidimensional stimulus set. The capacity was shown to be resistant to many
experimental manipulations including stimulus set size (e.g. Garner, 1953; Pollack, 1952), practice (e.g. Garner, 1953; Weber et al., 1977; Hartman, 1954), and across a wide range of modalities (such as electric shocks, saltiness, line length, hue, and loudness; e.g., Lacouture et al., 1998; Pollack, 1952; Garner, 1953; Miller, 1956).

A related paradigm to absolute identification is absolute pitch. Absolute pitch is very similar to absolute identification in that both require the participant to label stimuli that vary on a physical dimension with previously assigned labels; absolute pitch just restricts itself to tones of varying frequency (pitch), most commonly the frequencies that correspond with the Western musical scale (Takeuchi & Hulse, 1993; Yost, 2009). As is the case with absolute identification, absolute pitch is reported to be a very rare ability (Takeuchi & Hulse, 1993), although some exceptional people are able to easily identify over 50 tones (Miller, 1956), which is well above Miller’s upper limit of nine stimuli. Although the ability is rare, the rate of absolute pitch is reported to be as high as 1 in 7 for highly accomplished musicians (Baharloo et al., 1998).

A contribution of this thesis was to integrate absolute identification and absolute pitch, which are clearly similar but have been treated separately. Further, the exceptional ability of some people to far exceed Miller’s upper limit in identifying tones of varying frequency in absolute pitch tasks was considered possibly to be able to shed some light on whether that sort of capacity is possible across other modalities in an absolute identification task.

While absolute identification has not demonstrated the ability to identify uni-dimensional stimuli to the extent that absolute pitch has, recent research has shown that learning to identify uni-dimensional stimuli is possible. Despite the longstanding assumption that people cannot learn beyond Miller’s (1956) upper limit of nine
stimuli, recent work has shown otherwise. All three participants in Rouder et al.’s (2004) paper learned to identify beyond nine line lengths, with one of the participants learning to identify 20 stimuli. Dodds et al. (2011) extended these findings not only for line length, but also for dot separation, degree of angular rotation for a line, and tone frequency. Dodds et al. (2011) reported on one participant who was able to identify approximately 16 tones of varying pitch after 10 experimental sessions. Rae (2010) further extended these findings for tones of varying frequency and showed that two participants learned to identify an average of 29 stimuli after 10 experimental sessions.

These relatively new findings provide evidence that it is possible to exceed Miller’s (1956) upper limit across several modalities, along with the exceptional ability that is reported in absolute pitch tasks. This thesis therefore aimed to investigate three main questions: To what extent is the ability to exceed Miller’s (1956) limit learned? What quality do people have who can learn beyond nine stimuli? How far beyond Miller’s limit is it possible to go? Each of these questions is discussed below.

Answering the aims of this thesis

1. To what extent is the ability to exceed Miller’s (1956) limit learned?

The exceptional ability that some people are reported to have in absolute pitch tasks was an inspiration in investigating the question: To what extent is the ability to exceed Miller’s (1956) limit learned? For a long while absolute pitch was thought to be an innate quality that was unable to be learned. People were said to either possess or not possess absolute pitch (e.g. Revesz, 1953; Stumpf, 1883 in Takeuchi & Hulse, 1993). However, in a shift from the nature-only argument, most recent researchers
agree that there is a critical period, under the age of approximately seven or eight, when children can easily learn absolute pitch if exposed to enough musical training training (e.g. Levitin & Rogers, 2005; Takeuchi & Hulse, 1993). Some researchers also believe that learning is possible beyond the critical period (e.g. Levitin, 1994; Lundin, 1963).

Evidence that absolute pitch is, at least in part, a learned ability can be seen in studies where absolute pitch possessors are not uniform across their identification of tones. These people more accurately identifying notes they have had more exposure to such as Middle C, or other white notes on the piano (Athos et al., 2007; Boggs, 1907; Lundin, 1963; Miyazaki, 1988; Takeuchi & Hulse, 1993), and the note they tune their instruments to (Brammer, 1951; Ward, 1999). Furthermore, musicians generally are more accurate in identifying pitch on their own instrument (Pantev et al., 2000; Takeuchi & Hulse, 1993). Levitin (1994) proposed that the key element of learning to identify to label musical tones is the explicit pairing of the notes with the note name.

One of the issues in disentangling whether absolute pitch is innate or learned, or some combination of both, is that the people reported to have absolute pitch are typically musicians. With the average person having only a 1 in 10,000 chance of possessing absolute pitch (Takeuchi & Hulse, 1993), highly accomplished musicians are reported to likely have a 1 in 7 chance (Baharloo et al., 1998). This begs the question of a classic “chicken or egg” conundrum. It may be possible that people have learned to identify a large number of tones of varying frequency due to their prolonged musical training, or it may instead be the case that these same people possess a unique pitch identification ability which has facilitated their musical training and proficiency.
The aim described in Chapter 3 of this thesis was to investigate whether learning plays a role in the identification of tones of varying pitch by comparing the identification of Western and non-Western tones by both musicians and non-musicians. The justification for this was that musicians are repeatedly exposed to the discrete tone frequencies that correspond with the Western musical scale through many years of practice, while non-musicians are to a lesser extent through popular culture. If the ability to identify tones of varying frequency is, at least in part, a learned skill then it would stand to reason that Western tones would be more accurately identified than non-Western tones, particularly for musicians.

To investigate whether Western tones were better identified than non-Western tones, musicians and non-musicians were given a typical absolute identification task using tones of varying frequency. The Western tones matched notes used in the current Western scale, and the non-Western tones were based on prime number fundamental frequencies. As expected, musicians outperformed non-musicians in identifying tones, particularly for the Western tones. Musicians and non-musicians performed at very similar levels of accuracy in identifying the non-Western tones. This suggests that musicians’ superior ability to identify tones of varying frequency might be, at least in part, due to their long exposure to the Western scale during their years of training. It is unlikely that musicians, even ones who demonstrate an exceptional pitch identification ability, are only born with an innate ability to identify the arbitrary frequencies of the current Western musical tones to the exclusion of all others.

The aim of the work in Chapter 4 of this thesis was to further investigate whether learning plays a role in the identification of tones of varying pitch by examining whether native Mandarin speakers had an absolute pitch advantage.
compared with Western populations. Native tonal language speakers, such as speakers of Mandarin, have been reported to have superior levels of pitch identification ability compared with non-tonal language speakers (Deutsch, 2002, 2006; Deutsch et al., 2004; Deutsch et al., 2009; Levitin & Rogers, 2005; Pfordresher & Brown, 2009), as different tones distinguish different lexical meaning for the same word (Maddieson, 2013).

Two reasons that were considered for native tonal language speakers’ reported pitch identification advantage is that their early exposure to their language may provide two elements shown to correlate with superior pitch identification. These are early life training of pitch identification during the “critical period”, and then extended practice at pitch identification. People with high accuracy in absolute pitch tasks are often those who began their training in the critical period (Bachem, 1955; Bermudez & Zatorre, 2009; Deutsch et al., 2004; Levitin & Rogers, 2005; Miyazaki, 1988; Profita & Bidder, 1988; Russo et al., 2003; Sergeant, 1969; Takeuchi & Hulse, 1993). Higher rates of absolute pitch are also shown amongst individuals who have had extended musical training (Bachem, 1955; Baharloo et al., 1998; Profita & Bidder, 1988; Revesz, 1953).

To investigate whether native Mandarin speaking participants had an absolute pitch identification advantage, an experiment was conducted in China, where tertiary enrolled musicians and non-musicians completed an absolute identification task using tones of varying frequency. The results showed that music students performed with higher accuracy than social science students, although there was only evidence for this when feedback was provided. (For a discussion on whether feedback allows for a true account of absolute identification, see the section Additional issues raised from the results of this thesis below.) Only one participant, a musician, exceeded
Miller’s (1956) upper limit of nine stimuli for tones that were provided without feedback, to an equivalence of approximately 14 stimuli. A mixture model and Bayes factor analyses showed that the data are much more consistent with absolute pitch rates being equal in Western and Chinese populations, unlike the unequal rates that have been suggested in some previous research.

