WHEN LESS IS SOMETIMES MORE:
INVESTIGATING THE INTERPLAY BETWEEN META-COGNITION AND
MEMBER-TO-GROUP GENERALISATION

Kylie M’Intyre
BA. Ed. (Hons II)
BA. Arts (Hons I)

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in
the School of Psychology at The University of Newcastle, July 2010
STATEMENT OF ORIGINALITY

This thesis contains no material which has been accepted 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 this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

Signed .............................................
Date .............................................
I gratefully acknowledge the support and dedication of my supervisor Dr. Stefania Paolini. Her assistance and understanding has made this process a wonderful learning experience. I also acknowledge the support and feedback from Prof. Andrew Heathcote. I am forever grateful to Michael Smith whose love and support over the past few years has enabled me to survive this journey, and to Abbie who has enriched my life over the last two years. You are both my inspiration! I would also like to thank Dr. Glen McIntyre, who gave me the type of encouragement that only a brother can (honestly) provide. I also thank all of my family, especially Mum and Dad who encouraged me to love the process of discovery and to Marianne for her support over the past few years. Thanks also to my friends who tolerated me during this journey Peggy Aberg, Keiva Baker, Megan Etchells, Erica Harker and Tasmin Gillett. I would also like to extend my gratitude to all the staff at CCGS who have supported me in this journey. Finally, the financial support of The University of Newcastle Scholarship Central and of The University of Newcastle’s School of Psychology is also gratefully acknowledged.
# LIST OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement of Originality</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>List of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>ix</td>
</tr>
</tbody>
</table>

**Chapter 1: Member-to-group generalisation: Alternative Approaches to a Multi-Disciplinary Issue**

- The Pervasive Nature of Inductive Reasoning  
  - Member-to-Group Generalisation  
    - Traditional Approaches to Member-to-Group Generalisation  
      - A New Approach to Member-to-Group Generalisation  
- Measuring Generalisation  
  - Classic Measures  
  - Implicit Group Measures  
- Conclusions  

**Chapter 2: Generalisation Measures and Other Key Moderators: Meta-Analytic Review (1)**

- Past Reviews of Member-to-Group Generalisation  
  - The Present Review  
    - Method  
    - Results  
- General Discussion  
- Coda  

**Chapter 3: Member-to-Group Generalisation and Meta-Cognitions: Meta-Analytic Review (2)**

- Basic Meta-Cognitive Concepts Applicable to Member-to-Group Generalisation  
  - Meta-Analysis Study 2  
  - Method  

iv
<table>
<thead>
<tr>
<th>Chapter 8: Investigating the Source of Meta-Information and the Influence of Typicality (Study 5)</th>
<th>146</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>151</td>
</tr>
<tr>
<td>Results</td>
<td>154</td>
</tr>
<tr>
<td>Discussion</td>
<td>161</td>
</tr>
<tr>
<td>External vs. Internal Meta-Information</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 9: General Discussion: Summary of Findings, Applications, and Future Research</th>
<th>167</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of Key Findings</td>
<td>167</td>
</tr>
<tr>
<td>A Broader Theoretical and Empirical Approach to Member-to-Group Generalisation</td>
<td>170</td>
</tr>
<tr>
<td>Methodological and Social Implications</td>
<td>176</td>
</tr>
<tr>
<td>Limitations of the Present Work and Considerations for Future Work</td>
<td>178</td>
</tr>
<tr>
<td>Concluding Remarks</td>
<td>181</td>
</tr>
</tbody>
</table>

| References | 182 |

<table>
<thead>
<tr>
<th>Appendices</th>
<th>212</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix I: Concentrated-Dispersed Meta-Analysis</td>
<td>212</td>
</tr>
<tr>
<td>Appendix II: Pictures Included in the IAT</td>
<td>213</td>
</tr>
<tr>
<td>Appendix III: Experimental Profiles Used in Study 2</td>
<td>214</td>
</tr>
<tr>
<td>Appendix IV: Study 2 Mediation Analysis of the Easy-Thus-Liked Heuristic</td>
<td>217</td>
</tr>
<tr>
<td>Appendix V: Study 2 Mediation Analysis of the Easy-Thus-Typical Heuristic</td>
<td>218</td>
</tr>
<tr>
<td>Appendix VI: Study 3 Mediation Analysis of the Easy-Thus-Liked Heuristic</td>
<td>220</td>
</tr>
<tr>
<td>Appendix VII: Study 3 Mediation Analysis of the Easy-Thus-Typical Heuristic</td>
<td>221</td>
</tr>
<tr>
<td>Appendix VIII: Experimental Profiles Used in Study 5</td>
<td>223</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1: Tests Meeting Inclusion Criteria for Meta-Analytic Study 1 34
Table 2.2: Summary Data for Overall Generalisation and each Outcome Measure 37
Table 2.3: Summary Data for Sample Size Across all Outcome Measures 40
Table 2.4: Summary Data for Exemplar Typicality Across all Outcome Measures 41
Table 2.5: Summary Data for Exemplar Valence Across all Outcome Measures 43
Table 2.6: Summary Data for Out-Group Valence across all Outcome Measures 43
Table 2.7: Summary Data for Exemplar Valence across all Outcome Measures 44
Table 2.8: Summary Data for Out-Group Type across all Outcome Measures 45
Table 3.1: Coding of Tests Meeting Inclusion Criteria for Meta-Analytic Study 2 61
Table 3.2: Summary of Effect Sizes for Overall Generalisation and the Outcome Measures 64
Table 3.3: Summary of Effect Sizes for Cue Source 65
Table 5.1: Change in Group Measures Relative to Baseline as a Function of Number of Exemplars 92
Table 6.1: Outcome Measures as a Function of Meta-Information Cue, Information Source and Number of Exemplars 109
Table 7.1: Outcome Measures Among Active Elderly Participants as a Function of Meta-Information Cue and Number of Exemplars 130
Table 7.2: Outcome Measures as a Function of Meta-Information Cue and Number of Exemplars 140
Table 8.1: Outcome Measures as a Function of Information Source, Exemplar Typicality and Task Distraction 156
LIST OF FIGURES

Figure 2.1: Meta-regression as a Function of Total Sample Size 39
Figure 2.2: Meta-regression as a Function of Disconfirming Sample Size 40
Figure 6.1: Change in Group Stereotyping for Meta-Information Cue: Absent Participants as a Function of Information Source and Number of Exemplars 110
Figure 6.2: Change in Group Stereotyping for Meta-Information Cue: Present Participants as a Function of Information Source and Number of Exemplars 111
Figure 7.1: Change in Group Stereotyping for Participants Retrieving Active Elderly Exemplars as a Function of Meta-Information Cue and Number of Exemplars 130
Figure 7.2: Explicit Stereotype Change as a Function of Meta-Information Cue and Number of Exemplars 140
Figure 8.1: Stereotype Change by Exemplar Typicality and Task Distracter for Presented Information Participants 157
Figure 8.2: Stereotype Change by Exemplar Typicality and Task Distracter for Self-Generation Participants 158
ABSTRACT

Fluency is a meta-cognitive process whereby people apply the ease associated with retrieving or processing stimuli to a subsequent content-relevant judgment (Schwarz 1998; Winkielman & Cacioppo, 2001). In the present work, retrieval fluency and, to a certain extent, processing fluency are applied to the social cognitive process of member-to-group generalisation. During member-to-group generalisation, information about the individual members of a social group affects the judgment of the group as a whole (Paolini, Hewstone, Rubin, & Pay, 2004). In Chapter 1, I trace the development of member-to-group generalisation research and identify the contact hypothesis and the social-cognitive approach as two distinct approaches to generalisation. In Chapter 2, I meta-analytically review social cognitive studies that used an impression formation paradigm to investigate member-to-group generalisation. In Chapter 3, I review retrieval fluency research that had been applied to the investigation of social group judgments and meta-analytically review impression-formation studies that had investigated meta-cognitive processes. In Chapter 4, I introduce a dual process model of the effects of retrieval fluency on member-to-group generalisation and advance a self-generation paradigm to test for the influence of fluency on member-to-group generalisation. In Chapter 5, I present a first test of the new self-generation paradigm and demonstrate that retrieval fluency influences member-to-group generalisation. In Chapter 6, I combine the self-generation paradigm with the impression formation paradigm in a single research design and demonstrate that both approaches to retrieval fluency are appropriate to predict stereotype change. In Chapter 7, I test alternative accounts for the dual process model, and argue that retrieval fluency effects automatically influence subsequent out-group judgments unless participants are engaged in bias correction. In Chapter 8, I present a final test of fluency and demonstrate that both impression formation and self-generation paradigms are influenced by different sources of meta-information and by the typicality of the exemplar. In Chapter 9, I present a summary of the evidence, highlight the main themes of the research, and identify implications for future investigations and social policy interventions.
CHAPTER 1

MEMBER-TO-GROUP GENERALISATION: ALTERNATIVE APPROACHES
TO A MULTI-DISCIPLINARY ISSUE

In November 2005, NSW state parliament spent time discussing the issue of driving and parking in Sydney. On this topic, independent senator Peter Breen commented that Asian drivers “are among the worst in the world… they park anywhere”. He later claimed that he was just expressing views “about the way a certain group of people park” (Fisher, 2005, p.13). How can one generalise all Asians into the same category of poor drivers and terrible parkers? And how many good Asian drivers will Mr Breen need to come across before he changes his stereotyped view? Generalisation from instances to general properties is probably the simplest and most pervasive form of everyday inductive reasoning (Nisbett, Kranz, Jespon, & Kunda, 1983). Generalisation is key to the way people understand and approach the world in which they live (e.g., McCauley, 1987). For this reason, generalisation is a principle area of investigation to understand human social cognition.

The first chapter of this thesis identifies a specific type of generalisation: member-to-group generalisation (Paolini, Hewstone, Rubin, & Pay, 2004b). Member-to-group generalisation refers to cases in which information about individual members of a social group affects the judgment of the group as a whole. Within this chapter, I clarify that member-to-group generalisation has been studied from a contact hypothesis approach or a social cognitive approach. Then I propose that the contact hypothesis approach and the social cognitive approach may be integrated within a single meta-cognitive approach that builds on the strengths of each research tradition. Throughout this work, meta-cognition is defined as the set of cognitions that apply to one’s mental activities, or simply people’s cognitions about cognition (Rucker & Petty, 2004). I conclude this first chapter by looking at ways of measuring member-to-group generalisation.

The second chapter reviews the social cognitive literature that has adopted an impression formation paradigm to understand the process of member-to-group generalisation. Within the impression formation paradigm, participants are presented with information about the qualities and the behaviours of individual group members, often but not necessarily stereotype disconfirming in nature. They are asked to form an impression of the specific out-group member (or members) and are then measured for their out-group judgments. Changes in group judgments (or simply generalisation) are assessed in terms of intra-individual
changes in group stereotyping before and after receiving the group members’ information or in terms of differences in group stereotyping between experimental and no-information control conditions (e.g. Bodenhausen & Lichtenstein, 1987; Hewstone, Johnston, & Arid, 1992; Johnstone & Hewstone, 1992; Kunda & Oleson, 1995; Rothbart & John, 1985; Weber & Crocker, 1983). By using a meta-analytical approach to this literature, Chapter 2 provides a parallel for the meta-analytic reviews in the contact literature (Pettigrew & Tropp, 2000, 2006; Tropp & Pettigrew, 2004, 2005) and extends previous non-qualitative reviews conducted within the social-cognitive research area (Hewstone, 1994; Paolini, 2001). I conclude by arguing for an increased focus on meta-cognitive processes to broaden our understanding of this specific type of generalisation.

Chapter 3 focuses on the limited number of meta-cognitive investigations carried out in the generalisation literature. Within this chapter, I meta-analytically review the studies that have investigated meta-cognition within an impression formation paradigm to highlight themes that emerge from this literature. A specific meta-cognitive process, fluency, is identified as being able to further our understanding of member-to-group generalisation. In this work, fluency is broadly defined as the ease associated with processing or retrieving information (Winkielman & Cacioppo, 2001). Within Chapter 3, I highlight two different types of fluency: processing fluency, which refers to the speed and accuracy associated with processing information, and retrieval fluency, which refers to the speed and accuracy associated with retrieving information. I conclude Chapter 3 by suggesting that the impression formation paradigm is a suboptimal approach for investigating the effects of fluency on member-to-group generalisation.

In Chapter 4, I propose a new theoretical and methodological framework to account for fluency in member-to-group generalisation. In particular, I introduce a dual process approach to model the themes that have emerged from previous fluency research and to make testable predictions for the empirical studies included in this thesis. To test these hypotheses, I propose a ‘self-generation’ paradigm. Within this paradigm, participants are asked to retrieve from their memory known exemplars of an out-group before being asked to judge the out-group.

Chapters 5–8 report five empirical tests of fluency effects within a member-to-group generalisation framework. Chapter 5 prepares and tests the self-generation paradigm in the context of younger people’s attitude towards the elderly. Chapter 6 contrasts the self-generation paradigm with the more traditional impression formation paradigm by asking participants to either retrieve counterstereotypical exemplars from their memory or read about counterstereotypical group members. Chapter 7 presents two empirical studies investigating the explanatory power of the dual process account for retrieval fluency effects. In particular, Study 3 investigates alternative explanations for retrieval fluency and Study 4 investigates the
impact of a meta-cognitive cue on retrieval fluency. Chapter 8 uses a novel method to test for both processing and retrieval fluency.

Chapter 9 summarises and documents the main findings presented in the thesis. Within this final chapter, I argue that meta-cognition, in particular retrieval fluency, influences the extent of member-to-group generalisation. I argue that the evidence from my research supports a dual-process model. I conclude by making suggestions for future research and for the development of social policy targeting reduced stereotyping and discrimination.

I will begin this investigation of meta-cognitive influences on member-to-group generalisation by broadly discussing inductive reasoning and tracing the development of our current understanding of member-to-group generalisation.

The Pervasive Nature of Inductive Reasoning

Inductive reasoning involves forming general principles from a series of examples (Kruglanski & Thomson, 1999; Wagner & Parker, 1993). For many centuries, philosophers such as Aristotle and Hume have examined the properties and problems associated with inductive reasoning. Aristotle argued that the primary source of human beliefs come from inductive reasoning and Hume (1748/1955) argued that humans use inductive reasoning despite the inaccuracies associated with relying on previous experience to predict future events. Inductive reasoning has been identified as a key mental ability (Thurstone, 1938), and forms a part of many cognitive functioning theories (e.g., Cattell, 1971, Horn, 1994; Snow, 1980). Induction tasks have been of considerable interest to psychometricians (e.g. Carroll, 1993; Horn, 1994), information-processing psychologists (see, e.g., Sternberg, 1977; Whitely & Barnes, 1979), developmental theorists (Piaget, 1928; 1950) and learning theorists (Sutherland & Mackintosh, 1971; Trabasso, Bower, & Gelman, 1968). Inductive reasoning is essential for superior performance on IQ tests (e.g., WISC-IV-TR, WAIS), and in everyday situations (for a similar point, see Sternberg & Kalmar, 1998).

Within cognitive psychology, research on inductive reasoning was apparent as early as Lashley’s (1951) research on the problem of serial order in behaviour. By the 1960’s, inductive reasoning tests such as Milner’s (1964) multi-card task had become common. Inductive reasoning tests assess the extent to which individuals rely on this process and assume that superior performance requires individuals to identify underlying principles by formulating and testing hypothesis. Typically, cognitive models of inductive reasoning

---

1 In this classic inductive reasoning test participants are shown a series of cards depicting pictures of different colours, shapes and frequency. Participants are then provided with two additional cards and are required to select one of the two cards that match the original series. More than one trial is typically required as participants need to identify whether the concept involves colours, shapes, numbers, or some combination of the card dimensions. Information gained from subsequent trials allows the participant to rule out different hypothesis – for example colours might not matter, but shape and numbers do.
specify information processing components as well as cognitive strategies (e.g., Dillon & Wisher, 1981; Mulholland, Pellegrino, & Glaser, 1980; Restle, 1970; Simon, 1972; Sternberg, 1977; Whitely, 1977).

Within social psychology, social inductive reasoning is quite complex. Unlike cognitive inductive reasoning tasks that use simple objects, such as shapes, the investigation of social induction requires the consideration of more complex concepts and processes, such as exemplar information and self-relevance motives (e.g., see, Fielder & Walther, 2004; Medin & Schaffer, 1978, Posner & Keele, 1968; Smith & Medin, 1981). In addition, in social situations, it is typically not possible to gather all of the available and relevant information making social inductive reasoning more difficult than cognitive inductive reasoning (Fielder & Walther, 2004; Krueger & Clement, 1996).

Judgment construction of social groups involves a special type of social inductive reasoning called member-to-group generalisation. In the following section, I will trace the history of research surrounding member-to-group generalisation and consider the models adopted to represent this type of social inductive reasoning.

**Member-to-Group Generalisation**

Negative stereotypes about social groups (defined by race, gender, religion, occupation or other covert characteristics) can lead to prejudiced feelings and discrimination, ranging from mild social sanctions through to ethnic cleansing and genocide (Queller & Smith, 2002). Member-to-group generalisation explains, at least partly, the formation (Sherman, 1996), as well as the change (Rothbart, 1996), of people’s stereotypes. Understanding generalisation in the social domain has the potential to reveal the causative cognitive processes involved in utilising favourable or unfavourable out-group exemplar information when making decisions about out-groups and their members. From this research, social policy makers are better placed to reduce intergroup conflict and prejudice through relevant interventions.

**Traditional Approaches to Member-to-Group Generalisation**

Two separate research traditions have studied the process of member-to-group generalisation in social psychology: (1) research on intergroup contact and (2) research from the social cognitive approach. Intergroup contact research focuses on contact between members of different groups as a way of reducing out-group stereotyping and prejudice. From this perspective the basis of generalisation is “actual face-to-face interaction between members of different groups” (Tropp & Pettigrew, 2004, p. 253, emphasis partly removed). The social cognitive approach, on the other hand, focuses on the processing of written, audio or visual information about out-group members for its effects on group judgments. These two
research traditions are complementary and serve qualitatively different purposes. The intergroup contact approach looks at the application of generalisation and the social cognitive approach focuses on understanding the process of generalisation.

**Intergroup Contact Research**

Early work on the contact hypothesis was dominated by the assumption that establishing affective ties with specific out-group members would promote positive feelings toward the out-group as a whole (Tropp & Pettigrew, 2004). Allport (1954) formally advocated the idea that intergroup contact is an important means of reducing stereotyping and prejudice (for earlier instances, see Watson, 1947; Williams, 1947). Since Allport, the contact hypothesis has been extensively tested across a variety of situations, groups, and societies (for an early review, see Amir, 1976; for a recent review, see Pettigrew & Tropp, 2006).

Allport (1954) stressed that for contact to be successful in reducing negative attitudes the contact situation needs to include elements such as cooperative interdependence, whereby the individuals engaged in the contact situation depend on each others’ actions in order to enjoy positive outcomes (Gaertner, et al., 1999; Sherif, Harvey, White, Hood, & Sherif, 1961). Successful contact also requires individuals to be of equal status and of sufficient duration to enable friendships to develop (Cook, 1978). Allport also argued that there needed to be institutional support for intergroup contact. Through these elements, contact experiences were expected to reveal similarities amongst the contact partner and undermine hostile intergroup attitudes (Bramel, 2004).

Early commentaries on the contact hypothesis drew contradicting conclusions. Some indicated that contact leads to reduced prejudice (e.g. Jackson, 1993; Pettigrew, 1971; 1998; Riordan, 1978), while others suggested that contact has little, if any, impact on intergroup relations (e.g. Amir, 1976; Ford, 1986; for recent evidence, see Forbes, 2004). In attempting to resolve these conflicting conclusions, Pettigrew and Tropp (2000, 2006; see also, Tropp & Pettigrew, 2004, 2005) recently conducted an extensive meta-analysis. Their meta-analyses incorporated 713 independent samples from 515 studies of intergroup contact, beginning with research conducted in the 1940s and continuing until 2000. Pettigrew and Tropp found that contact experiences typically affect attitudes toward the specific out-group members but also generalise to the entire out-group. These effects occur reliably across a broad range of out-group targets and contact settings.