These results, that native tonal language speakers do not fit with the hypothesis for a large advantage for absolute pitch identification, may further support a key point about whether absolute pitch identification is a learned ability. Assuming that the lower than expected absolute pitch rates were either a result of strict experimental standards that absolute identification tasks demand, or a less selective participant sampling technique employed than in previous literature, being a native tonal language speaker does not allow for an advantage in absolute pitch identification. If speaking a tonal language only relies on relative pitch ability (Fromkin et al., 2010; Hove et al., 2010; Shellenberg & Trehub, 2008; Ye & Connine, 1999), it makes sense that this does not give speakers an absolute pitch identification advantage. However, if speaking a tonal language relies, even in part, on absolute pitch strategies (Deutsch, 2002, 2006; Deutsch et al., 2004; Deutsch et al., 2006; Deutsch et al., 2009; Gregerson et al., 1999; Gregerson et al., 2007; Lee & Lee, 2009; Lee & Hung, 2008; Levitin & Rogers, 2005; Pfordresher & Brown, 2009) then the question remains: Why does this ability not extend to the identification of tones of varying frequency? It is very possible that, as Levitin (1994) suggests, the essential aspect of learning to identify tones of varying frequency is the explicit pairing of the tone with the label, for which an absolute pitch template for tonal language speakers would not help.
In Chapter 5 of this thesis the investigation aimed to further question whether learning plays a role in absolute identification by examining whether explicit instructions about the pairing of stimuli with their labels, along with many practice sessions, increases novices’ ability to identify uni-dimensional stimuli. While exceptional cases of exceeding Miller’s (1956) upper limit of nine stimuli have been reported (e.g. Rouder et al., 2004; Dodds et al., 2011; Rae, 2010), the vast majority of participants are still unable to exceed Miller’s limit even with practice and experimental manipulations (e.g. Dodds et al., 2011; Rae, 2010; Garner, 1953; Pollack, 1952; Braida & Durlach, 1972; Lacouture et al., 1998; Weber et al., 1977; Hartman, 1954; Miller, 1956). In the long-term learning experiment reported in Chapter 5 of this thesis, motivated novice participants committed to between 20 and 29 sessions (81 and 140 blocks) with explicit instructions that the way to learn to increase accuracy was the direct pairing of the stimuli with their associated labels, as Levitin (1994) suggested.

Evidence suggests that musical proficiency is correlated with a superior ability to identify tones of varying frequency. The work described in Chapter 3 of this thesis and Rae (2010) showed that musicians were more likely to have superior absolute pitch identification ability. However, once again, this comes back to the chicken or the egg problem: Was it years of musical training that gave rise to superior pitch identification, or was it the ability to identify pitch that facilitated the years of training? Using novices for the experiment reported in Chapter 5 allowed for an investigation into whether an ability to acquire superior absolute pitch identification was possible in the absence of years of training. Furthermore the experiment described in Chapter 5 extended the learning experiment to lines of varying length, as both Rouder et al. (2004) and Dodds et al., (2011) reported success of exceptional
participants learning to exceed Miller’s (1956) upper limit of nine stimuli for this modality.

The results of the experiments described in Chapter 5 showed that learning to identify uni-dimensional stimuli beyond Miller’s (1956) upper limit of nine stimuli is possible for participants who have no modality-specific related training. Three of four participants learned to exceed Miller’s limit for lines of varying length with and without feedback, with one participant learning to identify an outstanding 16.3 stimuli (with feedback). However, although six of the seven participants who attempted to learn tones of varying frequency showed at least a slight improvement, none of them were able to exceed Miller’s upper limit for blocks without feedback by the end of the experiment, although two were able to when feedback was provided. Nonetheless, given that most participants were still showing an increase in accuracy up to their last session, it is possible that if they had been allowed more time there would have been more participants showing an ability to break Miller’s upper limit.

The results presented in Chapters 3, 4, and 5 of this thesis, along with recent research (Rouder et al., 2004; Dodds et al., 2001; Rae, 2010) show that the ability to identify uni-dimensional stimuli is, at least in part, a learned skill. This is in contrast to the previously long-standing assumption that Miller’s (1956) upper limit of nine stimuli was unable to be broken by practice. However, while there is strong evidence that learning is possible in absolute identification tasks, this conclusion is still not clear cut. As the findings in this thesis show, along with previous research (cited above), whether learning is possible depends on individual differences. Further, there are clear differences across modalities in how much learning is facilitated. The
difference in the ability to learn across modalities is discussed below in the section

_How far beyond Miller’s (1956) limit is it possible to go?_

Learning has shown to be possible in absolute identification tasks, although there are clearly differences in participants’ ability to do so. Rouder et al. (2004) showed that all three participants in their experiments using lines of varying length demonstrated some learning beyond Miller’s (1956) upper limit of nine stimuli, with one outstanding participant learning to identify 20 stimuli. Dodds et al. (2011) showed that learning beyond Miller’s (1956) upper limit of nine stimuli is possible for some, but not for others. Rae (2010) also reported on a few exceptional participants who were well able to exceed Miller’s limit, while showing that most participants were confined by his upper limit of nine stimuli. Chapter 5 of this thesis also reported on some participants being able to learn beyond Miller’s upper limit, while some could not. While these examples show that learning in absolute identification tasks is possible, it also raises the question of what enables these exceptional participants to learn beyond the upper limit (Miller, 1956) that was previously thought to be unbreakable. This leads on to the second point of interest that this thesis aimed to investigate.

2. _What quality do people have who can learn beyond Miller’s (1956) upper limit of nine stimuli that separates them from people who cannot learn?_

As discussed above, it is clear that some people have the ability to learn to identify uni-dimensional stimuli beyond Miller’s (1956) upper limit of nine in an absolute identification task. This is an important finding, given this limitation has been a key element of many theoretical accounts of absolute identification (including: Stewart et al., 2005; Petrov & Anderson, 2004; Marley & Cook, 1984;
and Brown et al., 2008). One of the questions this thesis aimed to investigate was what quality people have that enable them to break Miller’s limit, that people who are bound by the limit do not.

The work described in Chapters 3, 4, and 5 of this thesis, along with recent research (Rouder et al., 2004; Dodds et al., 2001; Rae, 2010) show a considerable variability in individual differences in the ability to learn to identify uni-dimensional stimuli. Dodds et al. (2011) noted that the participants who were able to learn beyond Miller’s (1956) upper limit of nine stimuli also had good initial performance. This was also the case for the two exceptional participants in Rae (2010) who learned to identify tones of varying frequency well beyond Miller’s upper limit. The findings of the work in Chapter 5 of this thesis, where participants completed between 84 and 140 blocks in an attempt to learn either tones of varying frequency or lines of varying length, were partly reminiscent of this.

However, the relationship between initial performance and amount of learning beyond Miller’s (1956) limit shown in Chapter 5 is not as clear cut as Dodds et al. (2011) might have suggested. The participant who showed the most learning for line lengths (both for blocks with and without feedback) started above average in their identification of the stimuli. Further, the two participants who learned to identify beyond Miller’s limit in the tones of varying pitch experiment for blocks with feedback also started their initial sessions above average (8.7 and 7.0 stimuli with feedback). However, two of the participants who learned to identify lines of varying length beyond Miller’s limit (for blocks with and without feedback) started with relatively low initial accuracy. This suggests that having a good initial accuracy does not guarantee high levels of learning in absolute identification tasks. Similarly, a low
initial accuracy does not necessarily determine that a participant will not demonstrate learning beyond Miller’s upper limit of nine stimuli.