Importantly, Tropp and Pettigrew (2005) argued that the different conclusions about intergroup contact in early research were partly due to an inability to distinguish prejudice measures between cognitive and affective lines. Their work demonstrates that contact research incorporating cognitive measures (e.g., trait rating scales) have smaller effect sizes than research using affective measures (e.g., feeling thermometers). This issue of including
different measures of generalisation and treating them as distinct will be further considered later in this chapter.

**Methodology.** The vastly dominant design used to test the contact hypothesis is cross-sectional correlational designs (Binder et al., 2009). An influential example of correlation research is the work by Islam and Hewstone (1993). Islam and Hewstone investigated prejudice between Hindu and Muslim students in Bangladesh and found that prejudice was lowest amongst the students who had the greatest amount of contact with the out-group. Correlation research has the merit of drawing on personal experiences but is limited in its ability to pinpoint causation (Binder et al., 2009; MacCullum & Austin, 2000). So, instead of contact reducing prejudice, it is possible that the opposite is occurring (i.e., reduced prejudice encourages contact; Pettigrew, 1998).

One way to address the problem of causation is through the use of a control group. For example, Maras and Brown (1996) arranged for a group of 9-year old school children to participate in an exchange program with a special needs school over a period of three months. Their research included a randomly selected control group of children who did not participate in the exchange. At the end of the experience, the exchange children showed more understanding of disabilities and held less homogeneous views of children with disabilities than the control children. The problem with these types of studies is that the control group, even when randomly selected, may still have had some degree of contact with the out-group (e.g., outside of regular school hours; Pettigrew & Tropp, 2000).

The rare experimental-based contact studies attempt to by-pass the problems of ‘causation’ and ‘residual contact’ by fully controlling for it (e.g. Brown, Vivian, & Hewstone, 1999; Cook, 1978; Desforges, et al., 1991; Wolsko, Park, Judd, & Bachelor, 2003). For example, Wolsko et al. studied generalisation by having White American students work on a task with a single out-group member (Latino confederate) of a stereotyped group (pleasant-but-lazy Latinos). In this experiment, participants were asked to work together as a group to complete a task and a group leader was selected by drawing a name out of a hat. This part of the process, however, was rigged so that the Latino confederate was always selected as the leader. The confederate acted in a stereotypical (pleasant and lazy) or counterstereotypical (pleasant and highly motivated) manner. Improved attitudes towards the confederate were generalised to the out-group when the individual was perceived as disconfirming the stereotype in the pleasant and highly motivated condition, but not in the stereotypical condition. Although by-passing the problems associated with causality and residual contact, these types of experimental contact studies often do not go as far as investigating the process that lead to generalisation.

**Summary.** The intergroup contact approach has proved useful for investigating intergroup relations (Pettigrew, 1986; 1998) and for encouraging stereotype reduction via
interpersonal experiences. However, research in this area does not identify the exact processes that lead to generalisation, although some advances in this direction have been recently made (e.g., Pettigrew & Tropp, 2006; Turner, Hewstone, & Voci, 2007). For more exact tests on the cognitive processes contributing to generalisation, I turn to the social cognitive approach.

**Social-Cognitive Approach**

Rothbart and John (1985) tackled the problem of weak and unpredictable generalisation in contact situations by bringing member-to-group generalisation into the framework of the categorisation process (Allport, 1954; Tajfel, 1981). After Rothbart and John’s initial theorisation, social cognitive researchers focused on identifying a number of psychological *mechanisms* responsible for stereotype formation and change, hence, explaining why sometimes generalisation occurs and why sometimes it does not (Hewstone, 1996).

Of particular relevance to the current thesis is Rothbart and John’s (1985) consideration of the role of exemplar typicality. Rothbart and John argued that when an individual is considered ‘typical’ of the group, generalisation to the out-group is increased, whereas when the out-group member is perceived as atypical, generalisation is reduced (Rothbart & Lewis, 1988; Wilder, Simon, & Faith, 1996). This is because typicality signals the cognitive connection of the exemplar to the group, thus resulting in enhanced generalisation for typical group member/s (or simply a typicality effect). Atypical group members, on the other hand, are unlikely to activate the group stereotype, thus leaving their attributes disassociated from the out-group representation (Johnston & Hewstone, 1992; Rothbart & John, 1985; Rothbart & Lewis, 1988). Due to the typicality effect, people generalise disconfirming attributes more from relatively typical exemplars than atypical exemplars.

Social cognitive researchers have proposed a number of different models of generalisation. Weber and Crocker (1983) identified three models: bookkeeping, conversion, and subtype (or prototype). The *bookkeeping model* (Rothbart, 1981) suggests that reduction in stereotyping is an incremental process in which each instance of stereotype relevant information is used to modify the existing stereotype. Any single piece of disconfirming evidence would be generalised to the group as a whole and elicit a minor change in stereotyping. Early commentaries argued that this model failed to clarify the exact instances that lead to the development of the group representation (Johnston & Hewstone, 1992). Instances could be behaviours or attributes of an individual group member or of all the group members. Nonetheless, the model predicts larger generalisation with a larger sample of disconfirmers (Hewstone, 1994).
The conversion model (Rothbart, 1981) portrays changes in stereotyping as a
dramatic process (see also, Henderson-King & Nisbett, 1996). Stereotyped beliefs are
expected to change drastically in response to extreme instances and remain largely unchanged
in the face of minor disconfirmation (Bodenhausen, Schwarz, Bless, & Wänke, 1995;
Hewstone, Hopkins, & Routh, 1992; Johnston & Hewstone, 1992). This is an intuitive model
that predicts extreme atypicality to have a stronger impact on the perception of the group than
moderate atypicality (Brown & Hewstone, 2005). The subtype model (or the prototype model)
predicts an advantage in generalisation of moderately atypical exemplars over extremely
atypical exemplars (Hewstone, 1994, 1996; Hewstone, Macrae, Griffiths, Milne, & Brown,
1994; Johnston & Hewstone, 1992; see also Huici, Ros, Carmona, Cano, & Morales, 1996;
Kunda & Oleson, 1997). Inspired by Rothbart and John’s (1985) model of social
categorization, this model suggests that moderately atypical individuals are more likely to
cause a re-evaluation of the stereotype than extremely atypical exemplars. This is because
extremely atypical individuals are so different from the group prototype that they are
excluded (or subtyped) from the stereotyped group.

These classic models made predictions about the out-group sample size and/or
exemplar typicality that govern generalisation. More recently, Bodenhausen et al. (1995)
developed the generalised appraisal model to incorporate contextually sensitive judgments,
and Schwarz and Bless (1992) advanced the cognitive-inclusion model to account for the
direction of generalisation. Broadly, the classic models assume that information about a
group’s central tendency and variability are stored in memory and are ‘permanently’ effected
by exemplar information. Whereas modern exemplar-based models (e.g., generalised
appraisal and inclusion-exclusion) assume that group representations are not permanently
stored in memory, but rather created on-line when a judgment is warranted (Garcia-Marques,
Santos, & Mackie, 2006). As a result these modern generalisation models are more able to
account for contextual variations.

In particular, the generalised appraisal model (Bodenhausen et al., 1995) starts from
the premise that out-group attitudes reflect a response to specific exemplars that are
cognitively accessible at the time of the judgment (e.g., Kahneman & Miller, 1986; Medin,
Altom, & Murphy, 1984; Smith, 1990, 1991, 1992; Smith & Zarate, 1992). The key to
Bodenhausen et al.’s (1995) model is not the typicality of the exemplar per se, but rather the
positivity of the exemplar. That is, irrespective of other exemplar characteristics, if the
exemplar is not well liked, then there will be no improvement of the out-group attitudes or
beliefs (p. 52). Bodenhausen et al. also argued that diverse out-groups that encompass a
number of different subgroups and a complex array of exemplars (i.e., internally diverse
groups) are unlikely to produce large generalisation simply because the individual members
share few common traits. In contrast, group schemas that are less internally complex should
be more sensitive to context cues and specific information and, thus, demonstrate larger generalisation effects. That is, generalisation should be largest when the group is small, homogenous and when the exemplar immediately accessible is perceived positively.

From a similar on-line view of stereotyping judgments, Schwarz and Bless (1992) suggested with their inclusion-exclusion model that the same out-group exemplar could be included or excluded from the group representation (Bless & Schwarz, 1998; for parallel hypotheses within person perception, see Stapel & Koomen, 1997, 1998). They argued that generalisation (or cognitive inclusion of the exemplar information in the group judgment) is the default process. So, under normal situations, an out-group exemplar should be included in the cognitive representation of the group. Certain contextual factors, however, may trigger exclusion of the exemplar material, resulting in a contrast effect (for a parallel discussion of contrast effects, see Martin, 1986). Hence, this model suggests that participants engage in bias correction when they suspect that the sample information had an unjustified influence on their view of the group as a whole.

Comparative tests of the bookkeeping, conversion and prototype models have typically relied on the concentrated-dispersed paradigm (for early examples of this paradigm, see Gurwitz & Dodge, 1977; Weber & Crocker, 1983). In this paradigm, all participants are presented with the same amount of exemplar information, however, the disconfirming information is either ‘concentrated’ in a few group members or ‘dispersed’ across many group members. For example, Weber and Crocker presented participants with individual behaviours performed by librarians or lawyers. One third of the presented behaviours were counterstereotypical, the rest of the information was either stereotypical or irrelevant. The total amount of counterstereotypical information was kept constant across conditions, but its pattern was varied systematically. In the dispersed condition, the counterstereotypical behaviours were distributed evenly across members, with each exemplar displaying the same number of counterstereotypical behaviours. In the concentrated condition, the counterstereotypical behaviours were clustered within fewer individuals. That is, in the dispersed condition a greater number of exemplars are presented as moderately atypical compared with the concentrated condition where a smaller number of exemplars are presented as extremely atypical. Participants then completed group measures of stereotyping. Weber and Croker found that disconfirming evidence changed the out-group stereotypes more when it was dispersed than when it was concentrated. They interpreted this evidence as consistent with the subtype model. More recently, Hewstone (1994) conducted a narrative review of the concentrated-dispersed literature and concluded that the subtype model was more strongly supported by this research than the conversion and bookkeeping models.

The concentrated-dispersed paradigm is problematic as a ‘single’ test for the generalisation process because it varies exemplar typicality and sample size of disconfirming
exemplars. As these two variables are both potentially relevant to the generalisation process (Hewstone, 1994; Hewstone et al., 1994; Kunda & Oleson, 1997; Paolini, 2001), in the following chapter I will review the influence of exemplar typicality and sample size on the generalisation process by focusing on research that does not adopt the concentrated-dispersed paradigm, thus providing a neater test of the different models of stereotype change.

Methodology. It is evident by now that social cognitive researchers investigate the way (often disconfirming) information regarding individual group members is processed, interpreted, stored, and utilised when building up group evaluations and judgements (Hewstone, 1989, 1994; Rothbart & John, 1985). Researchers in this tradition have typically adopted an impression formation paradigm (Garcia-Marques & Mackie, 1999).

Many impression formation studies investigate ‘natural’ groups, despite the inherent problem of variability in previous experiences with the out-group (e.g., Bodenhausen et al., 1995; Vescio, Sechrist, & Paolucci, 2003; Virj, Akehurst, & Smith, 2003). Drawing from Pettigrew and Tropp (2006) and Paolini, Hewstone, and Cairns (2007) I will argue that social groups can be broadly divided into two different categories: (1) affective groups (e.g., race, gender, ethnicity), where the representation is dominated by affect-laden information, and (2) cognitive groups (e.g., occupation or student academic majors), where the representation is dominated by beliefs and knowledge. Vescio et al. (2003) provide an example of generalisation research with an affective social group. These researchers asked participants to listen to an interview with an African American who was either described as stereotypical (control) or counterstereotypical. Participants generalised the counterstereotypical information by demonstrating less stereotypical group judgments than the control participants. Garcia-Marques and Mackie (1999, Study 1) provide an example of impression formation research using a cognitive social group. In their research participants were given information regarding eight members of one occupational group (computer programmers, construction workers, child-care professionals, or disco-bouncers) who were either stereotypical or counterstereotypical to the occupational stereotype. All participants were then asked to judge the target group. A pre-post treatment measure indicated that participants exposed to the counterstereotypical information generalised this information to the out-group by becoming less stereotypical in their judgments; similarly the participants exposed to the stereotypical exemplar information generalised this information to the out-group by becoming more stereotypical in their judgements. This evidence demonstrates that generalisation may occur in the direction of reduced as well as increased stereotyping depending on the exemplar information cognitively accessible at the time of the judgment.

In the study of natural groups, experimenters have very limited control over previous contact and previous knowledge the respondents may have about the natural group. Precautions against the confounding influence of these variables are difficult to implement.
and, consequently, are not usually taken (see e.g., Denhaerinck, Leyens, & Yzerbyt, 1989; Moreno & Bodenhausen, 1999). In an effort to avoid the problem of prior contact, some researchers adopted a minimal group paradigm (Tajfel, Flament, Billig, & Bundy, 1971). A minimal group is one that is artificially created for the purpose of the research. For example, research by Paolini, Crisp, and McIntyre (2009, Study 3) informed participants about an outgroup of ‘over-estimators’ of physical stimuli. This group was described as having certain characteristics and traits, but were typically known for their tendency to overestimate the presence of physical stimuli. Participants were then informed about the qualities of a counterstereotypical over-estimator and asked to indicate their group impressions. The advantage of this artificial group is that it allows for strict control of participants’ previous experience with the group under investigation, ensuring that any effects are due to the experimental manipulations (for an overview of artificial group studies, see Cadinu & Rothbart, 1996). However, as noted later in this thesis, generalisation from a counterstereotypical ‘over-estimator’ may not occur at the same rate as more realistic social groups, such as gender or religious groups.

Summary. Impression formation studies have the advantage of stricter control of the out-group information to which the participants are exposed (Garcia-Marques & Mackie, 1999), allowing for stringent empirical assessments of generalisation of sample information (e.g., Ames, 2004; Henderson-King & Nisbett, 1996; Rothbart & Lewis, 1998; Weber & Crocker, 1983). Using the range of impression formation paradigms, social cognitive researchers have identified a number of variables that moderate the process, such as exemplar typicality and disconfirming sample size. However, impression formation studies have also demonstrated that, even when people are presented with disconfirming information, stereotypes often remain unchanged (e.g., Banks, 1995; Huici et al., 1996; Judd & Park, 1993). Together, these studies demonstrate that generalisation occurs across a variety of outgroups. The integration of these diverse findings has resulted in rather sophisticated models about the dynamics of member-to-group generalisation. In the following chapter, I will look at the supportive data for each of these models.

A New Approach to Member-to-Group Generalisation

Stereotyping, prejudice and discrimination constitute major social problems that have been tackled through the contact and social cognitive approaches. In particular, contact researchers have articulated and tested social conditions that facilitate generalisation, while the social cognitive researchers have worked towards understanding the cognitive underpinnings of generalisation and variables that moderate generalisation. Despite the successes of these two approaches, stereotypes remain dominant in all societies and social groups remain in conflict. Throughout this thesis, I will argue that an investigation of meta-
cognition will advance our understanding of the process of member-to-group generalisation and improve the predictions and efficiency of social interventions.

Recently, social psychologists have become increasingly aware of the potential role played by meta-cognitive processes in influencing behaviour and judgments (for a review, see Jost, Kruglanski, & Nelson, 1998; see also, Metcalfe & Shimamura, 1994; Nelson, 1992). Little work, however, has specifically considered the impact of meta-cognition on the process of member-to-group generalisation (Paolini et al., 2009). Later in my work, I will demonstrate that there are a number of studies that have indirectly researched the effects of meta-cognition on member-to-group generalisation (e.g., Batson et al., 1997; Yzerbyt, Coull, & Rocher, 1999) and I will use a meta-analytical approach to synthesise its contribution. For now it suffices to say that this seminal meta-cognitive research indicates that generalisation might be shaped and stereotypes might be reduced by making individuals aware of their own thought processes. A better understanding of the effects of meta-cognition on the process of member-to-group generalisation might inform social policies that counter negative attitudes.

From a methodological point of view, the key difference between the contact and social cognitive methodologies is the interaction (intergroup contact) versus non-interaction (social cognitive) between the participant and the out-group members. One advantage of the intergroup contact approach is that it draws on real experiences with out-group members. One advantage of the social cognitive approach is the strict control inherent in the experimental situation, which allows for a clear understanding of the processes involved in member-to-group generalisation. I formalise a new generalisation paradigm that draws on both of these strengths (i.e., personal experience and experimental control) by asking participants to retrieve specific counterstereotypical out-group members from their memory. The retrieval instructions will be explicit, thus ensuring maximum control over the contents immediately accessible for generalisation while still allowing participants to rely on their own experiences with the out-group. Chapter 5 will pilot and test the self-generation paradigm that will be adopted throughout the experimental studies of this thesis.

Measuring Generalisation

Member-to-group generalisation is traditionally assessed along measures of group stereotypicality, group dispersion, and prejudice towards the group (Paolini et al., 2004b). As noted earlier, generalisation across the different outcome measures is assessed comparing a no-information group (typically a control group or a pre-information baseline) with an information group (typically an information group or a post-information group).
Chapter One

Classic Measures

Traditionally, *group stereotypicality* dominated the research on the generalisation process (Hewstone, 1996). Group stereotypicality taps the act of assigning a set of stereotyped attributes to individual members of a target group or to the group as a whole (Ashmore & Del Boca, 1981; Hewstone & Brown, 1986; Guinote, 2001). As stereotypes operate to simplify the perception of complex social situations (Macrae, Milne, & Bodenhausen, 1994; Stangor & Duan, 1991), they inevitably lead to reduced attention to stereotype inconsistent and individuating information (for reviews, see Fiske, 1998; Hamilton & Sherman, 1994; Hilton & von Hippel, 1996) and to limited openness to change.

Group stereotypicality was first measured by Katz and Braly’s (1933) trait listing task. In their original investigation, Princeton undergraduates were asked to ascribe a list of 84 trait adjectives to ten different ethnic groups. Since Katz and Braly, many variations of this measure have emerged (for an overview, see Park & Judd, 1990). For example, percentage estimate tasks require participants to estimate the prevalence of stereotypical attributes among out-group members (see e.g., Nisbett et al., 1983, Study 1; Wyer, Sadler, & Judd, 2002, Study 1). Trait rating tasks are another example; here participants indicate where an average group member, or the group in general, falls on rating scales anchored by stereotype relevant attributes (see e.g., Paolini et al., 2004b, Weber & Crocker, 1983, Study 1).

*Group dispersion* refers to the extent to which group members are perceived as being dispersed around their central tendency (Park & Judd, 1990). Dispersion is low when group members are perceived as being similar on an attribute. Under these conditions, group stereotypes are more confidently applied to (new) group members (Ryan, Judd, & Park, 1996) and group members are treated similarly (Guinote, 2001). When there are large discrepancies perceived amongst group members, dispersion is high, and stereotyping is generally lowered. Many group dispersion measures exist, such as general similarity tasks and range tasks (for an overview, see Park & Judd, 1990; see also, Maurer, Park, & Rothbart, 1995). General similarity tasks ask participants to rate how similar vs. different members are from each other (e.g. Maurer et al., 1995; Stroessner, Mackie, & Michalsen, 2005; Wyer et al., 2002, Study 1). Range tasks ask participants to estimate where the two most extreme group members fall on rating scales (e.g., Hewstone & Hamberger, 2000, Study 2; Paolini et al., 2004b, Study 2); the distance between the two points (range) provides a measure of dispersion.

*Prejudice* describes one’s overall evaluation of a group and encompasses how positively or negatively the out-group is perceived (Binder et al., 2009). Prejudice has been measured in a variety of ways, such as asking participants to express valenced judgments and valenced beliefs about a group (Bodenhausen, et al., 1995; Hamill, Wilson, & Nisbett, 1980) or complete feeling thermometer rating tasks (Brauer, Judd, & Jacquelin, 2001). Prejudice has also been measured through social distance scales where participants are asked to indicate
their willingness to interact with out-group members at various degrees of interpersonal closeness (Desforges, Lord, Pugh, Sia, Scarberry, & Ratcliff, 1997). Despite the visible differences amongst these measures, all of these assess how favourably or unfavourably the participants view the out-group.