There may still be some merit to the relationship between good initial performance and the ability to learn to identify uni-dimensional stimuli beyond Miller’s (1956) upper limit of nine stimuli, although only when averaged across participants. For example, when the participants from the experiments in Chapter 5 of this thesis are separated into those who did learn beyond Miller’s upper limit and those who did not, on average participants who demonstrated learning for blocks both with and without feedback had a higher initial accuracy than those who did not demonstrate learning. However, given that in the absolute identification paradigm the people who can learn to exceed Miller’s limit are seen as exceptional cases, averaging performance across all participants does not demonstrate the complexity of the range of abilities demonstrated in learning to identify uni-dimensional stimuli.

While not as clear cut as previous research has suggested (Dodds et al., 2011), there is still a loose relationship between the initial starting accuracy and ability to learn to identify uni-dimensional stimuli. The next logical question that is raised from these findings is what actually contributes to a participant’s above average starting accuracy, and further, the ability to learn beyond Miller’s (1956) upper limit of nine stimuli. The case of musical proficiency and absolute pitch ability might shed some light on this.

It has been suggested that absolute pitch, the ability to identify uni-dimensional tones of varying frequency, is very rare in the general population (1 in 10,000; Takeuchi & Hulse, 1993) but is greatly elevated amongst accomplished musicians (Baharloo et al., 1998). Rae (2010) showed that the highest performing participants, as in, those who were able to exceed Miller’s (1956) upper limit in identifying nine
tones of varying frequency, were musically proficient. Further, they all had a good initial starting accuracy. However, as discussed throughout this thesis, the relationship between musical proficiency and absolute pitch is not clear cut: Does advanced musical proficiency lead to a higher likelihood of developing absolute pitch? Or does absolute pitch ability mean that becoming musically proficient is more likely? What is also apparent is that musical proficiency does not guarantee having absolute pitch. This might also be extrapolated to other modalities in absolute identification. Perhaps there are some participants, who despite undertaking huge amounts of practice, might never be able to break Miller’s limit. If this is the case, then clearly there is another factor at play in determining whether a person is able to learn to identify beyond Miller’s upper limit of nine uni-dimensional stimuli.

The second aim of this thesis was to investigate what qualities people have who are able to learn to identify uni-dimensional stimuli beyond Miller’s (1956) upper limit of nine in an absolute identification paradigm. Having an above average starting accuracy, discussed above, might tell part of the story, however there is clearly more to this learning. Previous research has suggested that although the stimuli used in absolute identification tasks is physically uni-dimensional, perhaps some participants have a more complex psychological representation of the stimuli, giving an advantage in absolute identification. While the stimuli used in absolute identification tasks always vary on only one physical dimension, this does not guarantee the corresponding psychological representations are also uni-dimensional. Hue, for example, is perceived as a circle (Shepard, 1962; MacLeod, 2003) even though the corresponding physical stimuli vary on only one physical dimension (wavelength). Research has shown that the number of stimuli that can be reliably
identified increases exponentially as the number of independent dimensions increase (Eriksen & Hake, 1955; Miller, 1956; Rouder, 2001; Nosofsky & Palmeri, 1996).

To investigate the psychological representation of the physically uni-dimensional stimuli used in absolute identification tasks, multi-dimensional scaling techniques (Cox & Cox, 1994; 2001; Shepard, 1974) have been previously employed. Dodds et al. (2010) reported that multi-dimensional scaling techniques did not reveal a complex psychological representation for participants who were able to demonstrate learning of lines of varying length. Rae (2010), however, revealed that participants who were able to identify beyond Miller’s (1956) upper limit of nine tones of varying frequency had a helical psychological representation of the tones (as Bachem [1950] also suggested). Together, these findings suggested that perhaps having a multi-dimensional psychological representation of physically uni-dimensional stimuli is modality-dependent.

The work presented in Chapter 2 of this thesis aimed to address this possibility, using a Structural Forms Algorithm (Kemp & Tenenbaum, 2008; Dodds et al., 2012) to determine the psychological representation of stimuli in absolute identification tasks. The Structural Forms algorithm improved on previous attempts of multi-dimensional scaling techniques to determine the psychological representation of stimuli used in absolute identification tasks. Multi-dimensional scaling lacks a framework for inference and so conclusions about psychological representation of stimuli are based on subjective judgments (see Lee, 2001). Kemp and Tenenbaum’s (2008) algorithm, however, includes a framework of inference that allows probabilistic comparison of different structural forms based on a penalized likelihood.
The data from several multi-session absolute identification learning experiments (Dodds et al., 2001; Rae, 2010; Rouder et al., 2004) were used to construct confusion matrices for individual participants for their entire 10 hours of practice, and then also split into the first and last five hours of practice; thus giving each participants three confusion matrices. For each confusion matrix, the best chain structure (assumed for one-dimensional continua) and best ring structure (a slightly more complex form) was identified using Kemp and Tenenbaum’s (2008) Structural Forms Algorithm.

To recap the results of the learning experiments used for the analysis in Chapter 2, learning to identify stimuli beyond Miller’s (1956) upper limit of nine stimuli was demonstrated for lines of varying length, tones of varying frequency, and dots of varying separation, but was not shown for tones of varying intensity (Dodds et al., 2011; Rae, 2010; Rouder et al., 2004). The results of the analysis using Kemp and Tenenbaum’s (2008) Structural Forms Algorithm based on the confusion matrices constructed from this data showed that a more complex (ring) structure correlated with higher initial accuracy. Additionally, this correlated with the ability to learn to identify the uni-dimensional stimuli across the modalities of lines of varying length, tones of varying frequency, and dots varying in separation. Further, those who did demonstrate learning mostly showed a move from a uni-dimensional psychological representation (chain structure) to a more complex representation (ring structure) as learning improved. In contrast, not one participant demonstrated a ring structure for tones of varying intensity.

These results described in Chapter 2 of this thesis are consistent with Miller’s (1956) original proposition that absolute identification is subject to a severe learning capacity when the stimuli really are uni-dimensional. What the work in this thesis
contributes to this postulation is that while many of the modalities used in absolute identification tasks, such as lines of varying length and tones of varying frequency, are physically uni-dimensional, they are perceived as multi-dimensional, at least for some participants, thus allowing learning to take place beyond Miller’s upper limit of nine stimuli. The exception seems to be for tones of varying intensity, which has so far showed no learning, nor complex psychological representation. This is further discussed in the section Are tones of varying intensity (loudness) the only true uni-dimensional stimulus? below.

The work in Chapter 3 of this thesis further explored the dimensionality of stimuli in regards to whether extra physical dimensions improved accuracy in absolute identification tasks. Rae (2010) showed that although there was a relationship between ability to identify piano and sine tones, piano tones were more accurately identified. As previously stated, research has shown that the accuracy in identifying stimuli increases when the number of independent dimensions increase (Eriksen & Hake, 1955; Miller, 1956; Rouder, 2001; Nosofsky & Palmeri, 1996). Musical tones, such as piano tones, consist of a fundamental frequency (by itself, uni-dimensional) and a series of overtones (multi-dimensional). The experiment described in Chapter 3 used computer generated sine tones, with and without an additional overtone, in order to investigate whether a second physical dimension increased accuracy in an absolute identification task using tones of varying frequency.

The results of this experiment (Chapter 3) showed that there was not a significant difference in the identification of sine tones with or without an additional overtone (bi- and uni-dimensional stimuli respectively). Nor was there a significant interaction with musicality, that is whether or not the participant was a musician,
with dimensionality of the stimuli. These results suggest that although the tones with an additional overtone were physically multi-dimensional, they were most likely not physically separable. Therefore, adding the findings from Chapter 2 of this thesis (Structural Forms analysis), it is a likely assumption that it is the psychological representation of the stimuli that determines whether a participant is able to learn to identify beyond Miller’s (1956) upper limit of nine stimuli, and not the physical dimensionality.