Early research neglected key differences between these classic outcome measures, assuming that generalisation on any one measure was driven by a common process. More recently, however, researchers have become increasingly interested in assessing differences in measures (e.g. Maurer et al., 1995; Park, Wolsko, & Judd, 2001; Tropp & Pettigrew, 2004) and generalisation differences between each of these measures (e.g., Garcia-Marques & Mackie, 1999; Paolini, et al., 2004b). In the following chapter, I will review the literature that has assessed generalisation using the classic outcome measures. This analysis will enable me to address the moderating role of each of the outcome variables, thus identifying if each of the measures are equally sensitive to member-to-group generalisation. I will now introduce a new approach to assessing generalisation.

**Implicit Group Measures**

Explicit measures such as those described above are capable of measuring generalisation, but are subject to a range of social desirability considerations and demand characteristics as people are able to control their responses to explicit questioning (Devine & Monteith, 1999; Fazio & Dunton, 1997; Fiske, 1989; Plant & Devine, 1998). Implicit measures such as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; Greenwald & Banaji, 1995), the Go/ No Go Association Task (GNAT; Nosek & Banaji, 2001), completion of word fragments (Hense, Penner & Nelson, 1995, Sinclair & Kunda, 1999), assessments of abstract versus concrete language (von Hippel, Sekaquaptewa, & Vargas, 1995; 1997) are only minimally affected by efforts at providing socially desirable responses (for a review of implicit measures, see Fazio & Olson, 2003). This is because all of these implicit measures encourage a focus on response speed and/or accuracy, thus they limit conscious response control, providing an estimate of the measured construct without asking for a verbal report from the participant (Fazio & Olson, 2003).

For example, the IAT assesses automatic stereotyped associations by presenting pictures or words depicting an attitudinal object (e.g., pictures of overweight people) and pictures or words depicting an evaluative state (e.g., words such as ‘pleasant’, ‘unpleasant’) or a stereotypic state (e.g., ‘lazy’, ‘busy’). Participants are required to allocate quickly and accurately the pictures and words to a left or right hand category. An implicit stereotype is shown when participants respond more quickly to stereotype-consistent pairs of words than to stereotype-inconsistent pairs of words. For example, a person with a negative attitude towards
overweight people will be faster to pair images of overweight people with ‘unpleasant’ words than people with a positive attitude (e.g., Greenwald, Nosek, & Banaji, 2003).

Researchers from both the contact and social cognitive sides of the generalisation research have started including implicit measures to further understand the process of member-to-group generalisation (for the value in this approach, see Wilson, Lindsey, & Schooler, 2000). Contact researchers have found that prior contact with the out-group influences attitudes at an implicit level. For example, Tam, Hewstone, Harwood, Voci, and Kenworthy (2006) measured the quality and quantity of contact that young adults had with elderly people and then assessed out-group attitudes. Using the IAT, Tan et al. found that contact quantity (but not quality) had a positive effect on implicit associations, while contact quality had a positive effect on explicit ratings. Research adopting an impression formation approach has also found evidence of member-to-group generalisation on implicit measures. For example, Blair, Ma, and Lenton (2001) found that asking participants to visualise a strong (countert stereotypic) woman resulted in reduced implicit stereotyping using a gender-strength IAT when compared with control participants who visualised a stereotype-irrelevant scenario. The research reported in this thesis will always include the three classic measures (stereotypicality, dispersion and prejudice) and Studies 1 and 4 will also include the IAT as an implicit outcome measure. This multi-measure approach will make my research ideally positioned to identify possible dissociations between different features of group responding.

Conclusions

Member-to-group generalisation is a special type of inductive reasoning that has traditionally been researched through an intergroup contact or social-cognitive approach. In this chapter, I suggested that integrating the contact and social cognitive approaches through a meta-cognitive framework will extend and improve our understanding of the process of member-to-group generalisation. Hence, in the following chapter, I review the bulk of the literature that has adopted a social cognitive approach. In Chapter 3, I review the member-to-group generalisation literature that has included meta-cognitive manipulations. In Chapter 4, I present a theoretical and methodological framework that integrates meta-cognition and member-to-group generalisation research. In Chapter 5, I detail a new paradigm to test meta-cognitive member-to-group generalisation. This paradigm is then used across five empirical studies to test the role of meta-cognition in member-to-group generalisation.
CHAPTER 2

GENERALISATION MEASURES AND OTHER KEY MODERATORS:
META-ANALYTIC REVIEW (1)

If it looks like a duck, walks like a duck, and quacks like a duck – it’s probably a duck! Generalisation from instances (such as walking and quacking) to groups (such as ducks) is a form of inductive reasoning, which involves moving from specific observations to broader generalisations. Research in inductive reasoning has led to a better understanding of the processes involved in group stereotyping (McCrae, 1987; Tversky & Kahneman, 1971). A special type of inductive reasoning, introduced in Chapter 1, is the process of member-to-group generalisation. This refers to cases in which information about individual members of a social group affects the judgment of the group as a whole. Member-to-group generalisation not only allows for stereotype formation (Sherman, 1996), but also for stereotype revision (Rothbart, 1996).

In Chapter 1, I introduced two main approaches to member-to-group generalisation (also see Crocker, Fiske, & Taylor, 1984). One approach is through increased experience of out-group members in the form of intergroup contact. Intergroup contact has been thoroughly reviewed by Pettigrew and Tropp (2000, 2006, 2008; Tropp & Pettigrew, 2004, 2005). The other approach is social cognitive and investigates generalisation through exposure to (disconfirming) exemplar information (see Crocker et al., 1984). It is this social cognitive approach that I have reviewed herein. This review complements the work by Pettigrew and Tropp and together provides a comprehensive review of the stereotype change literature.

Social cognitive research on the process of member-to-group generalisation has typically adopted an impression formation paradigm (e.g. Bodenhausen & Lichtenstein, 1987; Garcia-Marques & Mackie, 1999; Johnston & Hewstone, 1992; Kunda & Oleson, 1995; Rothbart & John, 1985; Weber & Crocker, 1983; for a narrative review, see Paolini, 2001). This paradigm presents participants with information about out-group members and then measures for group perceptions (e.g., stereotypicality, group dispersion or prejudice). Generalisation is typically assessed by comparing a no-information control group with the experimental information group (e.g., Ames, 2004), or more rarely, through the use of pre and post information designs (e.g., Cameron & Rutland, 2006). In this chapter, I meta-analytically review the impression formation literature, with a focus on identifying if, and under what conditions, out-group sample information is generalised to the out-group as a whole.
Past Reviews of Member-to-Group Generalisation


Pettigrew and Tropp’s Reviews of the Contact Literature

Pettigrew and Tropp (2000, 2006) reviewed 515 intergroup contact studies involving over 250,000 participants (see also Tropp & Pettigrew, 2004, 2005). Research was included in the meta-analysis only if it involved “actual face-to-face interaction between members of clearly defined groups” (Pettigrew & Tropp, 2006, p. 754). Overall, Pettigrew and Tropp found that intergroup contact generalises to the out-group through a significant (though small) reduction in out-group prejudice (mean $r = -0.23$). In the analysis they included a variety of moderators to assess Allport’s (1954) optimal conditions to maximise generalisation (see also Pettigrew & Tropp, 2008). The assessment of moderators included coding for the generic vs. intimate nature of the contact interaction. Pettigrew and Tropp found that face-to-face contact was negatively related to prejudice (Pettigrew & Tropp, 2000), as was friendship-based contact experiences (Pettigrew & Tropp, 2000).

As noted in the previous chapter, Tropp and Pettigrew (2005) drew a distinction between outcome measures, defining cognitive outcome measures as one’s beliefs and perceptions about a group and defining affective outcome measures as one’s feelings and emotional responses to a group (Tropp & Pettigrew, 2005, p. 1147). Cognitive measures require people to provide a relatively detached evaluation of the group; on the other hand, affective measures require people to focus on their feelings toward the out-group members in the context of their relationships and experiences. Tropp and Pettigrew found a relative advantage for affective measures (such as prejudice) over cognitive measures (such as stereotyping), although both measures carried a significant reduction in negative out-group responses (Tropp & Pettigrew, 2005). Given that affective ties with out-group members are established through close contact, it is likely that such contact would be associated with greater attitudinal shifts on affective dimensions of prejudice than on cognitive dimensions. Tropp and Pettigrew’s approach to outcome measures is limited by their neglect for measures of dispersion, possibly reflecting little consideration of this outcome measure within the contact literature.

Hewstone’s Selective Review of the Social Cognitive Research

Hewstone (1994) reviewed impression formation paradigm research that focused on the generalisation of disconfirming information in the concentrated/dispersed paradigm. He
considered the three cognitive models of stereotype change, bookkeeping, conversion and subtype, and identified evidence supporting each of these models. Within the concentrated/dispersed literature, Hewstone identified that most generalisation occurred in the dispersed conditions, and used this evidence to argue that there was limited support for the conversion model and maximum support for the subtype model. Hewstone concluded that generalisation is maximised when exemplars are moderately atypical rather than extremely atypical, at least within concentrated-dispersed paradigms. However, as noted above, the concentrated/dispersed paradigm is limited as a thorough test of generalisation because it varies exemplar typicality and sample size.

By focusing exclusively on research adopting the concentrated/dispersed paradigm, Hewstone’s review offers only a selective view of the social-cognitive literature on member-to-group generalisation. The present review incorporates work from a wider number of impression formation studies, providing an updated and wider perspective on the social cognitive approach to generalisation.

**Paolini’s Narrative Review of the Social-Cognitive Research**

In her dissertation work, Paolini (2001, pp. 20-51) reviewed over two decades of social cognitive research on member-to-group generalisation. This review focused on a multitude of moderators of the generalisation process, such as member’s deviance (Garcia-Marques & Mackie, 1999, Study 1), member’s representativeness (Desforges et al., 1997), attributional cues (Garcia-Marques & Mackie, 1999, Study 3; Kunda & Oleson, 1995), strength of the stereotype (Kunda & Oleson, 1997), cognitive busyness (Yzerbyt et al., 1999), and direction of deviance (Garcia-Marques & Mackie, 1999, Study 2). Through this list of moderators, Paolini concluded that impression formation studies result in generalisation, but only under specific circumstances. Consistent with Hewstone (1994), she concluded that extremely atypical exemplars “rarely provide the best vehicle for stereotype disconfirmation” (Paolini, 2001, p. 50).

Paolini was able to identify and draw patterns between the various moderators. However, by not engaging in a quantitative review, the end result was a lengthy list of factors contributing or detracting from the generalisation process without systematic quantification of the generalisation process. The goal of this analysis was to update Paolini’s review, simplify the list of moderators, and quantify the size of generalisation effects and their moderators.

**The Present Review**

The previous reviews of the social cognitive approach to generalisation have used traditional qualitative approaches. Although several trends can be seen with narrative reviews (e.g. the benefit of dispersed information over concentrated information), there is still a
subjective judgment based on the researchers reading of the literature which limits definite conclusions about it. A meta-analysis offers several advantages over traditional narrative reviews. Firstly, it is very precise, as each study is represented as an effect size (Mullen, Brown, & Smith, 1992). Secondly, it requires a very clear set of inclusion criteria (Cortina, 2003). Thirdly, subtle differences can be identified between moderating variables that may not emerge in narrative reviews (Glass, 1977).

In this chapter, I have organised the literature to address two propositions (for a similar approach, see Paolini, 2001). The propositions address the following logically related questions: (1) Do people generalise sample information to the out-group in general? (2) What conditions, or factors, maximise the generalisation of information? The first proposition addresses the basic and simplest underlying assumption in the literature that out-group information is generalised to the out-group during impression formation research. The second proposition focuses on the different moderating variables that maximise generalisation.

**Proposition One: Testing Basic Generalisation**

The first proposition aims to address the question: Do people generalise information from out-group members to the out-group in general? Social cognitive researchers, supported by cognitive research on inductive reasoning (e.g., Chen, Mo, & Honomichl, 2004; Holyoak & Thagard, 1997; Ross & Kennedy, 1990), have assumed that people do generalise available evidence to generate a solution. However, moving from strictly cognitive inductive reasoning to social cognitive reasoning may require drawing new boundary conditions to the generalisation process. For example, Pettigrew (1981) noted that generalisation of social information may not occur due to substantial social support for stereotypes coming from media depictions, power relations and social norms (for a review of stereotypic media depictions, see Harris, 1994). In addition, generalisation of social information is a complex process that requires a number of factors to be successful. For example, Hewstone (1994) argued that exemplar attributes need to be perceived as relevant and associated with a typical group member. It is therefore important to start the investigation of member-to-group generalisation from an objective and quantifying assessment of the underlying assumption that out-group member information does indeed generalise to the out-group in general.

The first proposition requires comparing studies that use a no-information control group and an experimental group provided with out-group sample information. A within-subjects design is also appropriate to investigate the first proposition, whereby a pre-information measure is compared with a post-information measure. If generalisation occurs, the information group should express group judgments that incorporate or reflect the presented exemplar information. Proposition One will be tested using an overall generalisation effect size and separately across each of the outcome measures.
Chapter Two

20

This basic form of member-to-group generalisation should be evidenced by a significant positive overall generalisation effect. I expect to detect a moderating effect of outcome measures that differs from that detected by Tropp and Pettigrew’s (2005) analysis of generalisation through contact. Tropp and Pettigrew showed meta-analytically that the size of contact effects depends, among other things, on the outcome measure. As noted above, in their contact meta-analysis they found larger generalisation effect sizes for affective outcome measures. Face-to-face contact is an affective experience (Dovidio, Brigham, Johnson, & Gaertner, 1996; Stangor, Sullivan, & Ford, 1991). Thus, Tropp and Pettigrew found larger generalisation under conditions of match between an affective experience and an affective outcome measure. Research with the impression formation paradigm is typically conducted in a laboratory or classroom, that is, away from direct contact or exposure to the out-group member/s and with a focus on information (e.g., “please try and form an impression of each of these accountants”). This methodology should reduce affective influences on participants judgments of the out-group and encourage participants to remain emotionally detached from the out-group judgments. Hence, the impression formation paradigm should facilitate cognitive processes over affective processes (for a similar point, see Paolini et al., 2007). This is because when the experience with the out-group member is cognitive in nature, the cognitive aspects of the experience should become more accessible and the emotional experience less prominent (Wright, Aron, McLaughlin-Volpe, & Ropp, 1997). Thus, as in Tropp and Pettigrew I expect a match between experience type and measure type, however within an impression formation paradigm this should result in a generalisation advantage for cognitive outcome measures (stereotypicality and dispersion) over affective outcome measures (prejudice).

Indirect evidence for this ‘matching hypothesis’ was documented in a series of studies by Paolini and colleagues (2007; for ‘matching’ evidence in attitude change literature, see Fabringar & Petty, 1999; Haddock, Maio, Arnold & Huskinson, 2008). In Paolini et al.’s work, indirect intergroup friendship (a relatively cognitive form of contact that relies on the knowledge of intergroup friendships) was found to be more effective at changing out-group views among cognitively-orientated individuals and for cognitively-based group perceptions. In addition, direct intergroup friendship (a relatively affective form of contact that relies on experiences with intergroup friends) was found to be more effective among affectively-orientated individuals and for affective-laden group perceptions. This meta-analysis will test the cognitive portion of this matching hypothesis under a cognitively orientated paradigm.

**Proposition Two: Variables Influencing Generalisation**

Each of the studies included in the assessment of Proposition One will be coded for specific moderating variables to identify the conditions that maximise this basic form of
generalisation. This second proposition deals with variables that influence the strength and/or the direction of the effect of the out-group sample information on the group judgment (for a discussion of moderator variables, see Baron & Kenny, 1986; Howell, 2002). Since it is not statistically parsimonious to assess every suggested moderating variable, I have considered four moderating variables that are common to all of the studies (for a full list of moderating variables, see Paolini, 2001). These include: (1) size of the exemplar information; (2) typicality of the exemplar information; (3) exemplar and out-group valence; and (4) out-group type. The first three moderating variables are critical to the classic models of generalisation (bookkeeping, conversion and subtype) and to the generalised appraisal model. The fourth moderating variable was inspired by Tropp and Pettigrew’s (2005) analysis about the moderating role of measure type and Paolini et al.’s (2007) work on affective versus cognitive-based groups.

Sample size. Inductive reasoning research implies that people use logic to infer from particular examples to more general rules along a ‘law of large numbers’ principle. Within the context of member-to-group generalisation, this principle implies that the larger the number of exemplars the more generalisation should occur (Nisbett et al., 1983; Rothbart, 1981). There is discrete evidence that people generalise more from a larger sample of group members. Nisbett et al. (1983) tested this moderating variable by experimentally allocating participants to one of three conditions with varying sample sizes ($n = 1$ vs. $n = 3$ vs. $n = 20$). Participants in the single exemplar condition were asked to imagine they had just arrived at an unknown island in the Pacific where they came across a native who was obese (or brown in colour). They were then asked to indicate the number of other natives who might also be obese (brown). For the larger sample sizes ($n = 3$ or $n = 20$) the information was adapted to reflect a larger sample (e.g., three natives that were obese/brown). Nisbett et al. found that participants would more readily generalise the sample information to the group as the sample size increased. Accordingly, people complied with the statistical notion that larger samples are better for population estimates than smaller samples (for evidence of a sample size effect within contact literature, see Cernat, 2010).

More recently, Fiedler and Walther (2004) argued that out-group sample size is an important parameter in the process of making out-group judgments. However, they have done so from a very different premise to Nisbett et al. (1983). Their view is that stereotyping is a product of an informationally biased environment rather than a product of cognitively biased individuals. The key environmental factor in their Cognitive-Environmental Learning Approach (or CELA) model is the information sample size that is accessible to the perceiver. From this perspective, increasing exposure to exemplars should increase generalisation (or learning). Their model predicts that people generalise from out-group exemplar information, but the extent of such generalisation depends on the information available from
environmental exposure. Thus, research starting from a learning approach also suggests that larger information sample size should produce enhanced generalisation.

Sample size effects are also consistent with Rothbart’s (1981) bookkeeping model of stereotype change. Rothbart argues that as the size of the disconfirming sample information increases, there should be a corresponding increase in generalisation. It is important to note, in this model the information generalised to group judgments is always disconfirming in nature. This portion of my meta-analysis will also shed a light on the unique role of disconfirming sample size away from exemplar typicality.

By including sample size as one of the moderators, this review will identify support, or otherwise, for the basic law of large numbers. This law will find support if a meta-regression analysis reveals a significant effect using (total) sample size as the moderating variable. By including disconfirming sample size as a separate moderator, this review will identify support for the bookkeeping model. The bookkeeping model will find support if a meta-regression analysis reveals a significant effect when disconfirming sample size is included as the moderating variable in a meta-regression analysis. The key difference is that the sample size effect will code for the total sample size presented to the participants while the bookkeeping effect will code for the disconfirming sample size presented to the participants. In both sets of analysis, I expect to find that member-to-group generalisation increases as the (disconfirming) sample size increases.

Johnston and Hewstone (1992) reported research supporting the role of disconfirming sample size. In their research, psychology students were presented with information about eight physics students. In one group the participants learned about two extremely stereotype disconfirming physics students (concentrated condition) and in another group the participants learned about six moderately disconfirming physics students (dispersed condition). Outcome measures of group stereotypicality showed the greatest change amongst participants presented with the larger sample of disconfirming group members. Of course, within this study as in any concentrated/dispersed paradigm study, the role of the disconfirming sample size is confounded with exemplar typicality. By coding separately for disconfirming sample size and exemplar typicality, this meta-analysis will help clarify the different role of these two factors in the process of member-to-group generalisation. For now, it is suffice to note that support for the bookkeeping model will emerge if generalisation is enhanced as the sample size of disconfirming exemplars increases.