While a fairly direct relationship between absolute identification ability and having a more complex psychological representation of stimuli has been established in this thesis, this does not completely resolve the issue. There is still the question of what causes a person to have a more complex psychological representation of stimuli, compared with someone who does not. It is also unclear whether the complex psychological representation facilitates learning, or the act of learning enables the evolution of a more complex psychological representation. Future research directions are suggested towards the end of Chapter 6 that might help to disentangle this conundrum.

3. How far beyond Miller’s (1956) limit is it possible to go?

The final aim of this thesis was to investigate how far it is possible to learn beyond Miller’s (1956) upper limit of nine stimuli in an absolute identification framework. Earlier work in this thesis, along with relatively recent research (Rouder et al., 2004; Dodds et al., 2011; Rae, 2010) showed that it is possible to learn well beyond Miller’s limit for tones of varying frequency (up to 29 stimuli), lines of varying length (up to 20 stimuli), and dots varying in separation (up to almost 14 stimuli).
Chapter 5 in this thesis replicated the above findings, showing that learning beyond Miller’s (1956) upper limit of nine stimuli is possible for both tones of varying frequency and lines of varying length. Of the four participants who attempted to learn a stimulus set of 24 line lengths, three were able to exceed Miller’s limit for both blocks with and without feedback. The best performing participant learned to identify 16.3 stimuli for blocks with feedback and 13.5 for blocks without feedback. Tones of varying frequency showed less improvement in accuracy, with only two of seven participants showing any learning beyond Miller’s limit, and only for blocks with feedback. However, the amount of learning was not overly remarkable, with both participants learning to identify just below 10 stimuli.

While the amount of learning shown in Chapter 5 was not to the same high level reported in Rouder et al. (2004) or Rae (2010), the findings still allow some discussion. However, the amount of learning demonstrated would mostly likely have increased further, had the experiment allowed for more sessions. The three participants who learned to identify line lengths beyond Miller’s (1956) upper limit of nine were continuing to show strong learning right up to the last session. Similarly, two of the participants who attempted to learn tones of varying frequency were showing a good level of improvement up to the last session. Further, recruitment specifically targeted participants who had not had any modality specific training. Therefore, it could be speculated that at least a few of these participants ended the experiment below their maximum potential accuracy.

The results in Chapter 5, along with previous research (Rouder et al., 2004; Dodds et al., 2011; Rae, 2010) clearly show there are individual differences in the ability to identify uni-dimensional stimuli in an absolute identification task. The results in Chapter 5 also hint at another interesting point: There might be
fundamental differences between modalities used in absolute identification tasks. While there were only two participants who completed the experiment in both modalities, and results should be interpreted with caution, both participants showed learning to higher levels of accuracy for lines of varying length than tones of varying frequency. This may possibly be due to both participants having higher aptitudes for identifying lines over pitch on an individual level, but also suggests the possibility that lines of varying length are qualitatively different to identify than tones of varying pitch. Further, this raises the question of whether the modalities used in absolute identification tasks can be legitimately compared, and by extension, whether they can be bound by the same constraints, such as Miller’s (1956) processing capacity limit of 7±2. Perhaps theoretical accounts of absolute identification (Brown et al., 2008; Marley & Cook’s, 1984; Petrov & Anderson’s, 2004; Stewart et al., 2005) need to account for the variation in performance across modalities.

While this is obviously speculation, the idea has some merit. Clearly some modalities have shown more learning than others. Participants who have learned to identify lines of varying length have done so way beyond Miller’s (1956) upper limit of nine (Rouder et al., 2004; Dodds et al., 2011; the experiment in Chapter 5 of this thesis). Some participants who have learned to identify tones of varying frequency have done so to an even higher level (Dodds et al., 2011; Rae, 2010). Further, absolute pitch literature describes participants who are able to identify well beyond Miller’s limit (Levitin & Rogers, 2005), and Miller himself acknowledged some people are able to identify more than 50 musical notes. In contrast, any learning attempts of tones of varying intensity (loudness) have been severely limited, even despite earnest attempts (e.g. Garner, 1953; Weber et al., 1977; Dodds et al., 2011).
The highest possible learning capacity of each modality in absolute identification tasks is bound by physical limits and human perception. Identification of tones of varying frequency are bound by the range of human hearing (16 Hz to 20 kHz; Blauert, 1997), identification of lines of varying length are bound by visual processing capacity, and so on, all of which deteriorate across the aging process (Spear, 1993; Clinard, Temblay, & Krishnan, 2010). As well as being bound by the range of human perception, each modality used in absolute identification tasks would presumably also have different just noticeable differences (JNDs) not only across modalities, but also across participants. Some researchers have found cross-modality matches that suggest validity for equal-sensation functions across modalities. For example, Stevens (1956) found similar JNDs for loudness, vibration, and electric shock, and Stone and Bosley (1965) found similar JNDs for the odour of two acids. However, these have been limited to modalities that can be similarly measured. Further, Newman (1933) concluded that the JND is not a valid unit when comparing two psychological modalities. This might suggest that there is little concrete evidence that the potential for learning across the modalities used in absolute identification tasks have the same potential for learning. This is discussed further in the section Future research directions below.

A supposition based on the discussion above is that perhaps each of the modalities have different potential for psychological complexity. The research described in Chapter 2 (Dodds et al., 2012) has already hinted at this, as while lines of varying length, tones of varying pitch, and dots of varying separation showed a complex psychological representation for some participants, tones of varying intensity showed no such structure. So far the results lean toward a strong relationship between the ability to learn a modality and the psychological
representation of that modality. It would therefore make sense that if the modalities have different learning potentials, then the psychological representation for each of those modalities would also be different. Perhaps this reliant on the learning, or causing the learning – or some combination of both – depending on which way the relationship goes. Further, this warrants the question of whether each person has the same learning potential, and thus the same capacity for a complex psychological representation, for all modalities in absolute identification. Again, this is discussed further in the section *Future research directions* below.

The final point to be raised in answering the third question this thesis aimed to investigate, *How far beyond Miller’s (1956) limit is it possible to go?*, is that it might not be possible to ever know. While in theory we may speculate the potential of learning to identify uni-dimensional stimuli, the results for such a task are always likely to be confined to the artificial restraints of the laboratory in an experimental setting. Identifying uni-dimensional stimuli in isolation is not a real world task such as identifying faces, letters, distances between cars, and so forth, for which humans are much more clearly capable and reliant.

### Additional issues raised from the results of this thesis

*The use of feedback in absolute identification tasks*

One of the issues this thesis raised for discussion is the use of feedback in absolute identification tasks. All of the absolute identification tasks that were reported in Chapters 3 to 5 of this thesis used a combination of blocks with feedback and blocks without feedback. While typically only blocks with feedback have been used in absolute identification tasks (e.g. Hartman, 1954; Pollack, 1952; Garner, 1953; Lacouture, 1997; Rouder et al., 2004; Dodds et al., 2012), the results of
experiments from this thesis, along with Rae (2010) show that trial by trial feedback increases accuracy in identifying uni-dimensional stimuli.

The initial inspiration to conduct an absolute identification experiment using a combination of blocks with trial by trial feedback and blocks without any feedback for Rae (2010) came from absolute pitch literature. Musicians are explicitly trained in the use of relative pitch, that is, the ability to identify differences between notes, rather than individual notes in isolation (Miyazaki, 1995). Lockhead and Byrd (1981) suggested that any pitch identification experiment using feedback potentially provides participants with an opportunity to use relative pitch instead of, or in addition to, absolute pitch. Miyazaki (1988) proposed that a simple technique to control for relative pitch strategies is not to provide feedback at all. However, using feedback has its merits in absolute identification tasks, as is often used to aid with motivation, anchoring, and morale.

In order to determine whether relative pitch strategies have been used in an absolute pitch task that has used trial by trial feedback, Pikler (1955) suggested that analysing accuracy as a function of the previous interval may reveal that more commonly identified intervals, such as unisons, semitones, and octaves, have a higher accuracy rate if relative pitch is being used. This strategy was employed in Rae (2010) and it was found this was the case: unisons, semitones, and octaves were more accurately identified when trial by trial feedback was provided. This strategy is not conclusive in disentangling the use of absolute versus relative pitch strategies. However, it certainly provides strong evidence that relative pitch strategies are being employed instead of, or at the very least as well as, absolute pitch strategies when trial by trial feedback is provided.
For this reason, all absolute identification experiments conducted as part of this thesis used blocks both with and without trial by trial feedback. Accuracy for blocks with feedback was shown to be improved across all experimental tasks. The pitch identification task reported in Chapter 3 showed blocks with feedback were more accurately identified than blocks without, as did the pitch identification task reported in Chapter 4. Further, both the pitch identification task and the task using lines of varying length reported in Chapter 5 showed similar patterns of increased accuracy for blocks that provided trial by trial feedback. It must be noted that in Chapter 5 these results all included participants who were novices, that is, not trained in music (for the tones of varying frequency modality) or a related profession that would provide advantage (for lines of varying length). This suggests that the use of relative identification strategies relying on feedback are not limited to trained musicians in pitch identification tasks.

While there is only experimental evidence that shows accuracy is improved by the provision of trial by trial feedback for tones of varying frequency and lines of varying length (Chapter 3 to 5 of this thesis; Rae, 2010), it is likely that this is also the case for other modalities in absolute identification, such as dots varying in separation. Based on these findings, this calls into question whether previous research has actually been measuring absolute identification, or instead some combination of absolute and relative identification strategies. It is possible that the learning demonstrated in Rouder et al. (2004) and Dodds et al. (2012) was not due solely to absolute identification strategies, but instead with a mixture of absolute and relative identification strategies that could only have been unravelled by using a combination of blocks with and without trial by trial feedback.
While there is certainly enough evidence to suggest that the truer measure of absolute identification is without feedback, it does not suggest that all previous experiments that have measured accuracy only using feedback are now without merit. All of the experiments reported in this thesis (Chapters 3 to 5) and Rae (2010) showed that while feedback did improve accuracy, there was also a strong relationship between accuracy with and without feedback, both collectively and on an individual level. This suggests that the learning reported in previous literature, such as in Rouder et al. (2004) and Dodds et al. (2011), would likely have still been apparent, but possibly not to such a high level. It could be that the amount of learning would still be approximately the same as with feedback, but the level of accuracy would have been lower without feedback. Based on these findings and discussion, it would be recommended that future absolute identification experiments use a mixture of blocks with and without trial by trial feedback, as has been demonstrated successfully in this thesis. Blocks with feedback could still be included for the purpose of anchoring, learning, and motivation, but true absolute identification should be measured from blocks using no feedback.

A further implication for the findings that the use of feedback provides the opportunity for participants to use relative pitch strategies instead of, or in addition to, absolute pitch strategies is that theories of absolute identification might have to consider this. As presented in Chapter 1 (summarised from Donkin et al., 2015) there are several classes of theories of absolute identification that attempt to account for the benchmark phenomena associated with the paradigm. While all classes are likely to be somewhat affected by the findings from this thesis (e.g. Thurstonian models, Exemplar models, and Restricted Capacity Models) the class of absolute identification models most likely affected by the results are the Relative Judgement
Models (e.g. Holland & Lockhead, 1968; Laming, 1984; Stewart et al., 2005). As discussed in Chapter 1, relative judgement models’ defining feature is that decisions are based on the difference between current and previous stimuli (or responses), rather than being based directly on representation of the absolute magnitudes of stimuli. These models make use of the fact that trial by trial feedback directly affects the subsequent responses, and so the finding that performance on absolute identification tasks decreases in accuracy when feedback is withheld will likely need to be considered in future workings on the models.

**Are tones of varying intensity (loudness) the only true uni-dimensional stimulus?**

As discussed above in the section *How far beyond Miller’s (1956) limit is it possible to go?*, the issue of the potential differences across modalities used in absolute identification tasks was raised. Chapter 5 of this thesis reported several participants’ learning beyond Miller’s (1956) upper limit of nine stimuli for tones of varying frequency and lines of varying length, which is reminiscent of previous literature (Rouder et al., 2004; Dodds et al., 2011; Rae, 2010). Dodds et al. also reported on learning shown for dots of varying separation. While these three modalities showed different rates and levels of learning, they still demonstrated what was long thought to be impossible: Accuracy beyond Miller’s upper limit of nine stimuli is possible.

In contrast to this, a modality that has consistently shown resistance to learning is tones of varying intensity (loudness). Chapter 2 of this thesis used the data from Dodds et al.’s (2011) absolute identification learning experiments along with data from Rae (2010) and data from Rouder et al. (2004), to examine the psychological representation of the stimuli using Kemp and Tenenbaum’s (2008) Structural Forms.
algorithm. While the other three stimuli – lines of varying length, tones of varying frequency, and dots of varying separation – showed both learning beyond Miller’s (1956) upper limit of nine stimuli and a psychological complexity, at least for some participants, tones of varying intensity showed neither the ability to be learned, nor a psychological complexity. These results are consistent with previous attempts at learning tones of varying intensity that have failed (Garner, 1953; Weber et al., 1977; Shiffrin and Nosofsky, 1994).

The question posed from these findings is what makes the modality of tones of varying intensity unique. Results suggest that they are different to other modalities employed in absolute identification tasks, in that they are not able to be learned, and so far have not shown to elicit a complex psychological representation. Two issues can be raised from this. Firstly, it is possible that tones of varying intensity are not unique, but simply more research needs to be done across a wider range of modalities, such as saltiness, hue, and brightness, to determine if there are more uni-dimensional modalities that are truly unable able to be learned or elicit a complex psychological representation, thus being truly bound by Miller’s (1956) upper limit of nine stimuli.

No learning has been shown to date for tones of varying intensity, nor has a complex psychological representation. However, this does not mean that this will always be the case. For years, many researchers reported that other stimuli such as lines of varying length were not able to be learned in an absolute identification task, and it is only relatively recently that this has been proven to be untrue. Further, it is only in recent years that some modalities have shown the potential to have a complex psychological representation, using advanced techniques with the possibility for a framework of inference such as the Structural Forms algorithm reported in Chapter 2.
of this thesis (Kemp & Tenenbaum, 2008; Dodds et al., 2012). For example, Rouder et al. (2004; citing research: Durlach & Braid, 1969; Gravetter & Lockhead, 1973; Karpiuk, Lacouture, & Marley, 1997; Luce et al., 1976; Marley & Cook, 1984) reported that the perception of lines of varying length along with tones of varying intensity, are assumed not to correlate with a complex psychological representation. It is now clear that this is not the case for lines of varying length, as shown in Chapter 2 of this thesis. Instead, it is possible that future research techniques may also show that tones of varying intensity are able to elicit a similar psychological representation, at least for some participants. This is further discussed below in the section Future research directions.

**Perhaps the paradigm of absolute identification cannot account for exceptional levels of pitch identification ability described in absolute pitch literature**

The paradigm of absolute pitch reports the ability of some rare people who are able to identify every note in the musical scale (over 90 tones; Levitin & Rogers, 2005). This ability is said to be seemingly effortless and people with this ability often report it has always been there (Takeuchi & Hulse, 1993). This seems in contrast to the comparatively low processing capacity limitation of 7±2 that Miller (1956) reported by which absolute identification is bound.