**Exemplar typicality.** Brown and Hewstone (2005) reviewed the role of exemplar typicality across a mixture of research paradigms and concluded that high typicality of otherwise disconfirming members enhanced generalisation. Similarly, as noted earlier, research from a social cognitive approach has clarified that a _failure_ to generalise exemplar
information is often related to high exemplar atypicality (see comments in Chapter 1 on Rothbart & John, 1985).

Support for the important role of typicality in impression formation research comes from a clever design by Desforges et al. (1997, Study 1). Participants in Desforges et al.’s study were first screened for their pre-information prejudice towards a host of stereotyped social groups (former mental patients, welfare recipients, homosexuals, Palestine Liberation Organisation members, former heroin addicts, members of the Hell’s Angels or American Nazi Party). Next, in an ‘unrelated’ study, participants read the description of an individual whom they were supposed to work with in the next phase of the study. The individual was described as belonging to either one (or two) of the social groups for which they indicated a strong dislike during the pre-information phase of the experiment. In the two-group condition, half of the participants were also told that their partner had been chosen for the study because they were representative (i.e. typical) of one of the two groups. No real interaction ever took place, but participants were asked to imagine interacting cooperatively with the partner while watching a video displaying the alleged partner and another individual interacting together. A post treatment measure of participants’ prejudice towards the social groups followed. All the participants reported more positive feelings and reduced social distance towards the group(s) to which the partner belonged after the vicarious interaction, regardless of whether he belonged to one or two groups. However, the intra-individual change in group prejudice was significantly larger for the group from which the target was a representative member. This research is consistent with Rothbart and John’s (1985) socio-cognitive analysis of generalisation and suggests that information about the member’s typicality can boost generalisation from the sample to the group.

The relative advantage of moderately atypical information over extremely atypical information fits also with the subtype model (Brewer, Dull, & Lui, 1981; Hewstone, 1994; Johnston, Hewstone, Pendry, & Frankish, 1994; Rothbart & John, 1985; Taylor, 1981; Weber & Crocker, 1983). This model was detailed in Chapter 1. Despite the extensive empirical support for the subtype model, it is important to note that extremely atypical group members can still result in generalisation. Indeed, research by Yzerbyt et al. (1999, Study 1) indicated that, at least under conditions of cognitive load, extreme atypicality is able to promote generalisation. In their research participants were exposed to a 2-minute interview with an extremely atypical computer engineer, who was portrayed as being successful but also extroverted. Half of the participants were also asked to perform a computer-based tracking game while listening to the interview (load). Providing support for the conversion model, all the conditions provided with disconfirming information reported reduced group stereotyping. However, generalisation was greatest in the cognitive load condition demonstrating that
subtyping atypical exemplars is, at least partly, dependent on adequate cognitive resources (see also Richards & Hewstone, 2001).

Exemplar typicality will be coded to assess the support for the subtype vs. conversion models. Support for Hewstone and colleague’s (1992) subtype model will be identified if moderately atypical information is generalised more than extremely atypical information (Hewstone, Hopkins, & Routh, 1992a; Hewstone, Johnston, & Arid, 1992b). Support for Rothbart’s (1981) conversion model will be identified if participants generalise extremely atypical information more than moderately atypical information.

**Exemplar and out-group valence.** The generalised appraisal model predicts that out-group information will be generalised, but only when the exemplar information is positive in nature (see Bodenhausen et al., 1995). Allport (1954) supported the role of exemplar valence by specifying that positive contact serves to maximise change in intergroup attitudes. Furthermore, Paolini, Hewstone, Cairns, and Voci (2004a) found that contact resulted in strongest generalisation amongst people who had friends (a particularly positive form of contact) from the out-group. However, a review of psychological phenomenon by Baumeister, Bratslavsky, Finkenauer, and Vohs (2001) from cognitive, social and health psychology demonstrated a generalisation advantage of negative information. The advantage of negative information should reflect the fact that animals and humans have evolved to attend more strongly to negative than positive information. Within the contact tradition and in line with Baumeister et al., Paolini, Harwood and Rubin (under review) found that negative contact with out-group members caused higher category salience, resulting in worsened intergroup relations. Coding for the valence of the information provided to participants will meta-analytically test the role of positive versus negative exemplar information within impression formation research.

It is also possible that generalisation is influenced by the valence of the group. For example, Huici et al. (1996) provided participants with negative exemplar information about irresponsible teachers. They found that generalisation of the negative information occurred predominately amongst participants who already held a negative view of the group as a whole, but not amongst participants who had a positive view of the group as a whole, suggesting that the effect of exemplar valence might be moderated by group valence. Coding for both exemplar valence and group valence will increase our understanding of limitations on generalisation and enhance our understanding of the complex interplay between these variables.

**Out-group type.** If member-to-group generalisation is a basic principle, then generalisation effects should be similar across different types of out-groups. However, the generalised appraisal model (Bodenhausen et al., 1995) predicts that out-group type may influence generalisation. Bodenhausen et al. argued that the more variability that is associated
with an out-group the less likely there will be stereotype change after exposure to counterstereotypical information (cf. Hantzi, 1995). As noted earlier, Pettigrew and Tropp (2006) investigated the role of out-group type in their meta-analysis of the contact literature. Given the larger study pool, they were able to split their sample of studies into a number of social out-groups, including racial/ethnic and nonracial/nonethnic out-groups. Pettigrew and Tropp found statistically similar effect sizes amongst each of the out-group categories, with the largest effect amongst out-groups of a different sexual orientation ($d = -.563; r = -.271$), followed by physically disabled out-groups ($d = -.501; r = -.243$), then nonracial/nonethnic groups ($d = -.451; r = -.220$), racial/ethnic ($d = -.447; r = -.218$), mentally disabled ($d = -.423; r = .207$), mentally ill ($d = -.374; r = -.184$), and the elderly ($d = -.368; r = -.181$). Coding for the moderating variable of out-group type will provide a parallel to Pettigrew and Tropp’s analysis across impression formation studies.

There are limited direct tests on the role of out-group type within impression formation studies, partly because the paradigm is constrained by a focus on a single out-group (for emerging research on out-group to out-group generalisation in the contact literature, see Pettigrew, 2009; Tausch, Hewstone, Singh, Ghosj, & Biswas 2004). Hantzi (1995) provided some evidence of different generalisation rates across out-group types. Hantzi’s participants were first presented with a set of profiles about doctors, teachers, accountants, or architects in a concentrated-dispersed fashion. These groups were selected from pre-testing to represent different rates of familiarity (low for accountants and architects, high for doctors and teachers) and different rates of variability (low for doctors and accountants, high for teachers and architects). After the impression phase, participants rated the stereotypicality of the target group. Consistent with Hewstone’s (1994) review of the concentrated-dispersed literature, generalisation was enhanced amongst the dispersed conditions. Interestingly, this effect only held for familiar groups. For unfamiliar groups, generalisation was similar in the dispersed and concentrated conditions. From Hantzi’s research, it seems that the type of out-group influences the extent of generalisation.

Further moderating evidence for out-group type comes from Mackie, Allison, Worth, and Asuncion (1992), who found different rates of generalisation towards different out-groups. In their Study 1, participants first rated the perceived intelligence of various professional groups. They were then informed that a team of either teachers (stereotypically intelligent) or custodians (stereotypically non-intelligent) entered a college competition on problem solving and either succeeded or failed. The team performance was kept constant across groups, whereas the winning criteria were systematically varied across conditions making it easier or harder to succeed. Despite this, participants generalised the team success vs. failure information to the custodian out-group and rated them as being more intelligent after success than after failure. This finding did not hold for teachers, reflecting an asymmetry
in the effect as a function of group type and indicating that generalisation may be limited amongst some groups and enhanced for others. Coding for out-group type will allow for a systematic investigation of this potential moderating variable.

Due to the comparatively small number of studies available in the present research, out-groups have been coded along a two-level variable, as either cognitively-based out-groups (e.g., artificial groups, occupation groups and student academic majors) or affectively based out-groups (e.g., racial/ethnic groups, gender, elderly, or sexual orientation). This distinction is based on the prominence of beliefs versus emotions in people’s first responses to the group and associates more emotional responses to affective out-groups than to cognitive out-groups (Tajfel, 1982; Tajfel & Turner, 1986; for a similar distinction and operationalisation, see Paolini et al., 2007). It is worth noting that this difference also co-varies with the permanency of group boundaries. Cognitively based groups typically allow for movement between groups (e.g., students can change academic majors), whereas affectively based groups make it harder for movement between groups. In my work, these a-priori differences between cognitive and affective groups will also be corroborated empirically with a pilot study.

The categorisation of out-group type as either cognitive or affective is clearly different from the racial/ethnic versus nonracial/nonethnic coding strategy adopted by Pettigrew and Tropp (2006) and reflects a different emphasis and study availability in each research tradition. Because impression formation research is interested in the process of stereotype change, researchers have often avoided racial/ethnic groups which may shape affective responses and concerns of political correctness. Instead, researchers have used more cognitively based groups, such as student groups (e.g., Hewstone & Hamberger, 2000; Oakes, Haslam, & Reynolds, 1999; Quattrone & Jones, 1980) or artificial groups (e.g., Park & Hastie, 1987; Wilder et al., 1996). This makes Pettigrew and Tropp’s distinction between racial/ethnic vs. nonracial/nonethnic groups non-testable within the impression formation tradition. The alternative affective vs. cognitive split will result in a more balanced split of the available tests.

Due to the greater reliance on cognitive processes in impression formation research, I expect cognitive groups to be more open to generalisation effects than affective groups (for a similar application to artificial groups, see Paolini et al., 2009). Paolini et al.’s (2007) contact research on direct and indirect intergroup friendships considered the role of out-group type. They tested whether the type of friendship was direct or indirect and found that indirect friendship (a cognitive experience of out-group members) resulted in larger effects of generalisation when cognitive, rather than affective, out-groups were considered. While direct friendship effects were larger when affective out-groups were considered. These authors argued for a match between outcome measures and out-group type. For the current review this
matching hypothesis means that generalisation effects should be strongest amongst cognitive out-groups on cognitive measures and amongst affective out-groups on affective measures.

**Summary**

The advantage of a meta-analytic review is that it allows for cross study comparisons that are, otherwise, unachievable in single out-group studies. The purpose of this meta-analysis is to address two fundamental questions surrounding the member-to-group generalisation literature. Firstly, I will look at the underlying assumption that the presentation of exemplar material is generalised to out-group judgments. Secondly, I will look at four moderating variables that update and strengthen the previous narrative reviews (see, Hewstone, 1994; Paolini, 2001), allowing me to assess the support for different models of generalisation. I will now detail the technical overarching decisions surrounding my general meta-analytic approach.

**Overview of the Meta-analytic Approach and Selected Method**

**Rationale for selecting the Hedges and Olkin approach.** There are three major meta-analytic approaches. The oldest, developed by Rosenthal and Rubin (1978, 1988; Rosenthal, 1991), uses Fisher’s r-to-Z transformation for effect sizes and converts study outcomes to standard normal Z-scores for significance levels. A second approach by Hedges and Olkin (1985) converts study outcomes to standard units, using Cohen’s $d$ value, and then corrects for bias using the pooled standard deviation. A third approach, by Hunter and colleagues (Hunter & Schmidt, 1990; Hunter, Schmidt, & Jackson, 1982), corrects effect sizes for sources of errors. Schmidt and Hunter (1999) compared the three techniques using dummy variables provided by Johnson, Mullen and Salas (1995). Schmidt and Hunter found that the outcomes were ‘functionally identical for the three methods (p. 145)’.

To select an appropriate meta-analytical approach, I will follow Johnson and Eagly’s (2000) suggestion that the convention is to use Rosenthal and Rubin’s approach if most of the studies report correlations between two continuous variables (e.g., contact literature) and to use Hedges and Olkin’s (1985) $d$ if the majority of the studies report ANOVAs and $t$-tests for comparisons between means. As impression formation studies investigate comparisons between two groups and primarily report generalisation through ANOVAs or $t$-tests, Hedges and Olkin’s technique was adopted as my meta-analytic technique.

**Rationale for using Hedges’ $g$.** Hedges and Olkin (1985) recommend Cohen’s $d$ as the effect size estimate. Cohen’s $d$ is defined as the difference between two means divided by the standard deviation of the population (Cohen, 1988). In instances where the population standard deviation is unknown, a different estimate, Hedges’ $g$ (Hedges, 1981) is preferred over Cohen’s $d$. Hedges’ $g$ and Cohen’s $d$ are similar and produce very similar effect sizes
(e.g. Cooper & Hedges, 1994) and are algebraically equivalent to a correlation co-efficient. Importantly, Hedges’ $g$ corrects for bias associated with estimating the population standard deviation using a gamma function (Hedges, 1981). As the population standard deviation is often unknown in impression formation research, I used the less biased estimate and will report Hedges’ $g$. As Hedges’ $g$ is scale-free (Timm, 2004), I will also include an $r$ co-efficient for ease of interpretation.

**Rationale for adopting a random effects approach.** There are two statistical approaches that can be adopted when conducting a meta-analysis: a fixed effects approach and a random effects approach. A fixed effect analysis assumes that the only source of uncertainty is from sampling people into studies. This framework makes the assumption that any inferences from the current collection of studies will apply to future studies. In contrast, a random effect analysis assumes that there are two sources of variability: one due to the sampling of people into studies and one due to sampling variability (e.g., Hedges, 1994; Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Mosteller & Colditz, 1996; Raudenbush, 1994; for a review see, Normand, 1995). As this meta-analysis will synthesise data from studies with a range of participant populations, I adopted a random effects approach which corrects for variability across studies (for a similar point see, Cook et al., 1992).

A mixed-effects analysis was used to analyse the moderating variables. This approach uses a random effects model to combine tests within each subgroup and does not assume that test-to-test variance is the same across subgroups. This approach computes variability within subgroups, not pooled across subgroups as the fixed effects model does, resulting in slightly more conservative results than a fixed effects model.

**Rationale for reporting of outcome measures.** Meta-analytic reviews are often criticised for comparing ‘apples with oranges’ whereby studies measuring vastly different outcome variables are compared, resulting in a comparison between non-comparable items (Lipsey & Wilson, 2001). I will present an overall generalisation effect size that combines the data for the three primary outcome variables used in impression formation research: stereotypicality, dispersion and prejudice. The criticism of ‘apples and oranges’ will be addressed by separately analysing the data across each outcome measure (i.e., stereotypicality, dispersion and prejudice). Tropp and Pettigrew (2005) adopted a similar strategy describing an overall effect size and an effect size for different outcome measures.

**Summary.** Following the theory of Hedges and Olkin (1985) Hedges’ $g$ is used as the effect size statistic for each study, which I then meta-analysed using a random effects approach. The data was manipulated using the software program Comprehensive Meta-Analysis, version 2 (Bornstein, Hedges, Higgins, & Rothstein, 2005), and is presented as a combined generalisation effect and separately across outcome measures.
Inclusion Criteria for Identified Studies

Papers were included in this review if they met the following five primary criteria. First, the researchers provided participants with information about out-group members. This requirement included studies that involved the provision of information about cross-cutting categories, but excluded studies where the out-group exemplar information is implied and not directly provided to the participants (e.g., Hewstone et al., 1992a). Studies were also excluded if participants were members of the group under investigation (e.g., Stapel & Koomen, 1997).

Second, the researchers provided exemplar information either in a written format (on paper or computer), or via audio stimulus or videotape. This requirement includes information where the participants ‘overhear’ exposure to out-group information (e.g., Henderson-King & Nisbett, 1996), but excludes (contact) research that investigates face-to-face interaction with out-group members or where close proximity between the participant and members of the out-group existed (e.g., Henderson-King & Nisbett, 1996, Study 1; Wright et al., 1997).

Third, after participants were provided with the out-group exemplar information, the researchers needed to measure the perceptions of the entire out-group, through indicators of group stereotypicality, dispersion, and/or prejudice. Hence, studies were excluded if group stereotyping was measured only with respect to the specific sample of out-group members (e.g., Gross & Hardin, 2008; Stroessner & Mackie, 1992; Stroessner et al., 2005) or to a number of new group members smaller than the one presented in the impression phase (e.g., Gurwitz & Dodge, 1977).

Fourth, the study designs needed to include a no-information control condition, providing a comparative benchmark to identify whether generalisation had occurred in the experimental condition provided with out-group member information. Two designs using a no-information control condition are suitable: (1) tests adopting a between subjects design where a control condition is included with no information, and (2) tests adopting a within subjects design where pre and post information measures are used and the correlation data for the paired values is provided. In the first type of design, generalisation is reflected by a significant difference between the no-information control condition and the experimental conditions provided with the sample information. In the second type of design, generalisation is reflected by a significant difference between the pre (no information) measure and the post (exposure to out-group information) measures. Studies using pre and post information measures were only included when the correlation between the paired values was reported. This restriction was implemented to avoid inflating the sample size which leads to an over estimation of the contribution of the individual study towards the overall analysis (Lispey & Wilson, 2001). This restriction criterion resulted in three papers not being included in the meta-analysis (e.g., Cameron & Rutland, 2006, Studies 1 & 2; Garcia-Marques & Mackie, 2001; Mackie et al., 1992).
Fifth, group measures were reported independently for each condition. This point excluded studies where the group measures were not distinguishable from other dependent variables, or were reported only as measures of co-variation with other dependent variables (e.g., Jackson & Sullivan, 1988). This point also excluded studies where the group measures were collapsed across different conditions. For example, Garcia-Marques and Mackie (1999, Study 1) collapsed data from a no-information control group with a super-congruent group that received information about five confirming exemplars, one disconfirming exemplar and two stereotype irrelevant exemplars.

**Final Sample**

The final analysis across the two propositions included 29 individual papers from 1980-2009, with 46 individual studies and 66 independent tests. In this chapter I will refer to ‘tests’ rather than to ‘studies’ because any one study could yield three different effect sizes (one for each of the outcome variables, e.g., Paolini et al., 2004b). In most cases, only one effect size was derived from a single study. Studies with more than one test only contributed once to the overall generalisation effect to avoid inflating the effect size from that participant sample. The test that was selected for inclusion in the overall generalisation effect was based on the focus of the paper. For example, Paolini et al. (2004b) reported data for the outcome measures of stereotypicality, dispersion and prejudice. In this instance, the primary focus of Paolini et al.’s paper was on understanding the role of dispersion following exposure to information about a single disconfirming member; as a result the effect size from the test of dispersion was included in the overall generalisation analysis.

Combined, the studies represent responses from 3,338 participants with 38 tests conducted in the United States and 11 tests conducted in Europe. The remaining tests \((N = 17)\) were conducted in Canada, Australia, and New Zealand. The majority of tests were published in the *Journal of Personality and Social Psychology* \((N = 25)\), followed by the *Journal of Experimental Social Psychology* \((N = 7)\). The remaining tests \((N = 34)\) were from a variety of other journals. The tests involve a wide variety of out-groups, such as the elderly (Galinsky & Ku, 2004), occupational groups like nurses (Seta, Seta, & McElroy, 2003), lawyers (Garcia-Marques & Mackie, 2001), accountants (Paolini et al., 2004b); racially based groups (e.g. Virj, Akehurst, & Smith, 2003) and ethnicity based groups (e.g. Bless, Schwarz, Bodenhausen, & Thiel, 2001). A number of studies investigated artificial, or laboratory created, groups (e.g. Guinote, Judd, & Brauer, 2002; Thompson, Judd, & Park, 2000, Study 1). The majority of tests measured group stereotyping \((N = 30)\) tests, followed by prejudice \((N = 25)\) tests and then group dispersion \((N = 11)\) tests.
Method

Procedure

Locating relevant studies. Studies were retrieved from relevant databases (e.g., PsycINFO 1887-2009, ERIC, SocAbs, ProQuest, Google scholar). The search used 26 terms such as prejudice, stereotype, and names of prominent authors (e.g., Miles Hewstone; Ziva Kunda). To seek unpublished work, emails were sent to social psychologists’ networks asking for any empirical work in the area of member-to-group generalisation (e.g., Society of Australasian Social Psychologists, British Psychological Society’s Social Psychology Section; European Association of Social Psychology). I also conducted a retrospective search by combing the reference lists of located studies.