It is possible that the exceptional levels of pitch identification ability reported in absolute pitch literature cannot be completely accounted for by the paradigm of absolute identification, nor more broadly within the scope of the discipline of learning and cognition. Levitin and Rogers (2005) suggest that the answer may lie, at least in part, with a genetic disposition towards some of the underlying traits necessary for the development of absolute pitch. Zatorre (2003) argues that people
with absolute pitch show neuroanatomical differences as well as different results on neural tests when compared with people who do not have absolute pitch. This difference is especially evident on tests of working memory. When listening to transposed tone sequences, people without absolute pitch ability show a mismatched negativity and an attentive P3 evoked response potential. That is, they show the activation of working memory (Fujikoa, Trainor, Ross, Kakigi, & Pantev, 2004; Shahin, Bosnyak, Trainor, & Roberts, 2003). In contrast, people with absolute pitch seem to instead be using long-term memory (Hirose, Kubota, Kimura, Ohsawa, Yumoto, & Sakakihara, 2002), which Zatorre (2003) and Zatorre and Beckett (1989) suggest is because people with absolute pitch do not need to use working memory to keep a psychological representation of pitch active.

Another area of research that is often linked to superior absolute pitch identification ability is that of the savant-like quality some people have with the neurodevelopmental disorder of Autism. Several studies have shown that people with autism have a higher incidence of having absolute pitch ability, even when not trained as musicians. For example Heaton, Hermelin, and Pring (1998) found that musically naïve children with autism demonstrated a superior pitch identification ability compared with musically naïve mental-age matched controls. It has been suggested that this superiority might be due to an abnormally high sensitivity to subtle pitch differences (Bonnel, Mottron, Peretz, Trudel, Gallun, & Bonnel, 2003).

Furthermore, Williams Syndrome (a condition caused by a microdeletion of about 20 genes in the q11.23 region of one of their two chromosomes number seven) has also been shown to be associated with higher levels of absolute pitch abilities when compared with the general population. This is possibly in part due to an
extended “critical period” where acquisition of pitch identification ability normally takes place (Lenhoff, Parales, & Hickok, 2001).

The issues with the association of absolute pitch abilities being higher amongst people with autism and Williams Syndrome are too complex to outline in this thesis. However, it is still relevant to mention these issues as the ability for normal participants (that is, participants without any developmental disorders or syndromes) to surpass Miller’s (1956) upper limit of nine stimuli in an absolute identification task is still seen as exceptional. Perhaps the quality that people with autism or Williams Syndrome possess may hold the key for future studies, as it is still not known what actually causes some people to have a psychological representation of seemingly physically uni-dimensional stimuli, or the ability to learn to identify beyond Miller’s upper limit of nine.

Research (cited above) does suggest that having absolute pitch ability does relate to a genetic predisposition and neural correlates. However, as Zatorre (2003) suggests, even clear proof of a genetic link would not negate the importance of other factors in the absolute pitch acquisition process. Further, a genetic predisposition might be shown to be necessary, but is clearly not sufficient in itself (Levitin & Rogers, 2005). In any case, learning must be an essential aspect of the acquisition of absolute pitch as it would be virtually impossible for a person to innately know the arbitrary tones of the Western musical scale, complete with labelling, such as 440Hz is A4, without any environmental context or training. For this reason, studying absolute pitch within a cognitive framework and, more specifically, within the paradigm of absolute identification, still has merit. A more complete picture might be given in conjunction with other related psychological disciplines such as biopsychology and neuropsychology.
Future research directions

The discussion that has arisen based on the results of this thesis, in conjunction with previous cited research has not only attempted to answer the three questions posed at the beginning of this thesis (see section Answering the aims of this thesis above) but has also raised some points for future research. More specifically, the discussion earlier in Chapter 6 has raised questions for further investigation such as whether there are fundamental differences in the modalities used in absolute identification tasks, and whether the ability to exceed Miller’s (1956) upper limit of nine stimuli is associated with any other processing capabilities. Future research directions that might help to address these issues are outlined below.

It is suggested that more research needs to be done in order to investigate how far it is possible to learn in an absolute identification task. While Miller’s limit (1956) was long reported to be unbreakable, it is now clear that this is not the case, at least for some participants. Given the right circumstances, with a larger sample of highly motivated individuals, over an extended period of time, with increasingly large stimulus set sizes, perhaps it would be possible to reveal the true potential of learning for people who are able to break Miller’s upper limit of nine stimuli. Ideally this experiment would also be across a large number of modalities for comparison. This might show whether or not the modality of tones of varying intensity is truly bound by Miller’s (1956) limit, or whether it is able to be learned by a few rare individuals. This type of experiment could lead to another point of investigation: Whether people who have a complex psychological representation for one type of stimuli have it for other modalities. A long experiment such as the one suggested could provide many data points for confusion matrices to be constructed as was done in Chapter 2 of this thesis (Kemp & Tenenbaum, 2008; Dodds et al., 2012) which
enabled the use of a Structural Forms algorithm to identify the psychological complexity of the stimuli.

To extend this idea further, before an attempt at learning to identify the stimuli is made, it would be ideal to first calibrate the JNDs for each participant, and for each modality, possibly using a procedure such as the Method of Constant Stimuli. As discussed above, while it might not be possible to make direct cross-modality comparisons across the different stimuli used in absolute identification tasks using just noticeable differences (Newman, 1933), this could be used for a different, but somewhat related, purpose. It would be of interest to compare the complete range of each participant’s perception of each separate stimulus, and also to see how many discriminable stimuli they can distinguish, both across participants, and across modalities. This might lead to two interesting points for investigation. Firstly, it would allow for a direct comparison between modalities, especially in terms of what the maximum number of identifiable stimuli would be, even in theory if not practically. Secondly, this might allow for a more complete investigation of the relationship between stimuli and their potential for a complex psychological representation. That is, perhaps if there are only a small number of discernible levels within a modality, this might contribute to the explanation of why a complex psychological representation has not been elicited, when compared with a modality with a high number of discernible levels.

While the experiment described above sounds rather ambitious and may not be entirely practical, theoretically this would be a robust way to determine a number of factors that contribute to the ability to learn to identify beyond Miller’s (1956) upper limit of nine stimuli in an absolute identification task. Perhaps some subset of these tasks could be undertaken in an attempt to further understand the relationship
between learning potential and psychological representation across modalities, at least for the exceptional participants who are able to demonstrate learning beyond Miller’s (1956) upper limit of nine stimuli. Further, as discussed previously in Chapter 6, to reveal the complete picture of the ability to identify uni-dimensional stimuli beyond Miller’s upper limit this may require collaboration with other related psychological disciplines such as neuropsychology and biopsychology.

A future research direction suggested by an examiner of this thesis is that perhaps participants might be able to improve their learning of uni-dimensional stimuli by having specific training of a multidimensional psychological representation of the stimuli. The results of the Structural Forms analysis in Chapter 2, along with the findings from Rae (2010; discussed in Chapter 1) show that people who have a more complex psychological representation of physically uni-dimensional stimuli have a more accurate identification of that stimuli. This is particularly apparent for tones of varying frequency that has been shown to have a helical representation, especially for trained musicians (see discussion in Chapter 1). It is likely that musical training explicitly trains participants for a multidimensional psychological representation as the cyclical octave structure lends itself to this, and this is likely where the helical representation arises from. Given this, it is likely that the superior performance of musicians in identifying tones is not only attributable to extended practice of labelling tones, but also for their formation of the underlying complex psychological representation of the tones. This quite possibly might explain the relatively poor performance of the musicians in Chapter 3’s work on non-Western tones: Musicians did not have the underlying helical structure supporting the identification of familiar Western tones.
A future experiment, suggested by one of the examiners of this thesis, is to run a similar experiment to the one in Chapter 5 where participants complete many hours of training of uni-dimensional stimuli. However, instead of labelling the lines of varying length with arbitrary numbers (such as 1 through 24), they could be given labels that imply a helical or circular structure. For example, stimulus 1 might be named A-Dax, stimulus two A-Fep, stimulus three A-Vuf, and so on, and then stimulus 7 might be named B-Dax, stimulus eight B-Fep, stimulus nine B-Vuf, and so on. If participants are able to explicitly learn to impose a multidimensional psychological representation onto stimuli, then it would stand to reason that their performance in identifying the lines of varying length would improve even more so than the learning demonstrated in the experiment in Chapter 5. A Structural Forms analysis (Kemp & Tenenbaum, 2008) such as the one conducted in Chapter 2 of this thesis could reveal whether the participants had managed to internalise such a representation.