Coding of each study. Each study was independently coded by myself and also by one of two female psychology research assistants. The research assistants were trained following the principles outlined by Lipsey and Wilson (2001, pp. 88-90). This started with training in the use and understanding of the coding protocol. Next, both research assistants coded a small set of studies and the results were compared and discussed to resolve any problems. The research assistants then independently coded half of the papers. Three meetings were held during coding where we discussed the papers and the coding results. Inter-rater agreement on each item was identified and discrepancies were resolved through discussion.

Sample size was coded by recording both the total number of exemplars and the number of disconfirming exemplars presented to participants. The total sample size ranged from 1 to $n = 30$ (Hoffman & Hurst, 1990) and the disconfirming sample size ranged from 0 to $n = 15$ (Hoffman & Hurst, 1990). Inter-rater agreement rate on total sample size was 95%, while the inter-rater agreement rate on disconfirming sample size was 90% (for use of percentage inter-rater agreement, see Lipsey & Wilson, 2001; also see Bettencourt, Charlton, Dorr, & Hume, 2001). The majority of discrepancies between coders on these variables were due to typographical errors when entering the data onto the electronic database. These errors were easily recognised and reconciled through discussion.

Typicality was tightly defined in the coding protocol to minimise discrepancies. Individual exemplars were coded as being moderately atypical when 50% or less of the information was stereotype disconfirming. Exemplars were considered to be extremely atypical when more than 50% of the information was disconfirming. When the information presented to the participants was stereotype confirming, the exemplars were coded as typical. When more than one exemplar was presented in the sample, a mode of the sample typicality was used, so if two exemplars were moderately atypical and one was extremely atypical, the sample was coded as moderately atypical. The overall inter-rater agreement on this variable was 80%, and discrepancies were resolved through discussion.
Exemplar valence was defined in the coding protocol as the evaluative direction (positive, negative, mixed, neutral) of the exemplar information. For example, if participants were provided with the sample information of an ‘irresponsible’ person, exemplar valence was coded as negative (i.e., irrespective of whether this information is stereotype consistent or inconsistent). Alternatively, if participants were provided with information of a ‘hardworking person’, then exemplar valence was coded as positive. The overall inter-rater agreement on this item was 76%. The valence of the out-group was also coded. Out-groups were classified as being positive, negative or neutral in valence. The overall inter-rater agreement on the out-group valence item was 95%.

The decision to allocate out-groups to either an affective or cognitive category was guided by a pilot study. I expected that groups associated with higher feelings and emotions would include groups such as gender and ethnicity. While groups associated with thoughts and cognitions would include occupational groups and student majors. Participants in Pilot Study 1 were six individuals (2 males, 4 females) from the same community as my experimental participants and had a mean age of 26.67 years (2.33 SD). Participants were approached in a public shopping centre and asked to complete a 5 minute anonymous survey. They were asked to read a list of 25 out-groups that had been used in impression formation research and rank them from 1 – 25, assigning the highest ranks (e.g., first, second, etc.) to the groups most associated with their emotions and feelings (i.e., affective groups) and assigning lowest ranks to the groups that were most associated with their thoughts and beliefs (i.e., cognitive groups; measures adopted from Paolini et al., 2007). The individual ranks were averaged across participants to create a mean ranking. Of the top 12 (emotionally-based) groups, 11 were groups not linked to employment/academic groups and were coded in the meta-analysis as being affective. These groups included: Black people, people with a disability, elderly people, females, homeless people, homosexuals, immigrants, males, mentally ill people and religious groups. The bottom 13 (cognitively-based) rankings were employment/academic groups and were coded in the meta-analysis as being cognitive. These included: accountants, child care workers, cleaners, doctors, education students, physics students, politicians, psychology students and teachers. In the coding protocol, out-group type was then coded by categorising each group under investigation as being either an affective social group (e.g., racial/ethnicity, gender, disability, age) or a cognitive group (e.g., occupation/student groups, artificial). The inter-rater agreement on this variable was 92%.

The one exception to this affective/cognitive division being split by employment/academic groups was the out-group of ‘bouncers’. Pilot study participants included this group within the top 12 mean ranks. This was most likely attributed to a highly emotive media story being popularised at the time of data collection, where a young Australian tourist was beaten to death by a Greek bouncer while on holidays in Mikonos, in the Greek Islands. Since bouncers are an employment group, I decided to code the out-group of bouncers as cognitive.
**Effect size calculation and analysis.** Hedges’ $g$ was calculated as the effect size using the effect size calculator developed by Wilson (2001). Each effect size was estimated from relevant condition means, $t$-test for independent samples, $F$ statistics for main effects, or $F$ statistics for interactions accompanied by means and sample sizes. Wherever the reported data included the standard deviations for the two conditions (i.e., the no-information control group and the out-group exemplar information experimental group), the pooled standard deviation was computed based on these results. However, when standard deviations were not reported, an estimate of the pooled standard deviation was obtained from analysis of variance results (Lipsey & Wilson, 2001). There were five tests (out of 66; 12%) that reported no significant differences between control and experimental conditions (i.e. absence of generalisation effects), and failed to supply any further statistics (e.g., Batson et al., 1997; Kunda & Oleson, 1997, Study 4). These were conservatively set to an effect size of zero (Wolf, 1986; also see, Bettencourt et al., 2001). When the exact number of participants allocated to each condition was not specified, an estimate was computed by dividing the number of participants between conditions. Inter-rater agreement for the calculation of the $g$ value was 94% with discrepancies resolved through discussion.

Table 2.1 displays the details of the included tests and their coded properties.

**Results**

**Overview**

The results for the meta-analysis are presented in two sections. The first section presents the findings for Proposition One, which looked at the basic effect of member-to-group generalisation. The second section presents the findings for Proposition Two, which looked at the moderating variables expected to influence member-to-group generalisation.

**Proposition One**

Within Proposition One, I first present an overall generalisation effect size and then the effect size across each outcome measure. The use of the random effects model was supported by the heterogeneity of the fixed effects model on the overall effect size ($Q (50) = 95.73, p < .001$). Tests measuring generalisation for each outcome measure were also all heterogeneous: stereotypicality measures ($Q (29) = 70.37, p < .001$), dispersion measures ($Q (11) = 28.59, p < .002$) and prejudice measures ($Q (24) = 46.69, p < .005$). The summary data for the overall effect size and each outcome measure is presented in Table 2.2.

---

3 Wilson’s effect size calculator (2001) refers to Hedges’ $g$ as Cohen’s $d$. 

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study</th>
<th>$g$</th>
<th>$r$</th>
<th>N</th>
<th>DV</th>
<th>Size</th>
<th>Disc size</th>
<th>Prop discon</th>
<th>Typic</th>
<th>Eval</th>
<th>Gval</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ames (2004)</td>
<td>2</td>
<td>.36</td>
<td>.18</td>
<td>102</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Ames (2004)</td>
<td>3</td>
<td>.43</td>
<td>.21</td>
<td>72</td>
<td>S</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Batson et al. (1997)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>46</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Batson et al. (1997)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>48</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Blair et al. (2001)</td>
<td>1</td>
<td>.66</td>
<td>.31</td>
<td>39</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Blair et al. (2001)</td>
<td>2</td>
<td>.51</td>
<td>.25</td>
<td>80</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Blair et al. (2001)</td>
<td>5</td>
<td>.38</td>
<td>.19</td>
<td>128</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Bless &amp; Schwarz (1998)</td>
<td>1</td>
<td>.76</td>
<td>.36</td>
<td>48</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Bodenhausen et al. (1995)</td>
<td>1</td>
<td>.16</td>
<td>.08</td>
<td>46</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Bodenhausen et al. (1995)</td>
<td>2</td>
<td>.59</td>
<td>.28</td>
<td>63</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Bodenhausen et al. (1995)</td>
<td>3</td>
<td>.25</td>
<td>.12</td>
<td>63</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Cohrs &amp; Asbrook (2009)</td>
<td>1</td>
<td>.94</td>
<td>.43</td>
<td>78</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Corely &amp; Pollack (1996)</td>
<td>2</td>
<td>.71</td>
<td>.33</td>
<td>78</td>
<td>P</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Corely &amp; Pollack (1996)</td>
<td>2</td>
<td>.92</td>
<td>.42</td>
<td>42</td>
<td>P</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Dasgupta &amp; Greenwald (2001)</td>
<td>1</td>
<td>.91</td>
<td>.41</td>
<td>33</td>
<td>P</td>
<td>20</td>
<td>10</td>
<td>½</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>Dasgupta &amp; Greenwald (2001)</td>
<td>2</td>
<td>.89</td>
<td>.41</td>
<td>26</td>
<td>P</td>
<td>20</td>
<td>10</td>
<td>½</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>Desforges et al. (1997)</td>
<td>1</td>
<td>.75</td>
<td>.35</td>
<td>30</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Desforges et al. (1997)</td>
<td>1</td>
<td>.65</td>
<td>.31</td>
<td>48</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Desforges et al. (1997)</td>
<td>2</td>
<td>.62</td>
<td>.29</td>
<td>56</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Desforges et al. (1997)</td>
<td>2</td>
<td>.30</td>
<td>.47</td>
<td>56</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Duval et al. (2000)</td>
<td>1</td>
<td>.61</td>
<td>.29</td>
<td>37</td>
<td>S</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>Esses &amp; Dovidio (2002)</td>
<td>1</td>
<td>.87</td>
<td>.41</td>
<td>56</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Garcia-Marques &amp; Mackie (1999)</td>
<td>2</td>
<td>.53</td>
<td>.26</td>
<td>40</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>½</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>C</td>
</tr>
</tbody>
</table>

Notes: $g$ = Hedges’ $g$ effect size for each sample; $N$ = Number of participants; DV = outcome measure (S = Stereotypicality; D = Dispersion; P = Prejudice); Size = total out-group sample size; Disc Size = Disconfirming sample size; Prop discon = proportion of sample that is disconfirming; Typic = Typicality of exemplars (1 = typical; 2 = moderately atypical; 3 = extremely atypical; 4 = other/unspecified); Eval = Valence of the exemplar information (1 = positive; 2 = negative; 3 = mixed; 4 = unspecified); Gval = Valence of the out-group (1 = positive; 2 = negative; 3 = mixed); Type = Out-group type (A = affective; C = cognitive); * Not included in overall generalisation analysis.
Table 2.1 (continued)
Tests Meeting Inclusion Criteria for Meta-Analytic Study 1

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study</th>
<th>g</th>
<th>r</th>
<th>N</th>
<th>DV</th>
<th>Size</th>
<th>Disc size</th>
<th>Prop discon</th>
<th>Typic</th>
<th>Eval</th>
<th>Gval</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garcia-Marques &amp; Mackie (1999)*</td>
<td>2</td>
<td>.49</td>
<td>.24</td>
<td>40</td>
<td>D</td>
<td>2</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Garcia-Marques &amp; Mackie (1999)</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>S</td>
<td>8</td>
<td>3</td>
<td>$\frac{3}{8}$</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Garcia-Marques &amp; Mackie (1999)*</td>
<td>3</td>
<td>.45</td>
<td>.22</td>
<td>75</td>
<td>D</td>
<td>8</td>
<td>3</td>
<td>$\frac{3}{8}$</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Hamill et al. (1980)</td>
<td>1</td>
<td>.63</td>
<td>.30</td>
<td>86</td>
<td>S</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Henderson-King &amp; Nisbett (1996)</td>
<td>1</td>
<td>.58</td>
<td>.28</td>
<td>48</td>
<td>S</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Henderson-King &amp; Nisbett (1996)*</td>
<td>1</td>
<td>.51</td>
<td>.25</td>
<td>48</td>
<td>P</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Hoffmann &amp; Hurst (1990)</td>
<td>1</td>
<td>.49</td>
<td>.24</td>
<td>80</td>
<td>S</td>
<td>30</td>
<td>15</td>
<td>$\frac{1}{2}$</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Huici et al. (1996)</td>
<td>1</td>
<td>.70</td>
<td>.33</td>
<td>41</td>
<td>S</td>
<td>6</td>
<td>3</td>
<td>$\frac{1}{2}$</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Huici et al. (1996)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>50</td>
<td>S</td>
<td>6</td>
<td>3</td>
<td>$\frac{1}{2}$</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Kashima &amp; Kashima (1993)</td>
<td>1</td>
<td>.54</td>
<td>.26</td>
<td>40</td>
<td>D</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Kashima &amp; Kashima (1993)</td>
<td>1</td>
<td>.58</td>
<td>.28</td>
<td>40</td>
<td>D</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Kunda &amp; Oleson (1995)</td>
<td>1</td>
<td>.86</td>
<td>.40</td>
<td>32</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Kunda &amp; Oleson (1997)</td>
<td>1</td>
<td>1.22</td>
<td>.52</td>
<td>38</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Kunda &amp; Oleson (1997)</td>
<td>2</td>
<td>.61</td>
<td>.29</td>
<td>47</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Kunda &amp; Oleson (1997)</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>Paolini et al. (2004b)</td>
<td>1</td>
<td>.88</td>
<td>.40</td>
<td>40</td>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)*</td>
<td>1</td>
<td>.55</td>
<td>.27</td>
<td>40</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)</td>
<td>1</td>
<td>.28</td>
<td>.14</td>
<td>40</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)</td>
<td>2</td>
<td>.88</td>
<td>.40</td>
<td>40</td>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)*</td>
<td>2</td>
<td>.11</td>
<td>.05</td>
<td>40</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)</td>
<td>2</td>
<td>-.26</td>
<td>-.13</td>
<td>40</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)</td>
<td>3</td>
<td>-.20</td>
<td>-.10</td>
<td>40</td>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)*</td>
<td>3</td>
<td>.16</td>
<td>.08</td>
<td>40</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
</tbody>
</table>

Notes: $g$ = Hedges’ $g$ effect size for each sample; $N$ = Number of participants; DV = outcome measure (S = Stereotypicality; D = Dispersion; P = Prejudice); Size = total out-group sample size; Disc Size = Disconfirming sample size; Prop discon = proportion of sample that is disconfirming; Typic = Typicality of exemplars (1 = typical; 2 = moderately atypical; 3 = extremely atypical; 4 = other/unspecified); Eval = Valence of the exemplar information (1 = positive; 2 = negative; 3 = mixed; 4 = unspecified); Gval = Valence of the out-group (1 = positive; 2 = negative; 3 = mixed); Type = Out-group type (A = affective; C = cognitive); * Not included in overall generalisation analysis.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Study</th>
<th>$g$</th>
<th>$r$</th>
<th>$N$</th>
<th>DV</th>
<th>Size</th>
<th>Disc size</th>
<th>Prop discon</th>
<th>Typic</th>
<th>Eval</th>
<th>Gval</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paolini et al. (2004b)*</td>
<td>3</td>
<td>-1.14</td>
<td>-0.07</td>
<td>40</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)</td>
<td>4</td>
<td>1.05</td>
<td>0.46</td>
<td>50</td>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)*</td>
<td>4</td>
<td>-0.72</td>
<td>-0.34</td>
<td>50</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Paolini et al. (2004b)*</td>
<td>4</td>
<td>-0.30</td>
<td>-0.15</td>
<td>50</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Penn et al. (2003)</td>
<td>1</td>
<td>0.48</td>
<td>0.23</td>
<td>77</td>
<td>S</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Penn et al. (2003)</td>
<td>1</td>
<td>0.66</td>
<td>0.31</td>
<td>77</td>
<td>P</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Rothbart &amp; Lewis (1988)</td>
<td>1</td>
<td>1.10</td>
<td>0.48</td>
<td>48</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Rothbart &amp; Lewis (1988)</td>
<td>1</td>
<td>0.40</td>
<td>0.20</td>
<td>48</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Seta et al. (2003)</td>
<td>5</td>
<td>0.15</td>
<td>0.08</td>
<td>52</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Stratton et al. (2006)</td>
<td>1</td>
<td>0.14</td>
<td>0.07</td>
<td>28</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Stratton et al. (2006)</td>
<td>2</td>
<td>-0.80</td>
<td>-0.37</td>
<td>17</td>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>Stratton et al. (2006)*</td>
<td>2</td>
<td>0.32</td>
<td>0.16</td>
<td>17</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>Vescio et al. (2003)</td>
<td>1</td>
<td>0.53</td>
<td>0.26</td>
<td>64</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>Virj et al. (2003)</td>
<td>1</td>
<td>-0.25</td>
<td>-0.13</td>
<td>50</td>
<td>P</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>Weber &amp; Crocker (1983)</td>
<td>1</td>
<td>2.06</td>
<td>0.72</td>
<td>40</td>
<td>S</td>
<td>30</td>
<td>10</td>
<td>$\frac{1}{3}$</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>Wyer et al. (2002)</td>
<td>1</td>
<td>1.33</td>
<td>0.55</td>
<td>75</td>
<td>S</td>
<td>12</td>
<td>8</td>
<td>2$\frac{1}{3}$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Wyer et al. (2002)*</td>
<td>1</td>
<td>0.69</td>
<td>0.33</td>
<td>75</td>
<td>D</td>
<td>12</td>
<td>8</td>
<td>2$\frac{1}{3}$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Wyer et al. (2002)</td>
<td>2</td>
<td>1.10</td>
<td>0.48</td>
<td>101</td>
<td>S</td>
<td>12</td>
<td>8</td>
<td>2$\frac{1}{3}$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Wyer et al. (2002)*</td>
<td>2</td>
<td>1.41</td>
<td>0.58</td>
<td>101</td>
<td>D</td>
<td>12</td>
<td>8</td>
<td>2$\frac{1}{3}$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Yzerbyt et al. (1999)</td>
<td>1</td>
<td>-0.22</td>
<td>-0.11</td>
<td>27</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
</tbody>
</table>

Notes: $g =$ Hedges’ g effect size for each sample; $N =$ Number of participants; DV = outcome measure (S = Stereotypicity; D = Dispersion; P = Prejudice); Size = total out-group sample size; Disc Size = Disconfirming sample size; Prop discon = proportion of sample that is disconfirming; Typic = Typicality of exemplars (1 = typical; 2 = moderately atypical; 3 = extremely atypical; 4 = other/unspecified); Eval = Valence of the exemplar information (1 = positive; 2 = negative; 3 = mixed; 4 = unspecified); Gval = Valence of the out-group (1 = positive; 2 = negative; 3 = mixed; 4 = unspecified); Type = Out-group type (A = affective; C = cognitive); *Not included in overall generalisation analysis.
Table 2.2
Summary Data for Overall Generalisation and each Outcome Measure

<table>
<thead>
<tr>
<th></th>
<th>N (N%)</th>
<th>Total Ps</th>
<th>Fail-safe n</th>
<th>g</th>
<th>r</th>
<th>SE</th>
<th>95% CI</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>51 (77)</td>
<td>2,680</td>
<td>2,329</td>
<td>.54</td>
<td>.26</td>
<td>.06</td>
<td>.43/.65</td>
<td>9.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stereotypicality</td>
<td>30 (45)</td>
<td>1,625</td>
<td>732</td>
<td>.50</td>
<td>.24</td>
<td>.08</td>
<td>.34/.65</td>
<td>6.31</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Dispersion</td>
<td>11 (17)</td>
<td>563</td>
<td>119</td>
<td>.59</td>
<td>.28</td>
<td>.15</td>
<td>.30/.89</td>
<td>3.90</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prejudice</td>
<td>25 (38)</td>
<td>1,150</td>
<td>300</td>
<td>.42</td>
<td>.21</td>
<td>.08</td>
<td>.26/.58</td>
<td>5.04</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Notes. N = Number of tests; N% = percentage of the 66 total tests; Ps = Participants; g = mean effect size weighted by the inverse variance; SE = standard error of g; 95% CI = 95% confidence interval of g; Z = the standard score; p = probability of g.

Overall generalisation effect. Six studies presented data on multiple outcome measures (e.g., Paolini et al., 2004b). As detailed above, data from any one participant was only included once within the analysis of the overall generalisation effect (Lipsey & Wilson, 2000). The included outcome measure was the one identified by the researchers as the primary focus of their paper. This culling left 51 tests that contributed to the overall generalisation effect. The removed tests were still included in the individual outcome measure analysis.