A further future research direction that would also be interesting would be to investigate the retention of learning of the participants who have shown to exceed Miller’s (1956) upper limit of nine stimuli in absolute identification tasks. It would be worthwhile to see if the participants who were able to break Miller’s limit for lines of varying length, tones of varying pitch, and dots of varying separation (Rouder et al., 2004; Dodds et al., 2011; Rae, 2010; the findings from Chapter 5 of this thesis) were able to demonstrate the high levels of accuracy in identifying the uni-dimensional stimuli that they were previously reported to do at the end of the learning tasks, or at least demonstrate a starting accuracy above their previous starting point. Further, it would also be of interest whether these participants were able to reach higher levels of accuracy more quickly than they did previously. These
results could give some insight into how the information in the learning task in absolute identification experiments is encoded; whether the ability to identify the stimuli during the experimental sessions is due to some type of working memory mechanics, or if it is being more deeply encoded in long-term memory.

A final direction to be discussed for future research is whether the ability to learn to identify uni-dimensional stimuli beyond Miller’s (1956) upper limit of nine, as well as eliciting a complex psychological representation of the stimuli, is related to higher level cognitive skills such as executive function. Recent research (Evans, Rae, Bushmakin, Rubin, & Brown, 2017) has identified a link between individual differences in the speed-accuracy trade off, that is the quantitative balance between caution and urgency in decision making tasks (for more information see: Wickelgren, 1977; Starns & Ratcliff, 2012; Evans & Brown, 2016), with the personality trait need for closure, where people have, amongst other qualities, high levels of decisiveness (Kruglanski, 1989; Webster & Kruglanski, 1994). Similarly, it is possible that the exceptional ability to identify and hold a complex psychological representation of uni-dimensional stimuli may be related to higher order cognitive skills such as executive function.

Executive function refers to higher-level abilities in areas such as problem-solving and mental flexibility (Shannon & Thomas-Duckwitz, 2011). More specifically, metacognitive skills are strategies such as self-regulation, coordination, planning and self-monitoring, and are applied either consciously or unconsciously during cognitive processes (Patterson, 2011; Zhou & Cunningham, 2011). They have been reported as being highly relevant to learning (Flavell, 1976, 1979; for more information see Veenman, Elshout, & Meijer, 1997). Performance on a neuropsychological assessment such as the Wisconsin Card Sorting Test (Grant &
Berg, 1948) might reveal a correlation between the ability to learn to identify beyond Miller’s (1956) upper limit of nine stimuli in an absolute identification task and executive function abilities. More specifically, it might show a link between those participants who are able to elicit a complex psychological representation of physically uni-dimensional stimuli and good performance on an executive function task. Further, neuroimaging studies and functional magnetic resonance imaging (fMRI) procedures have shown connections to executive function and metacognition (Cunningham & Zhou, 2011; Osaka, 2007; Smith & Jonides, 1997). As discussed above in the section Additional issues raised from the results of this thesis, perhaps the collaboration with other disciplines within the realm of psychology, such as biopsychology and neuropsychology, might be necessary to give a complete picture of what affords some people the ability to generate complex psychological representations of the uni-dimensional stimuli in absolute identification tasks.

Conclusion

While it was long thought that Miller’s (1956) processing capacity limit of $7 \pm 2$ was unable to be exceeded in identifying uni-dimensional stimuli in an absolute identification task, recent work has proven otherwise. While it is still the exception, rather than the rule, breaking Miller’s limit is not as rare an ability as previously thought. The ability appears to be due to individual differences such as initial performance and the complexity of the psychological representation of the stimuli. The origin of this exceptional ability is still not understood, but evidence suggests there is a learning component, although there are hints that there might also be an innate element that is required for learning to be possible. It is also recommended
that future studies measure absolute identification where no feedback is given, as this seems to be the truer measure of the ability.
References


Dodds, P., Rae, B., & Brown, S. D. (2012). Perhaps undimensional is not


Pollack, I. (1952). The information of elementary auditory displays. *Journal of


Appendices
Participant Questionnaire

The following questions relate to any musical training or experience you may have had:

Have you ever learned to play a musical instrument (including vocals/singing)? Yes / No

If yes, which instrument/s?

To what level have you learned (e.g. AMEB grades, or years playing)

How old were you when you start learning these instrument/s?

Have you ever had training in music theory or musicianship? Yes / No

If yes, please describe:

Have you ever had aural training in music? Yes / No

If yes, please describe:

How would you rate your overall level of musical proficiency? (please circle)

Beginner Intermediate Advanced Virtuoso

The following questions relate to languages that you can speak:

Before you were 8 years old, what was the main language/s that you spoke:

After you were 8 years old, what is the main language/s that you have spoken:

What other languages have you spoken throughout your life (and to what level of fluency?)

Please turn over for more questions

The following questions relate to your age and hearing ability:
Please circle the age group that best describes you:

Under 18   18 to 25   26 to 35   36 to 45   46 to 55   56 to 65   Over 66

Have you ever had hearing difficulties? Yes / No

    If yes, please describe:

The following question relates to how you perceive your pitch identification ability:

Absolute Pitch (also sometimes known as “Perfect Pitch”) is the ability to identify a tone that is sounded in isolation. This is often tested by hearing a note played on a musical instrument (such as a piano) where a participant would try to name the note they heard.

If you were played 10 separate notes, how many of these would you expect to be able to label correctly? (please circle)

    0   1   2   3   4   5   6   7   8   9   10

Thank you for taking the time to complete this questionnaire
Appendix 2: Participant Questionnaire from Chapter 4

音调语言问卷调查

Tonal Language (Al) Questionnaire

语言(LANGUAGES)

(以下问题是针对你在8岁前所讲的语言)
(These questions are about languages you spoke before you were 8 years old)

1. 你在8岁前讲得最多的语言是？
   (What was the language you spoke the most before you were 8 years old?)

2. 你在8岁前还讲其他语言吗？
   (Did you speak any other languages before you were 8 years old?)

3. 你的家人有讲其他语言吗 (你不会讲的语言)？
   (Were there languages spoken by anyone in your household, but that you did not speak yourself?)

(以下问题是关于你从8岁到现在所讲的语言)
(These questions are about languages you have spoken since you were 8 years old until now)

4. 从8岁到现在，你主要讲的语言是
   (Which is the main language you have spoken since you were 8 years old?)

5. 从8岁到现在，你有没有还讲过其他语言
   (Have you spoken any other languages since you were 8 years old?)

以下问题是关于你去年一年里所讲的语言
(These questions are about the languages you have spoken in the last year)

6. 去年一年里你主要讲的语言是
   (Which is the main language you have spoken in the last year?)

7. 去年一年里你有没有还讲过其他语言
   (Have you spoken any other languages in the last year?)
人口统计信息 (DEMOGRAPHICS)

以下是关于你居住地的问题
(These questions are about where you have lived)
8. 请列出你在8岁前居住过的国家
(Please list the country/countries you lived in before you were 8 years old)
9. 请列出你在8岁以后居住过的国家
(Please list the country/countries you lived in since you were 8 years old)

音乐素养 (MUSICAL EXPERIENCE)

一下问题是关于你已有的音乐知识或训练
(These questions are about any musical experience or training you have had)
10. 你在8岁以前是否接受过以下音乐训练
(Do you have any of the following musical training before you were 8?)
□ 乐器或声乐课程 (Instrument or voice lessons) 如果有，是什么乐器 (If yes, which instrument/s)
□ 音乐理论和鉴赏课程 (Theory or musicianship lessons)
□ 考试和成绩 (Exams/grades) 如果有，到什么级别 (If yes, to which level)

11. 你在8岁以后是否接受过以下音乐训练
(Do you have any of the following musical training after you were 8?)
□ 乐器或声乐课程 (Instrument or voice lessons) 如果有，是什么乐器 (If yes, which instrument/s)
□ 音乐理论和鉴赏课程 (Theory or musicianship lessons)
□ 考试和成绩 (Exams/grades) 如果有，到什么级别 (If yes, to which level)

绝对音准 (ABSOLUTE PITCH)

以下问题有关于绝对音准，又称完美音准
(These questions are about absolute pitch - which is also called “perfect pitch”)
12. 你认为自己有绝对音准吗? (请圈出) ( 有 / 没有 / 我不清楚)
(Do you know if you have absolute pitch (Please circle) Yes / No / I don't know)
13. 如果上一题你回答了“有”：如果你听到10个音符，你预计你能完全正确地识别几个?
(If you answered “yes” to question 12: If you heard 10 notes, how many of those would you expect to be able to identify correctly? Please circle.)

请圈出：
0 1 2 3 4 5 6 7 8 9 10

谢谢您花时间完成这份问卷。
(Thank you for taking the time to fill out this survey)
Appendix 3.1

Learning experiment of lines of varying length. Each participant completed 20 sessions with seven blocks in each. All blocks gave trial by trial feedback except for the final (seventh) block of each session which had no feedback at all.

▲ = above Miller’s upper limit of 9 stimuli

<table>
<thead>
<tr>
<th>Participant</th>
<th>Feedback</th>
<th>Proportion Correct (Equivalent Stimuli)</th>
<th>Total increase in proportion correct from first to last sessions (Equiv. stimuli)</th>
<th>One tailed one sample t-test of mean of last sessions (μ = .375; equiv. to 9 stimuli)</th>
<th>Binomial test (p = .375; equiv. to 9 stimuli)</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean of First Sessions (1 &amp; 2)</td>
<td>Mean of Last Sessions (19 &amp; 20)</td>
<td>df</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>A</td>
<td>with</td>
<td>.222 (5.3)</td>
<td>.410 (9.8▲)</td>
<td>11</td>
<td>2.26</td>
<td>.022*</td>
</tr>
<tr>
<td></td>
<td>without</td>
<td>.192 (4.6)</td>
<td>.442 (10.6▲)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>with</td>
<td>.302 (7.2)</td>
<td>.678 (16.3▲)</td>
<td>11</td>
<td>20.96</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td></td>
<td>without</td>
<td>.325 (7.8)</td>
<td>.563 (13.5▲)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>with</td>
<td>.268 (6.4)</td>
<td>.522 (12.5▲)</td>
<td>11</td>
<td>9.10</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td></td>
<td>without</td>
<td>.250 (6.0)</td>
<td>.525 (12.6▲)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>with</td>
<td>.239 (5.7)</td>
<td>.326 (7.8)</td>
<td>11</td>
<td>-2.85</td>
<td>.992</td>
</tr>
<tr>
<td></td>
<td>without</td>
<td>.275 (6.6)</td>
<td>.221 (5.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3.2
Learning experiment of tones of varying pitch. Each participant completed 20 sessions with seven blocks in each. All blocks gave trial by trial feedback except for the final (seventh) block of each session which had no feedback at all.
▲ = above Miller’s upper limit of 9 stimuli

<table>
<thead>
<tr>
<th>Participant</th>
<th>Feedback</th>
<th>Proportion Correct (Equivalent Stimuli)</th>
<th>Total increase in proportion correct from first to last sessions (Equiv. stimuli)</th>
<th>One tailed one sample t-test of mean of last sessions (μ = .375; equiv. to 9 stimuli)</th>
<th>Binomial test (p = .375; equiv. to 9 stimuli)</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean of First Sessions (1 &amp; 2)</td>
<td>Mean of Last Sessions (19 &amp; 20)</td>
<td>df</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>A</td>
<td>with</td>
<td>.363 (8.7)</td>
<td>.412 (9.9) ▲</td>
<td>.049 (1.2)</td>
<td>11</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.213 (5.1)</td>
<td>.229 (5.5)</td>
<td>.016 (0.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>without</td>
<td>.293 (7.0)</td>
<td>.397 (9.5) ▲</td>
<td>.104 (2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>with</td>
<td>.196 (4.7)</td>
<td>.250 (6.0)</td>
<td>.054 (1.3)</td>
<td>11</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.232 (5.6)</td>
<td>.263 (6.3)</td>
<td>.031 (0.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>with</td>
<td>.163 (3.9)</td>
<td>.163 (3.9)</td>
<td>0 (0)</td>
<td>11</td>
<td>-9.23</td>
</tr>
<tr>
<td></td>
<td>without</td>
<td>.163 (3.9)</td>
<td>.163 (3.9)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3.3

Learning experiment of tones of varying frequency. Participants chose to complete between three and seven blocks per session, and also chose to end the experiment when they felt they had learned as much as possible or no longer wanted to continue. All blocks gave trial by trial feedback except for the final block of each session which had no feedback at all (block three to seven, depending on the length of the session). ▲ = above Miller’s upper limit of 9 stimuli

<table>
<thead>
<tr>
<th>Participant</th>
<th>Number of sessions</th>
<th>Number of blocks</th>
<th>Feedback</th>
<th>Proportion Correct (Equivalent Stimuli)</th>
<th>Total increase in proportion correct from first to last sessions (Equiv. stimuli)</th>
<th>One tailed one sample t-test of mean of last sessions (μ = .375; equiv. to 9 stimuli)</th>
<th>Binomial test (p = .375; equiv. to 9 stimuli)</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>18</td>
<td>86</td>
<td>with</td>
<td>.305 (7.3)</td>
<td>.354 (8.5)</td>
<td>.049 (1.2)</td>
<td>.0004</td>
<td>.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>without</td>
<td>.250 (6.0)</td>
<td>.275 (6.6)</td>
<td>.025 (0.6)</td>
<td>.999</td>
<td>.0006</td>
</tr>
<tr>
<td>G</td>
<td>31</td>
<td>123</td>
<td>with</td>
<td>.254 (6.1)</td>
<td>.335 (8.0)</td>
<td>.081 (1.9)</td>
<td>.001</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>without</td>
<td>.145 (3.5)</td>
<td>.188 (4.5)</td>
<td>.043 (1.0)</td>
<td>.999</td>
<td>.0007</td>
</tr>
<tr>
<td>H</td>
<td>29</td>
<td>112</td>
<td>with</td>
<td>.221 (5.3)</td>
<td>.365 (8.8)</td>
<td>.144 (3.5)</td>
<td>.0007</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>without</td>
<td>.220 (5.3)</td>
<td>.238 (5.7)</td>
<td>.018 (.4)</td>
<td>.999</td>
<td>.0003</td>
</tr>
<tr>
<td>J</td>
<td>20</td>
<td>84</td>
<td>with</td>
<td>.280 (6.7)</td>
<td>.314 (7.5)</td>
<td>.034 (0.8)</td>
<td>.0007</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>without</td>
<td>.208 (5.0)</td>
<td>.250 (6.0)</td>
<td>.042 (1.0)</td>
<td>.999</td>
<td>.0003</td>
</tr>
</tbody>
</table>