The mean effect size was significant ($M_g = .54, r = .26, p < .001$), indicating that out-group member information does generalise to the out-group. Cohen (1977, 1988) indicates that $d = .54$ would be a medium effect size, since the $g$ estimate is very similar to $d$ (albeit slightly more conservative), this effect magnitude is also considered medium. The confidence limits for $g$ (.43/.65) suggest that even in the worst case scenario, this effect size would still be small.

Publication bias was assessed using Cooper’s (1989) fail-safe n method, which estimates the number of null tests required to nullify the effect. The obtained mean effect of member-to-group generalisation was unlikely to be due to publication bias as 2,329 null tests are required to nullify this effect.

It is worth noting that the obtained effect size for Proposition One is slightly larger than the basic contact effect size identified by Pettigrew and Tropp’s (2000) meta-analysis ($r = -.23$). Together with Pettigrew and Tropp (2000), it is apparent that inductive reasoning does occur in the social domain, and that information exposure and face-to-face interactions both produce generalisations from the exemplar to the out-group.

Stereotypicality. The results indicate that exemplar information is generalised to the out-group on the outcome measure of stereotypicality ($M_g = .50, r = .24, p < .001$). Cooper’s (1989) fail-safe n assessed publication bias and indicated that this effect is unlikely to be due to publication bias. It is interesting to note that stereotypicality measures are the more frequently adopted outcome measure, with almost half of the tests in this meta-analysis adopting this measure to assess group judgments.

Dispersion. Dispersion has the strongest generalisation effect ($M_g = .59, r = .28, p < .001$). However, it must be noted that the difference between dispersion and the other two
outcome measures was not statistically reliable (both $Qs (1) < 1$). In addition, dispersion has been the least adopted outcome variable, and subsequently is comparatively less stable (e.g., there is a high standard error, which is almost double the standard error rate for stereotypicality and prejudice). The large confidence interval (.30/.89) suggests that the advantage of dispersion (over the other outcome measures) in detecting generalisation should be taken cautiously. More studies using dispersion as the outcome measure would be valuable in further understanding the status of this measure.

**Prejudice.** Across the three outcome variables prejudice had the lowest effect size. However, it is important to stress that generalisation along this variable was still significant ($M_g = .42, r = .21, p < .005$) indicating that out-group information generalises to measures of prejudice (Table 2.2). Although this is a more modest effect size (Cohen, 1977, 1988) than that for stereotypicality and dispersion, it is still highly significant suggesting that generalisation does occur when assessing it along prejudice. As predicted, this more modest effect size may be tied to the affective nature of prejudice measures and its mismatch with the type of experience employed in most social cognitive studies (Paolini et al., 2007).

**Proposition One: Concluding Comments**

Focusing on the basic generalisation effect size from impression formation studies that included a no-information control group revealed that people do generalise information about individual out-group members to the out-group (e.g., Duval, Ruscher, Welsh, & Catanese, 2000). The generalisation of sample information to the out-group has been an implicit assumption amongst member-to-group generalisation researchers, and this meta-analysis now provides clear evidence for the soundness of this approach. Altogether, this initial evidence suggests that the inclusion of exemplar information in the out-group judgment may be a default outcome (Schwarz & Bless, 1992; for a similar point see, Paolini 2001).

The effect size for generalisation across impression formation research was found to be slightly larger than the size identified by Pettigrew and Tropp’s (2000; 2006) meta-analysis ($r = .26$ vs. $r = -.23$). Together with Pettigrew and Tropp’s data, this meta-analytic study indicates that people are able and willing to modify their group judgments and that this occurs both after exposure to information about out-group members as well as after face-to-face encounters with out-group members.

The pattern of findings across outcome measures in Proposition One identified a relative (although not statistically reliable) advantage for cognitive outcome measures (dispersion and stereotypicality) over more affective outcome measures (prejudice). This pattern is consistent with work by Tropp and Pettigrew (2005) and by Paolini and colleagues (2007). In the contact literature Tropp and Pettigrew (2005) found an advantage of affective outcome measures and concluded that “affective outcomes of intergroup contact are more likely
to generalise than cognitive outcomes” (p. 1150). As noted by Paolini et al. (2007), the advantage of affective measures in contact research is most likely due to the match between affective outcome measures and the affective basis of the experience. Put differently, the greater generalisation potential of affective measures across the contact literature should reflect the fact that intergroup contact experiences are affective in nature. As people engage in actual face-to-face contact with out-group members, feelings and emotions are likely to be evoked and this matches the affective basis of prejudice measures. Extending this matching-hypothesis to impression formation studies, I confirmed that cognitive outcome measures are more effective at identifying generalisation of cognitive based experiences with out-group members than affective measures. As people engage in impression formation tasks, cognitive processes such as information processing are likely to be evoked and this matches the cognitive basis of stereotyping and dispersion measures.

**Proposition Two: What Factors Maximise Generalisation?**

Proposition Two used the studies presented in Table 2.1 to investigate the role of sample size, exemplar typicality, exemplar and out-group valence, and out-group type as moderating variables of generalisation effect sizes.

**Sample size.** The total sample size was fitted as a continuous moderator variable using a fixed effects meta-regression analysis. Total sample size resulted in a significant effect, $Q (1) = 8.48, p < .01$. These findings indicate that there is a significant sample size effect, whereby larger sample sizes result in greater generalisation. The slope depicting the meta-regression is displayed in Figure 2.1, and the summary statistics are presented in Table 2.3. This analysis confirms the role of the law of large numbers in impression formation research with greater generalisation in the larger sample size (Nisbett et al., 1983).
The number of stereotype disconfirming exemplars (or disconfirming sample size) was also fitted as a continuous moderator variable using a fixed effects meta-regression analysis. Disconfirming sample size also resulted in a significant effect, $Q(1) = 8.96$, $p < .01$. These results replicate the pattern that emerged in the total sample size effect (for summary data, see Table 2.3; for graphical representation, see Figure 2.2). This sample size effect supports Rothbart’s (1981) bookkeeping model of stereotype change confirming that the larger the number of disconfirming exemplars the larger the stereotype change.

Table 2.3  
*Summary Data for Sample Size Across all Outcome Measures*

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$N$ (N%)</th>
<th>$Q$</th>
<th>Slope</th>
<th>SE</th>
<th>95% CI</th>
<th>Z</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>51 (77)</td>
<td>8.48</td>
<td>.02</td>
<td>.01</td>
<td>.01/.03</td>
<td>2.91</td>
<td>.004</td>
</tr>
<tr>
<td>Stereotypicality</td>
<td>30 (45)</td>
<td>8.81</td>
<td>.02</td>
<td>.01</td>
<td>.01/.03</td>
<td>2.97</td>
<td>.003</td>
</tr>
<tr>
<td>Dispersion</td>
<td>11 (17)</td>
<td>4.61</td>
<td>.04</td>
<td>.02</td>
<td>.004/.08</td>
<td>2.15</td>
<td>.03</td>
</tr>
<tr>
<td>Prejudice</td>
<td>25 (38)</td>
<td>3.60</td>
<td>.03</td>
<td>.01</td>
<td>.06/.50</td>
<td>1.90</td>
<td>.06</td>
</tr>
<tr>
<td>Disconfirming sample</td>
<td>45 (68)</td>
<td>8.96</td>
<td>.04</td>
<td>.01</td>
<td>.01/.06</td>
<td>2.99</td>
<td>.003</td>
</tr>
<tr>
<td>Stereotypicality</td>
<td>26 (39)</td>
<td>9.55</td>
<td>.04</td>
<td>.01</td>
<td>.02/.07</td>
<td>3.09</td>
<td>.002</td>
</tr>
<tr>
<td>Dispersion</td>
<td>9 (14)</td>
<td>5.44</td>
<td>.06</td>
<td>.03</td>
<td>.01/.11</td>
<td>2.33</td>
<td>.01</td>
</tr>
<tr>
<td>Prejudice</td>
<td>23 (35)</td>
<td>2.41</td>
<td>.05</td>
<td>.03</td>
<td>-.01/.10</td>
<td>1.55</td>
<td>.12</td>
</tr>
</tbody>
</table>

*Notes.* $N =$ Number of tests; $N/$ = percentage of the 66 total tests; SE = standard error of $g$; 95% CI = 95% confidence interval of $g$; $Z =$ the standard score; $p =$ probability of $g$.

A ratio between the number of disconfirming exemplars and the total sample size was also created to gauge the degree to which disconfirming exemplars might stand out within the overall presented information. This ratio was then run as a meta-regression; the analysis returned non-significant results, the estimated slope was -.10 with a 95% CI at -.32/.10, $Z < 1$, $p = .33$. This indicates that the ratio of disconfirming exemplars to total exemplars is not important, rather it is the absolute number of disconfirming exemplars and the absolute number of exemplars presented that drive generalisation effects.

*Figure 2.2.* Meta-regression as a function of disconfirming sample size.
Looking closely at the data as it was distributed across outcome measures, it is interesting to note that sample size did not impact on generalisation across prejudice measures. There was a marginally significant effect on total sample size and a non-significant effect of disconfirming sample size on prejudice measures. This finding supports the conclusion from Proposition One that affective outcome measures may not be applicable to the more ‘cognitive’ tasks associated with impression formation research. The general advantage of cognitive measures to detect generalisation as the total sample size and disconfirming sample size increases will be discussed more thoroughly below in the concluding comments section.

**Typicality.** There were 7 tests that presented stereotypical exemplars, 13 tests that presented moderately atypical exemplars and 24 tests that presented extremely atypical exemplars as defined by my coding protocol. Recall that tests were coded as extremely (moderately) atypical when more than half (50% or less) of the exemplar information was disconfirming in nature. Seven tests were left un-categorised due to insufficient information, and were excluded from this section of the analysis. The summary data for typicality across outcome variables is presented in Table 2.4. The difference between the three levels of typicality on the total generalisation effect size was not statistically reliable ($Q (2) < 1, ns$), indicating that exemplar information is generalised to the out-group at similar rates irrespective of typicality.

<table>
<thead>
<tr>
<th>Table 2.4</th>
<th>Summary Data for Exemplar Typicality Across all Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$ (N%)</td>
</tr>
<tr>
<td>Total</td>
<td>51 (77)</td>
</tr>
<tr>
<td>Stereotypical</td>
<td>7 (14)</td>
</tr>
<tr>
<td>Moderately atypical</td>
<td>13 (25)</td>
</tr>
<tr>
<td>Extremely atypical</td>
<td>24 (47)</td>
</tr>
<tr>
<td>Stereotypicity</td>
<td>30 (45)</td>
</tr>
<tr>
<td>Moderately atypical</td>
<td>13 (20)</td>
</tr>
<tr>
<td>Extremely atypical</td>
<td>13 (20)</td>
</tr>
<tr>
<td>Dispersion</td>
<td>11 (17)</td>
</tr>
<tr>
<td>Moderately atypical</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Extremely atypical</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Prejudice</td>
<td>25 (38)</td>
</tr>
<tr>
<td>Stereotypical</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Moderately atypical</td>
<td>11 (17)</td>
</tr>
<tr>
<td>Extremely atypical</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

**Notes.** $N =$ number of tests; $N\% =$ percentage of the 66 total tests; $SE =$ standard error of $g$; $95\%CI =$ 95% confidence interval of $g$; $Z =$ the standard score; $p =$ probability of $g$.

The analysis of typicality across each of the outcome measures revealed that for cognitive outcome measures (stereotypicality and dispersion) extremely atypical exemplars were more effective at producing generalisation than moderately atypical exemplars. The
extreme-moderate difference on stereotypicality measures did not reach standard levels of significance, $Q (1) = 1.95, p = .16$, and was non-significant on the dispersion measures, $Q (1) < 1$. It is worthwhile to note that the dispersion measures consisted of a small sample set and should be interpreted with caution. As an overall pattern, I interpret the slight advantage of extremely atypical exemplars on cognitive measures as consistent with Rothbart’s (1981) conversion model.

On the affective outcome measure of prejudice, extremely atypical exemplars did not generalise to the out-group, although this was obtained from a small sample of tests ($n = 2$). The difference between the three levels of exemplar typicality on this outcome measure was non-significant, $Q (2) < 1$. Altogether, measures of prejudice followed a similar pattern of results as the cognitive outcome measures, although the small sample size limits the reliability of these findings.

In summary, the moderating variable of exemplar typicality displayed a similar pattern of results across all of the outcome measures. These findings are in contrast to Hewstone (1994), Brown and Hewstone (2005), and Paolini (2001) who concluded that moderately atypical exemplars are generalised more strongly to the group than extremely atypical exemplars. The reasons for this difference are explored more thoroughly in the concluding comments section.

**Exemplar and out-group valence.** Exemplar valence was assessed with 30 tests using positive exemplar information ($M_g = .46, r = .22, p < .001$), 10 tests using negative exemplar information ($M_g = .57, r = .27, p < .001$), and 10 tests using a mixture of positive and negative information ($M_g = .76, r = .36, p < .001$). One test was excluded due to insufficient information. Table 2.5 provides the summary data. Full comparisons were not always possible within each measure type due to insufficient data, but the pattern within measures typically mapped the overall effects. There was a tendency for a superior effect of mixed valence information, although the difference between the three levels of this moderating variable was statistically non-significant, $Q (2) < 1, ns$.

When focusing on the positive vs. negative difference, stereotypicality measures reflected a relative advantage of negative exemplar information over positive exemplar information, and this difference was marginally significant, $Q (1) = 3.33, p = .07$. This is inconsistent with the generalised appraisal model (Bodenhausen et al., 1995). Instead, it is consistent with Baumeister et al.’s (2001) review of valence asymmetry evidence. In line with Baumeister et al.’s parallel work in related areas, this meta-analytic evidence suggests that negative information is more strongly attended, processed and recalled than positive information, and as a consequence, has greater impact on judgments and behaviours.
Table 2.5
Summary Data for Exemplar Valence Across all Outcome Measures

<table>
<thead>
<tr>
<th></th>
<th>N (N%)</th>
<th>g</th>
<th>r</th>
<th>SE</th>
<th>95% CI</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>30 (45)</td>
<td>.46</td>
<td>.22</td>
<td>.07</td>
<td>.33/.59</td>
<td>7.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Negative</td>
<td>10 (15)</td>
<td>.57</td>
<td>.27</td>
<td>.12</td>
<td>.33/.80</td>
<td>4.75</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mixed</td>
<td>10 (15)</td>
<td>.76</td>
<td>.36</td>
<td>.16</td>
<td>.45/.92</td>
<td>4.80</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Stereotypicality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>14 (21)</td>
<td>.31</td>
<td>.15</td>
<td>.09</td>
<td>.13/.49</td>
<td>3.41</td>
<td>= .001</td>
</tr>
<tr>
<td>Negative</td>
<td>8 (12)</td>
<td>.58</td>
<td>.28</td>
<td>.12</td>
<td>.35/.81</td>
<td>4.95</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mixed</td>
<td>7 (11)</td>
<td>.82</td>
<td>.38</td>
<td>.22</td>
<td>.39/1.25</td>
<td>3.77</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Dispersion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>5 (8)</td>
<td>.41</td>
<td>.20</td>
<td>.32</td>
<td>-.21/1.03</td>
<td>1.30</td>
<td>= .19</td>
</tr>
<tr>
<td>Mixed</td>
<td>6 (9)</td>
<td>.71</td>
<td>.33</td>
<td>.16</td>
<td>.40/1.01</td>
<td>4.57</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Prejudice</td>
<td>25 (38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>20 (30)</td>
<td>.37</td>
<td>.18</td>
<td>.10</td>
<td>.19/1.66</td>
<td>3.92</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Notes. N = number of tests; N% = percentage of the 66 total tests; SE = standard error of g; 95% CI = 95% confidence interval of g; Z = the standard score; p = probability of g.
Different subscripts indicate marginally different effects, $Q\ (1) = 3.33, p = .07$.

Out-group valence was assessed with 18 tests using a positively valenced out-group ($M_g = .51, r = .25, p < .001$), 22 tests using a negatively valenced out-group ($M_g = .50, r = .24, p < .001$), and 11 tests using an out-group of neutral or mixed valence ($M_g = .66, r = .31, p < .001$). Table 2.6 provides the summary data.

Table 2.6
Summary Data for Out-Group Valence across all Outcome Measures

<table>
<thead>
<tr>
<th></th>
<th>N (N%)</th>
<th>g</th>
<th>r</th>
<th>SE</th>
<th>95% CI</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>18 (27)</td>
<td>.51</td>
<td>.25</td>
<td>.10</td>
<td>.32/.69</td>
<td>5.28</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Negative</td>
<td>22 (33)</td>
<td>.50</td>
<td>.24</td>
<td>.01</td>
<td>.35/.65</td>
<td>6.46</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mixed</td>
<td>11 (17)</td>
<td>.66</td>
<td>.31</td>
<td>.02</td>
<td>.39/.93</td>
<td>4.80</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Stereotypicality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>18 (27)</td>
<td>.34</td>
<td>.17</td>
<td>.10</td>
<td>.15/.54</td>
<td>3.46</td>
<td>= .001</td>
</tr>
<tr>
<td>Negative</td>
<td>6 (9)</td>
<td>.51</td>
<td>.25</td>
<td>.09</td>
<td>.32/.69</td>
<td>5.36</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mixed</td>
<td>6 (9)</td>
<td>.91</td>
<td>.41</td>
<td>.05</td>
<td>.47/1.34</td>
<td>4.11</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Dispersion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>6 (9)</td>
<td>.58</td>
<td>.28</td>
<td>.18</td>
<td>.23/93</td>
<td>3.21</td>
<td>= .001</td>
</tr>
<tr>
<td>Mixed</td>
<td>5 (8)</td>
<td>.58</td>
<td>.28</td>
<td>.27</td>
<td>.04/1.11</td>
<td>2.11</td>
<td>= .04</td>
</tr>
<tr>
<td>Prejudice</td>
<td>25 (38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>4 (6)</td>
<td>.01</td>
<td>.005</td>
<td>.27</td>
<td>-.52/.54</td>
<td>.05</td>
<td>ns</td>
</tr>
<tr>
<td>Negative</td>
<td>17 (26)</td>
<td>.46</td>
<td>.22</td>
<td>.08</td>
<td>.29/.62</td>
<td>5.42</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mixed</td>
<td>4 (6)</td>
<td>.76</td>
<td>.36</td>
<td>.19</td>
<td>.38/1.14</td>
<td>3.94</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Notes. N = number of tests; N% = percentage of the 66 total tests; SE = standard error of g; 95% CI = 95% confidence interval of g; Z = the standard score; p = probability of g.
Different subscripts indicate marginally different effects, $Q\ (2) = 5.19, p = .08$. 

There was no statistical difference between the three levels of out-group valence, \( Q(2) = 1.13, p = .57 \). However, out-groups of neutral or mixed valence displayed the strongest generalisation effect size. The advantage of mixed groups is consistent with Park and Hastie’s (1987) research. Park and Hastie found that participants were more likely to generalise traits to a group that was perceived as high in variability, than to a group perceived as low in variability. According to Park and Hastie, this occurs because disconfirming information is more readily included within a group that is already perceived as variable.

To further understand valence effects, I investigated exemplar valence at each level of out-group valence (see Table 2.7). Interestingly, I found that when there was a match between the valence of the exemplar and the valence of the out-group, generalisation was of a similar size (positive exemplar-positive out-group, \( M_g = .53, r = .26, p = .001 \); negative exemplar-negative out-group, \( M_g = .55, r = .27, p = .003 \)). Moreover, the match between mixed exemplar and mixed out-group produced the largest effect size (\( M_g = .88, r = .40, p < .001 \)). However, when the valence of the exemplar information disconfirmed the valence of the out-group, there was an advantage of negative exemplar information over positive exemplar information. That is, a negative exemplar disconfirming a positive stereotype generalised more (\( M_g = .57, r = .27, p < .001 \)) than a positive exemplar disconfirming a negative stereotype (\( M_g = .44, r = .21, p < .001 \)). This seems to be consistent with Baumeister et al.’s (2001) general framework, as participants were more willing to change their views of positive groups by attending to negative exemplar information than to improve negative out-group views after exposure to positive exemplar information.

<table>
<thead>
<tr>
<th>Table 2.7</th>
<th>Summary Data for Exemplar Valence across all Outcome Measures</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( N )</td>
<td>g</td>
<td>r</td>
<td>SE</td>
<td>95% CI</td>
</tr>
<tr>
<td>Positive out-group</td>
<td></td>
<td>18 (27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive exemplar</td>
<td>9 (14)</td>
<td>.53</td>
<td>.26</td>
<td>.15</td>
<td>.23/.83</td>
</tr>
<tr>
<td></td>
<td>Negative exemplar</td>
<td>4 (6)</td>
<td>.57</td>
<td>.27</td>
<td>.17</td>
<td>.24/.91</td>
</tr>
<tr>
<td></td>
<td>Mixed exemplar</td>
<td>4 (6)</td>
<td>.68</td>
<td>.32</td>
<td>.34</td>
<td>.02/1.34</td>
</tr>
<tr>
<td>Negative out-group</td>
<td></td>
<td>22 (33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive exemplar</td>
<td>16 (24)</td>
<td>.44</td>
<td>.21</td>
<td>.007</td>
<td>.29/.58</td>
</tr>
<tr>
<td></td>
<td>Negative exemplar</td>
<td>6 (9)</td>
<td>.55</td>
<td>.27</td>
<td>.19</td>
<td>.19/.92</td>
</tr>
<tr>
<td>Mixed out-group</td>
<td></td>
<td>11 (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive exemplar</td>
<td>5 (8)</td>
<td>.42</td>
<td>.21</td>
<td>.21</td>
<td>.01/.82</td>
</tr>
<tr>
<td></td>
<td>Mixed exemplar</td>
<td>6 (9)</td>
<td>.88</td>
<td>.40</td>
<td>.14</td>
<td>.61/1.15</td>
</tr>
</tbody>
</table>

Notes. \( N \) = number of tests; \( N\% \) = percentage of the 66 total tests; SE = standard error of \( g \); 95\% CI = 95\% confidence interval of \( g \); Z = the standard score; p = probability of \( g \).

To summarise, what has emerged from this portion of the analysis is a general advantage for negative exemplar information and negative out-groups over positive exemplar information and positive out-groups. This finding is consistent with Paolini et al.’s (under
review) finding for contact literature that people are naturally biased towards worsening group judgments, through inclusion of negative information and experiences, than towards improving group judgments. However, across analyses, there was an overall generalisation advantage of mixed valence exemplars from an out-group of a mixed valence. I will come back to these results more extensively in the concluding comments section.

**Out-group Type**

There were 28 tests that used affective out-groups and 23 tests that used cognitive out-groups. The summary data are shown in Table 2.8. The mean effect size for affective groups was significant ($M_g = .47, p < .001$) as was the mean effect size for cognitive groups ($M_g = .64, p < .001$). The greater generalisation for cognitive groups over affective groups did not reach standard significance levels ($Q(1) = 2.19, p = .14$).

<table>
<thead>
<tr>
<th></th>
<th>N (N%)</th>
<th>g</th>
<th>r</th>
<th>SE</th>
<th>95% CI</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>51 (77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Affective</strong></td>
<td>28 (42)</td>
<td>.47</td>
<td>.23</td>
<td>.06</td>
<td>.35/.59</td>
<td>7.54</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Cognitive</strong></td>
<td>23 (35)</td>
<td>.64</td>
<td>.30</td>
<td>.10</td>
<td>.45/.83</td>
<td>6.59</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Stereotypicity</strong></td>
<td>30 (45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>10 (15)</td>
<td>.50</td>
<td>.24</td>
<td>.08</td>
<td>.34/.66</td>
<td>6.12</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Cognitive</td>
<td>20 (30)</td>
<td>.50</td>
<td>.24</td>
<td>.12</td>
<td>.26/.73</td>
<td>4.16</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Dispersion</strong></td>
<td>11 (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>10 (15)</td>
<td>.68</td>
<td>.32</td>
<td>.14</td>
<td>.42/.94</td>
<td>5.04</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Cognitive</td>
<td>25 (38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>20 (30)</td>
<td>.48</td>
<td>.23</td>
<td>.08</td>
<td>.33/.64</td>
<td>6.10</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Cognitive</td>
<td>5 (7)</td>
<td>.16</td>
<td>.08</td>
<td>.07</td>
<td>-.34/.67</td>
<td>&lt; 1</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 2.8

Summary Data for Out-Group Type across all Outcome Measures

Notes. N = number of tests; N% = percentage of the 66 total tests; SE = standard error of g; 95% CI = 95% confidence interval of g; Z = the standard score; p = probability of g.

Overall, these results are consistent with my prediction derived from Tropp and Pettigrew (2005) and Paolini et al. (2007) that, due to the cognitive basis of the impression formation paradigm, cognitive out-groups should be more open to change than affective out-groups. The advantage of generalisation towards cognitive out-groups also supports Bodenhausen et al.’s (1995) claim that less complex cognitive-based groups are more open to change than more complex, variable affective-based social groups.

Focusing on the added complexity of measure type, affective measures seem to capture more generalisation within impression formation studies investigating affective out-groups than cognitive out-groups (although the small sample of cognitive tests limited the ability of the random effects approach to identify any significant difference between these two levels, $Q(1) = 1.37, p = .24$). Altogether, this pattern of results is consistent with Paolini et al.’s (2007) general hypothesis of greater generalisation when a match exists between exemplar experience and out-
group type. The cognitive experience of receiving information changed views of cognitive groups; while changes in prejudice were more easily detected with an affective than cognitive group.

Proposition Two: Concluding Comments

The second proposition asked: What conditions, or factors, maximise the generalisation of exemplar information? Through the analysis of four different moderating variables, several interesting trends emerged. I summarise these themes below.

The meta-regression conducted for sample size revealed support for the law of large numbers and a general benefit of increased size of disconfirming exemplars, supporting the bookkeeping model of stereotype change (Rothbart, 1981). A closer look at the moderating effect of measure type, however, revealed that this effect was driven by cognitive outcome measures, not affective outcome measures. This finding is consistent with the general trend that emerged in Proposition One, of an advantage for cognitive outcome measures over affective outcome measures. This cognitive-advantage trend is predicable opposite to that identified by Tropp and Pettigrew’s (2005) meta-regression. Tropp and Pettigrew found that affective measures (in their research separately coded as measuring ‘emotions’ or ‘favourability’) were able to detect large contact effects, while cognitive measures (in their research coded separately as measuring ‘stereotypes’ or ‘beliefs’) detected only small contact effects. However, this discrepancy between my results and Tropp and Pettigrew can be reconciled invoking Paolini et al.’s (2007) prediction of a match between exemplar experience and outcome measure type. In both Tropp and Pettigrew and my analysis generalisation was maximised under conditions of match. That is, in contact research there is a match between an affective experience with affective measures, while in impression formation research there is a match between a cognitive experience and cognitive measures. To summarise these analyses, as the sample of total exemplars and disconfirming exemplars provided to impression formation participants increases, generalisation becomes larger on cognitive outcome measures.

The moderating role of exemplar typicality provided a fair test of the predictions of typicality by avoiding the confound associated with the concentrated/dispersed paradigm between sample size and typicality. Although the overall analysis identified that each level of typicality (stereotypical, moderately and extremely atypical) was significantly generalised to the out-group, there was a pronounced advantage of extremely atypical exemplars on cognitive outcome measures, in particular for dispersion measures. This conclusion is consistent with Rothbart’s (1981) conversion model but contradicts Hewstone (1994) and Paolini’s (2001) narrative reviews of the concentrated-dispersed literature.

Hewstone (1994) indicated that, although, there was support for each of the cognitive models, he argued for a distinct advantage of the subtype model, whereby moderately atypical
exemplars are more effective in promoting generalisation than extremely atypical exemplars. The conclusions made by Hewstone were based, however, on a different sample of studies with a distinctively different operationalisation of exemplar typicality. Specifically, he reviewed studies that used the concentrated-dispersed paradigm where typicality is manipulated using a small number of extremely atypical exemplars or a larger number of moderately atypical exemplars. To account for result discrepancies between Hewstone’s review and the present review, I conducted a mini meta-analysis on the concentrated-dispersed literature\(^4\). Studies were coded such that the participants provided with the larger sample size of moderately atypical exemplars (dispersed condition) were treated as comparison group one; while the participants provided with the smaller sample of extremely atypical exemplars (concentrated condition) were treated as comparison group two. Consistent with Hewstone’s original conclusions, a positive effect size emerged, \((M_g = .65, r = .31, p < .001)\)\(^5\). This finding indicates greater generalisation amongst the larger, moderately atypical sample (dispersed condition). This additional meta-analysis on the concentrated-dispersed literature demonstrates that the conclusions drawn by Hewstone (1994) are applicable within the specific literature considered in his review. However, in light of my disconfirming sample size meta-regression, it seems reasonable to argue that the advantage of dispersed information reflects a disconfirming sample size effect rather than, as traditionally argued, a typicality effect.

The analysis of out-group valence revealed an advantage for groups of mixed valence. This finding is consistent with Park and Hastie’s (1987) work on perceived group variability (cf. Hantzi, 1995). Park and Hastie manipulated group variability and demonstrated that when a group is perceived as high in variability, disconfirming information is more readily generalised to the out-group than when a group is perceived as low in variability. In line with the conclusions of Park and Hastie, it seems important for social policy makers to encourage people to perceive groups as highly variable as this perception is more likely to facilitate generalisation of disconfirming out-group information.

The complex interplay between exemplar valence and out-group valence was also considered in this second proposition. A general trend emerged showing that negative exemplar information was more likely to be generalised to an out-group than positive exemplar information. This finding is consistent with a general valence asymmetry as discussed in Baumeister et al. (2001), who reviewed many domains of psychology and concluded that, possibly because of its evolutionary significance, negative information is more strongly processed and included in judgments than positive information. Baumeister et al.’s evolutionary

---

\(^4\) This meta-analysis included 21 tests, with 1,092 participants. The selected studies met the general selection criteria adopted for Meta-Analysis Study 1, with the exception of a no-information control group. In this mini meta-analysis all participants were presented with disconfirming exemplar information. See Appendix I for a table displaying the included studies and the coded variables.

\(^5\) The fail-safe \(n\) indicated that 539 null tests would be required to nullify the effect.
explanation, based on the generalisation advantage of negative information occurring across a variety of psychological phenomenon, reflects an evolutionary advantage of responding efficiently to stimuli that highlights a problem for the organism’s well-being. Within an intergroup perspective it has been argued that this advantage of negative information is due to the increased salience of the in-group-out-group distinction when exposed to negative out-group exemplars (see also, Paolini et al., under review).

Finally, the moderating variable of out-group type was considered. It was found that affective groups were more likely to demonstrate generalisation on affective outcome measures, while cognitive groups demonstrated an especially strong effect of generalisation on the cognitive measure of perceived out-group dispersion. Overall, these findings reiterate the conclusions by Paolini et al. (2007) that outcome measures should be matched to the participant experience and to the affective vs. cognitive nature of the out-group. More generally, these results highlight that generalisation effects can be enhanced through aligning type of experience with the out-group member, out-group type and outcome measure type.

General Discussion

Default Exemplar Inclusion and Beyond

This meta-analysis demonstrated that participants provided with out-group exemplar information readily generalise the information to the out-group. Hence, it can be argued that when information is presented, it is incorporated within the out-group judgment in a default manner (see also, Paolini, 2001). It is important to recognise that within this set of studies participants were clearly encouraged to engage in member-to-group generalisation. There were no additional tasks required of the participants beyond that of forming an impression of out-group members through the presented information. This simple methodology is likely to facilitate conversational rules of “relevance” (Grice, 1975). That is, these participants might have interpreted the researcher’s instructions and procedure as a request to regard the out-group information as relevant, valid and a useful basis for their subsequent judgment. As a result, participants make use of the exemplar information and include it in their out-group judgment. Interestingly, negative exemplar information was more readily included in out-group judgments than positive exemplar information. This was interpreted through the framework of Baumeister et al.’s (2001) “bad is better than good” interpretation of psychological phenomena. Hence, negative out-group exemplars can be added to Baumeister et al.’s lengthy list of everyday events that are more thoroughly attended to when presented in negative terms rather than positive terms.

I believe that Schwarz and Bless’ (1992; Bless & Schwarz, 1998) inclusion-exclusion model might provide a useful umbrella to further reconcile these diverse conclusions. As part of their model, Schwarz and Bless argue that the “default operation is to include the (exemplar)
information” (p. 169). This default inclusion was supported through the results of Proposition One, where the exemplar information was found to generalise to out-group judgements across each of the different outcome measures.

Schwarz and Bless (1992) also contend that the direction and extent of generalisation is contingent on contextual cues, such as varying levels of exemplar typicality (for evidence of contextual influences in out-group judgments, see Kervyn, Yzerbyt, Demoulin, & Judd, 2008). Adding to Schwarz and Bless’ contextual cues, I propose that different meta-cognitive processes may operate as a cue to facilitate cognitive inclusion or exclusion of the exemplar information. This idea will be thoroughly investigated in the following chapter, where I will investigate impression formation studies that have included additional meta-cognitive task cues. In particular, I will focus on cues that encourage participants to ‘include’ or ‘exclude’ exemplar information within their out-group judgment. A goal of the following empirical studies will be to focus on meta-cognitive cues that facilitate the inclusion of exemplar information, and subsequently expand our understanding of the diverse ways people compile their out-group judgments when faced with out-group member information.

**Updating the Previous Reviews**

The present meta-analysis paralleled the meta-analysis on contact by Pettigrew and Tropp (2000, 2006; Tropp & Pettigrew, 2004, 2005) and found similar levels of generalisation to that identified within the contact literature (r = .26 vs. r = .23, respectively). Similar to Tropp and Pettigrew (2005), I also found evidence of a moderating effect of measure type, however in the impression formation research there was an overall advantage of cognitive outcome measures (see Proposition One). In the analysis of Proposition Two, I found an advantage of affective outcome measures with research involving affective out-groups. In drawing this diverse evidence together, I have demonstrated a general benefit of matching cognitive groups to cognitive measures and affective groups to affective measures. Together with the meta-analyses by Tropp and Pettigrew, this provides a synthetic picture of the affective/cognitive interface across affective based phenomena (i.e., contact experiences) and cognitive based phenomena (i.e., impression formation).

This meta-analysis on the basic generalisation effects and several pertinent moderators also extends the two previous narrative reviews of impression formation research by Paolini (2001) and Hewstone (1994). By adopting a meta-analytic technique, I was able to provide a quantitative extension to Paolini’s review, allowing for greater precision in conclusions. Together, both this review and Paolini’s conclude that basic generalisation occurs as a default process.

Paolini’s review of moderating factors of the process of member-to-group generalisation highlighted a number of different themes, and identified many different variables
that influence generalisation. My analysis of the second proposition focused on a smaller set of moderating variables than the narrative review by Paolini. Moreover, because I focused on moderators that applied to all impression formation studies, rather than to a smaller subset, the results of Proposition Two are not immediately comparable to those of Paolini.

Through the analysis of typicality, I detected an advantage of extremely atypical exemplars across the different outcome measures. This goes against the conclusions drawn by both Hewstone (1994) and Paolini (2001), who argued for an advantage of moderately atypical exemplars in member-to-group generalisation research. As noted before, the conclusions made by Hewstone were based on the concentrated-dispersed literature. I do not dispute these conclusions. In fact, I also found support for the dispersed condition when meta-analytically reviewing the same concentrated-dispersed data. Through a separate analysis of sample size and exemplar typicality, however, I have argued that the evidence of concentrated-dispersed paradigms support the bookkeeping model’s sample size effect rather than the subtype model’s argument for an advantage of moderately atypical exemplars. This is an important extension on Hewstone’s review.

**Issues for Future Research**

There is now meta-analytical evidence that generalisation occurs at similar levels after exposure to out-group exemplars from either a contact-based paradigm (e.g., Pettigrew & Tropp, 2006) or an impression formation paradigm (the present analysis). In Chapter 4, I will propose to combine these two different approaches to member-to-group generalisation within a new integrative paradigm. Such a paradigm will rely on participants’ direct contact experiences with out-group members within a modified impression formation paradigm.

**Investigating a variety of outcome measures.** Throughout this chapter, I have argued for a benefit of matching outcome measures to the cognitive-affective basis of the out-group experience. At times, the analysis of the moderating role of the outcome measures was constrained by a small sample size, particularly the investigation of perceived out-group dispersion. It is important for future research to include a variety of outcome measures to confirm that cognitive measures are more appropriate to assess cognitive approaches to generalisation (i.e., impression formation) and that affective measures are more appropriate to assess affective approaches to generalisation (i.e., contact literature).

**Automaticity of out-group judgments.** The default approach to generalisation identified here suggests that generalisation of exemplar information to out-groups may be automatic. This suggestion builds on stereotype activation research using implicit outcome measures showing that stereotypes are, at least in part, activated automatically (Banaji & Greenwald, 1994; Banaji & Greenwald, 1995; Banaji, Hardin, & Rothman, 1993; Bargh, 1994; Devine, 1989; Dovidio, Evans, & Tyler, 1986; Fiske, 1989; Geis, 1993; Gilbert & Hixon, 1991;
Perdue & Gurtman, 1990; Pratto & Bargh, 1991). It is reasonable to assume that if a stereotype is automatically activated then generalisation may also be automatic.

Within the meta-analysis, implicit measures were not included as a stand-alone outcome measures as these measures are (still) very infrequently included within the impression formation paradigm. To test the idea of automaticity, I will include implicit stereotyping measures in Studies 1 and 4. Intergroup generalisation evidence points towards automatic generalisation. For example, Dasgupta and Greenwald (2001) found that White American participants exposed to a positive Black American exemplar were less negative on measures that assessed for an automatic bias (for an applied example, see Maio, Haddock, Watt, & Hewstone, 2008). Thus building on this newer approach to outcome measures, I will also include an automatic assessment of stereotyping within this research.

Coda

From this meta-analysis a number of points should now be clear. Firstly, member-to-group generalisation is a default process, possibly even an automatic process, which is best viewed through the lens of Schwarz and Bless’ (1992) inclusion-exclusion model. Secondly, the next step for this research should focus on understanding the meta-cognitive cues that facilitate inclusion or exclusion of member information. Thirdly, exposure to exemplar information can improve out-group attitudes across a variety of outcome measures. Finally, member-to-group generalisation is maximised when the cognitive-affective out-group experience is matched to the cognitive-affective outcome measure.
CHAPTER 3

MEMBER-TO-GROUP GENERALISATION AND META-COGNITIONS: META-ANALYTIC REVIEW (2)

Having reviewed the member-to-group generalisation literature that adopted an impression formation paradigm, the stage is now set to extend this research into a newer area of research: meta-cognition. Meta-cognition is broadly defined as people’s cognitions about cognition (Rucker & Petty, 2004) and as noted by Jost et al. (1998) “metacognitive phenomena pervade nearly every aspect of personal and interpersonal life” (p. 151). An example of metacognition in action is when we decide to write down the date of an appointment because we believe that we may forget it. This familiar act reflects the fact that we have a lay theory, or meta-cognition, about our cognition, in this instance our ability to remember, or not, the appointment. Despite the familiarity of such an example, it is only in the last twenty years that psychologists have begun a serious investigation on the impact of metacognition on social judgments (for a review, see Jost et al., 1998). Jost et al. (1998) argued for an expanded approach to meta-cognition within social psychological research, suggesting that “social psychology has much to offer theory and research on metacognition” (p. 151).

This thesis responds to Jost et al.’s (1998) call for an expanded approach to meta-cognitive research by applying it to member-to-group generalisation, thus moving from a strictly cognitive analysis of generalisation to a meta-cognitive analysis. In so doing, this thesis engenders a broader perspective on the process of member-to-group generalisation (see also, Paolini et al., 2009). In particular, throughout this thesis I will look at the meta-cognitive process of fluency. Fluency refers to the additional information associated with the experience of information processing or information retrieval. In essence, fluency informs judgments and encompasses the degree of activation, speed, and effort involved in processing and/or retrieving stimuli (see also, Winkielman, Schwarz, Fazendeiro, & Reber, 2003). In general, judgments associated with high fluency tend to be inline with implications of the accessible content but tend to be opposite to (or less like) the accessible content when fluency is low (Sanna, Schwarz, & Stocker, 2002; Schwarz et al., 1991; for a review see Schwarz, 1998).

There is limited explicit evidence to support the involvement of meta-cognition in member-to-group generalisation. However, this does not mean that the literature is void of examples. There are several studies that have investigated meta-cognition but not framed their research in these terms (e.g., Batson et al., 1997; Galinsky & Ku, 2004). A notable exception is
the work by Paolini et al. (2009). Paolini et al. (2009) used a standard impression formation paradigm in which participants were provided with a written profile of an individual group member described in stereotypical or counterstereotypical terms and were then asked to judge the group as a whole. Across three studies, Paolini and colleagues manipulated people’s motivation to monitor member-to-group generalisation by making some participants accountable for their group judgment (high level of motivation) and other participants non-accountable (low motivation). In Studies 2 and 3, Paolini et al. also manipulated the perceived validity of the member information by providing some participants with standard information (high level of perceived validity) and having other participants think they had received additional subliminal, thus potentially unreliable, member information (low perceived validity). Specifically, Paolini et al. conveyed information to the low perceived validity participants that the exemplar information was untrustworthy, that is, these participants were provided with information about the information, or meta-information. In this thesis, meta-information is defined as ‘information about information’.

Paolini et al. (2009) found that when participants were not focused on the accuracy of their judgments (i.e., low motivation) and the exemplar had high perceived validity, changes in group stereotyping reflected the accessible member’s information. However, when participants had increased motivation to monitor the judgment process or when the exemplar information had low perceived validity, the available member’s information was excluded from the group judgments. In essence, Paolini et al. found that when meta-cognitive cues signal that the exemplar information is satisfactory, then generalisation is facilitated. On the other hand, when meta-cognitive cues signal that the exemplar information is poor, then generalisation is inhibited. This suggests that meta-cognitive processes are able to interrupt the default process of generalisation (see Chapter 2).

Two important concepts emerge from Paolini et al.’s (2009) research. Firstly, meta-information was identified as an important moderator of member-to-group generalisation. Specifically, Paolini et al. argued that the direction of the meta-information can facilitate either inclusion or exclusion of the exemplar information. Secondly, Paolini argued that the source of the meta-information might also moderate the extent to which people generalise. In particular, Paolini et al. argued that the source of the meta-information may be external to the participant and fall “under the direct and exclusive control of the experimenter (p. 684)” or may be internal to the participants and fall “under the realm of the decision maker (p. 684)”. I will follow up on both the direction and source of the meta-information by structuring my review of meta-cognitive impression formation literature around these two concepts.

In the next section of this chapter, I will review the fluency literature and impression formation literature that has implicitly or explicitly considered the direction of meta-information (i.e., inclusion vs. exclusion) and the source of the meta-information (i.e., internal
Firstly, I will start by discussing the fluency research that has been applied to the broader domain of social judgments. Then, I will focus on the impression formation research that has investigated the source and direction of meta-information in member-to-group generalisation. I have decided to investigate the effects that the meta-cognitive process of fluency has on member-to-group generalisation, because I believe that this research will further inform social cognitive research, and similarly enhance our understanding of fluency.

Basic Meta-Cognitive Concepts Applicable to Member-to-Group Generalisation

Fluency is a broad, encompassing label referring to the ease of processing and/or retrieval of information used for subsequent judgments. In essence, this meta-cognitive process provides additional information to the perceiver about the actual process of accessing the stimulus material. Throughout this thesis, I will argue that fluency meta-information forms an important part of formulating an out-group judgment. To help achieve this, I will now review the fluency research that has assessed social judgments after the retrieval of out-group information.

Information vs. Meta-Information

From a meta-cognitive perspective, judgements are influenced by two types of information: (1) judgment relevant information and, (2) the meta-information associated with this information (Caruso, 2008; Winkielman & Cacioppo, 2001; for reviews, see Reber, Schwarz, & Winkielman, 2004; Schwarz, 1998). Schwarz and his colleagues investigated the role of meta-information in subsequent judgments (Schwarz et al., 1991). For this, they introduced in the literature the few-versus-many paradigm. Within this paradigm, participants are asked to retrieve from memory either a few or many different pieces of information about a subject area and then are asked to complete judgments about the subject area. For example, Schwarz et al. (1991, Study 1) asked participants to retrieve 6 (few condition) or 12 (many condition) examples of their own past assertive behaviour. Following retrieval, all participants were asked to judge their own level of assertiveness.

Schwarz et al. (1991) predicted that if participants were relying on the available information then assertiveness judgments should be higher in the many-exemplar condition. Alternatively, if participants were relying on the retrieval meta-information then assertiveness judgments should be higher in the few-exemplar condition. This is because the meta-information associated with the retrieval task (i.e., easy vs. difficult) was expected to inform the participant’s judgment, whereby participants in the easy (difficult) condition infer that their memory must contain many (few) additional examples. In support of a retrieval fluency interpretation, Schwarz et al.’s participants in the few-condition evaluated themselves as more assertive than participants in the many-condition.
Dijksterhuis, Macrae, and Haddock (1999) extended the few-versus-many paradigm to the investigation of social groups. Dijksterhuis et al. built on research showing that prejudiced people find it easy to generate negative stereotypical material (e.g., Augoustinos, Ahrens, & Innes, 1994; Kawakami, Dion, & Dovidio, 1998; Lepore & Brown, 1997; Locke, MacLeod, & Walker, 1994), and proposed that retrieval fluency effects should be moderated by individuals’ level of prejudice. Hence, they first tested participants’ levels of prejudice toward women using the Modern Sexism Scale (Swim, Aikin, Hall, & Hunter, 1995). After completing unrelated filler tasks, participants were asked to generate either three or eight traits that men and women differ upon just before describing a female secretary and her lifestyle. The open ended descriptions of the female secretary were rated for stereotypicality by two judges who were blind to conditions.

Consistent with a retrieval fluency account, Dijksterhuis et al. (1999) found a retrieval fluency effect for low prejudice participants. Low prejudice participants generating three male vs. female traits were less stereotypical than participants generating eight traits. On the other hand, high prejudice participants did not differ in their exemplar descriptions as a function of the number of traits recalled. Hence, high prejudice participants who were expected to find it easy to produce stereotypical material failed to demonstrate any impact of the meta-cognitive information. In addition, a manipulation check confirmed that in both the few and many condition, the high prejudiced participants did not display any discrepancy in self-reported difficulty in retrieving three or eight traits, suggesting that they may have more examples accessible in memory that reduce the differential appreciation of retrieval ease. As a result, amongst participants low in prejudice, the meta-information associated with the few-exemplar condition indicated that the task was easy, while for those in the many-exemplar condition the meta-information indicated that the task was hard. On the other hand, due to the high accessibility of available examples, the meta-information associated with the task amongst participants high in prejudice indicated that the task was easy irrespective of the number of examples they retrieved.

Dijksterhuis et al.’s (1999) work has the merit of demonstrating that stereotype relevant judgements are built from two sources of information: the information and the meta-information. However, Dijksterhuis et al.’s paradigm is not ideal for use in the empirical investigation of retrieval fluency and member-to-group generalisation. The most obvious limitation is that this research looks at a different kind of generalisation: Group-to-member, rather than member-to-group generalisation. From early models of social categorisation (Rothbart & John, 1985), we know that these two forms of generalisation are not equal. In addition, while Dijksterhuis et al.’s outcome measure has been employed in a number of studies (e.g., Dijksterhuis & van Knippenberg, 1996; 1998; Macrae, Stangor, & Milne, 1994), there is some discontinuity between the retrieved content (gender-differing traits) and the outcome...
measure (description of a female secretary). Dijksterhuis et al. argue that using unrelated measures is evidence of the generalisability of fluency findings. In my research, I will use a more direct approach. I will adopt a closer matching of the retrieved material to subsequent judgments (e.g., Schwarz et al., 1991), to offer a more precise assessment of the relationship between exemplar content and judgment decision. In addition, in Study 1, I will explicitly test for the generalisation of retrieval fluency effects to exemplar-unrelated dimensions.

From Paolini et al.’s (2009) research we know that (external) meta-information can facilitate the cognitive inclusion or exclusion of the exemplar information in an impression formation paradigm. What is missing from Dijksterhuis et al.’s (1999) research is clarification of the direction of the generalisation effect produced by the internal meta-information. As Dijksterhuis et al. did not include a baseline condition in their design, it is unclear if the meta-information leads to inclusion or exclusion from social judgments relative to the benchmark. Sanna, Kennedy, Change and Miceli (2009) investigated the role of meta-information in hindsight bias and included a control condition in their research. In Study 1, participants were asked to consider a recent college basketball game, in which Carolina defeated Duke. Participants were asked to generate either three reasons (easy condition), 12 reasons (difficult condition), or no reasons (control condition) indicating why Carolina won the game. Sanna et al. (2009) found that participants in the easy condition demonstrated greater hindsight bias compared to the control condition. Participants in the difficult condition demonstrated less hindsight bias compared to the control condition. This suggests that within the domain of hindsight bias, internal meta-information is included in the subsequent judgment when the retrieval task is perceived as easy, but is excluded from the subsequent judgment when the task is perceived as difficult. It is important to test if inclusion and exclusion of (internal) meta-information extends to the area of social judgments, and this will be achieved by including a control group in all studies. In so doing, I will be able to identify the direction of the effect of meta-information.

**Source of the Meta-Information**

Paolini et al. (2009) argued that the source of the meta-information might be internal or external. Internal meta-information comes from the decision maker, and includes considerations surrounding the ease with which judgment relevant information comes to mind. This ‘internal’ meta-information would serve as a “basis to infer whether the information is representative of the group or not and thus whether it can be relied on when judging the group” (Paolini et al., 2009, p. 684). External meta-information, on the other hand, comes from information extraneous to the participant, and includes experimenter controlled variables. For example, some might require the participant to engage in some additional task, such as Maurer, Park and Rothbart (1995) who required participants to physically sort the presented exemplars into
different sub-groups. It is worthwhile to note that within an impression formation paradigm, the exemplar information is always presented ‘externally’ by the experimenter. This implies that any manipulation of internal meta-information must also be injected artificially by the experimenter, making the investigation of internal meta-information using the impression formation paradigm a sub-optimal approach.

Interestingly, the internal/external distinction between meta-information cues maps onto two distinct areas of fluency research: (1) retrieval fluency research, which typically deals with internal meta-information cues, and (2) processing fluency research, which typically deals with external meta-information cues (Hertwig, Herzog, Schooler, & Reimer, 2008; Reber et al., 2004; Winkielman et al., 2003). From this perspective, retrieval fluency includes the ease with which internal mental operations concerned with retrieving information from memory are conducted (Schwarz, 1998); whereas processing fluency encapsulates the ease with which mental operations concerned with understanding an external stimulus are conducted (Reber et al., 2004, see also, Whittlesea, 1993; Winkielman et al., 2003). I elaborate on this distinction below.

**Internal meta-information.** Fluency research manipulating internal cues has typically adopted the few-versus-many paradigm and focused on retrieval fluency. Retrieval fluency has been investigated across a variety of judgment domains. An expanded list would include: attitude strength (Haddock, Rothman, Reber, & Schwarz, 1999), self judgments (Schwarz et al., 1991), childhood memory (Winkielman, Schwarz, & Belli, 1998), hindsight bias (Sanna et al., 2002; Sanna et al., 2009), group-based judgments (e.g., Rothman & Hardin, 1997), judgments of health risks (Rothman & Schwarz, 1998), likelihood estimation (Wänke et al., 1995), interpersonal closeness (Broemer, 2001), self-doubt (Hermann, Leonardelli, & Arkin, 2002), persuasion (Tormala, Petty, & Briñol, 2002; Wänke, Bless, & Biller, 1996), and judgments of stereotypicality (Dijksterhuis et al., 1999). Generally, judgements tend to be consistent with the retrieved content when retrieval fluency is high. Together, the broad application and consistent findings of retrieval fluency research demonstrate the important role of this meta-cognitive tool in the judgment process. My research will extend our understanding of retrieval fluency by testing if this effect is applicable to member-to-group generalisation.

To my knowledge, Rothman and Hardin (1997, Study 1) carried out the only investigation to date of retrieval fluency research within a member-to-group framework. Rothman and Hardin adopted Schwarz et al.’s (1991) few-versus-many paradigm and asked participants to recall three (few) or six (many) stereotype unrelated behaviours (impolite actions) performed by members of the same sex (in-group condition) or opposite sex (out-group condition). Participants (52 male, 44 female) were then asked to judge the target group on the trait of politeness. Rothman and Hardin found a significant interaction between group type (in-group vs. out-group) and number of behaviours (three vs. six). Participants in the out-group
condition generalised more in the three behaviours condition than in the six behaviours condition, thus demonstrating a classic retrieval fluency effect. Participants in the in-group condition generalised more in the six behaviours condition than in the three behaviours condition, thus demonstrating a sample size effect. Rothman and Hardin concluded that this demonstrated that out-group judgments are based on retrieval fluency, while in-group judgments are based on retrieved content.

Although, Rothman and Hardin (1997) have the merit of having extended the use of the few-versus-many paradigm to the member-to-group generalisation domain, their extension remains sub-optimal. A problem with Rothman and Hardin’s method lays in the nature of their retrieval request. In their work, participants were not required to retrieve an exemplar; they were required to retrieve specific behaviours. The use of behaviours is potentially limiting in member-to-group generalisation research due to their transient nature. Evidence indicates that traits are more readily used than behaviours in making inferences (e.g., Gawronski, 2003). This means that the retrieval task of Rothman and Hardin may have unduly underestimated changes in out-group judgements and the magnitude of the retrieval fluency effect. My research improves on this limitation by asking participants to retrieve exemplars that display counterstereotypical traits, rather than counterstereotypical behaviours.

**External meta-information.** Fluency research relying on manipulation of external meta-cognitive cues has typically focused on processing fluency. Processing fluency refers to the characteristics surrounding the interpretation of the stimulus, such as speed and accuracy of processing (Rubin, Paolini, & Crisp, 2010). Processing fluency has been investigated in a range of areas such as image processing (Bornstein, Kale, & Cornell, 1990; Desimone, Miller, Chelazzi, & Lueschow, 1995; Halberstadt & Rhodes, 2003; Jacoby & Dallas, 1981), word processing (Whittlesea, 1993), migrant bias (Rubin et al., 2010) and stimulus familiarity (Alter & Oppenheimer, 2008). For example, Whittlesea (1993, Experiment 5) found that high processing fluency results in positive evaluations surrounding word judgment. They embedded words in an expected context (e.g., “stormy seas tossed the boat”) or in an unexpected context (e.g., “stormy seas tossed the lamp”). Participants judged the expected words (e.g. boat) more positively than unexpected words (e.g. lamp), suggesting that positive stimulus judgments are tied to the meta-information associated with high processing fluency (see also, Winkielman et al., 2003). In general, judgments associated with high processing fluency tend to result in a positive evaluation of the stimulus (Alter & Oppenheimer, 2008; Bless & Fielder, 1995; Carver & Scheirer, 1990; Ramachandran & Hirstein, 1999; Schwarz, 1990; Vallacher & Nowak, 1999).

Member-to-group generalisation studies that manipulated cognitive load can, in my view, be interpreted through the lens of processing fluency. That is, studies that have made it difficult for participants to process the exemplar stimuli through the use of a cognitive load
have, in essence, manipulated processing fluency. For example, Garcia-Marques and Mackie (1999, Study 3) manipulated both retrieval fluency and processing fluency (although their research was not explicitly framed in these terms). In their research, participants were given information regarding eight members of one occupational group (computer programmers, construction workers, child-care professionals, or disco-bouncers). Three of the eight were stereotype incongruent; three were stereotype congruent; and the remaining two were described in stereotype neutral terms. The fluency manipulations were implemented at either the impression or judgement phase: some participants rehearsed a 9-digit number while reading the sample information (low processing fluency), others while rating the group (low retrieval fluency), and some did not receive any competing task (high processing and retrieval fluency). A set of a priori contrasts revealed that generalisation on both group stereotypicality and dispersion measures was apparent only when processing fluency was high. Amongst participants low in processing and retrieval fluency there was no effect of the sample information on the group judgment. From this processing fluency interpretation, it is apparent that under conditions of high fluency people rely on the immediately accessible content and under conditions of low fluency people exclude such content from the judgment at stake.

**Summary**

Limited member-to-group generalisation research has been couched in terms of meta-cognition. The exception, by Paolini et al. (2009), flagged the importance of two key concepts: (1) the direction of the meta-information, and (2) the source of the meta-information cue. In the section above I elaborated on these concepts with a focus on fluency research. What has emerged from this section of the review is that the standard few-versus-many paradigm may require some adjustments (i.e., a control condition, a specific retrieval request matching the judgment, and a focus on traits) before being successfully applied to a member-to-group generalisation framework. To carry out a preliminary test of the relevance of meta-information direction and source for member-to-group generalisation, and prior to embarking on a pointed and new investigation, I will meta-analytically review the impression formation literature that has investigated (often implicitly) meta-information.

**Meta-Analysis Study 2**

I will now review impression formation studies that have investigated the direction and source of meta-information. Through this meta-analysis, I aim to identify seminal themes and findings surrounding meta-cognition and member-to-group generalisation. As noted earlier, the majority of these studies have not been framed in terms of a meta-cognitive investigation (for an exception, see Paolini et al., 2009). Despite this lack of a specific meta-cognitive focus, I would like to argue that these impression formation studies have nonetheless implemented a
variety of techniques to manipulate the direction and source of meta-information cues. This re-
interpretation of existing literature is in line with Jost et al.’s (1998) argument that much social
cognitive research can be considered social meta-cognitive research.

Impression formation studies were included in Meta-analysis Study 2 if the participants
were all presented with the same amount of out-group information, but one or more groups
received an additional meta-cognitive task cue. From a meta-cognitive standpoint, I believe that
because the exemplar information presented to participants does not vary, any difference
between conditions must be attributable to the presentation of the additional meta-cognitive
task cue. Although the exact cues presented vary across studies (for information about the exact
nature of the cue, see Table 3.1), I believe that they all share a (mostly unstated) goal: to change
people’s cognitions.

The direction of the meta-information cues will be coded by identifying whether the
meta-cognitive cue is expected to facilitate or inhibit member-to-group generalisation, that is,
whether the cue is expected to encourage inclusion or exclusion of the exemplar information
from the group judgment (also see Schwarz & Bless, 1992; Paolini et al., 2009). Cues that
facilitate generalisation should work in conjunction with the default process of generalisation,
resulting in a strong generalisation effect of the exemplar information to the group judgment.
Cues that inhibit generalisation should work against the default process of generalisation,
resulting in reduced generalisation of the exemplar information to the group judgment. Put
differently, if counterstereotypical exemplar information is presented with a facilitating meta-
cognitive task cue this should encourage stereotype change. On the other hand, if
counterstereotypical exemplar information is presented with an inhibiting meta-cognitive task
cue this should encourage stereotype maintenance. Hence, I expect a significant difference in
generalisation effects between cues that facilitate inclusion of the exemplar material (or
inclusive cues) and cues that inhibit generalisation and encourage exclusion of the exemplar
material (or exclusive cues).

This review will code the source of the meta-information cue presented to the
participants as either internal or external. However, as noted above, the impression formation
paradigm relies on information that is externally provided by the experimenter which
artificially injects internal meta-information cues into the experimental paradigm. Despite the
‘artificial’ nature of internal meta-information in impression formation studies, I will code the
meta-information as ‘internal’ when the meta-information is controlled by the exemplar
information (e.g., exemplar typicality) or controlled by the decision maker (e.g., retrieval
fluency). Meta-information will be coded as ‘external’ when it is controlled by the
experimenter (e.g., request to engage in additional task such as cognitive load or sorting tasks).
This analysis will provide a preliminary investigation into the impact of internal and external
### Table 3.1

**Coding of Tests Meeting Inclusion Criteria for Meta-Analytic Study 2**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study</th>
<th>$g$</th>
<th>$r$</th>
<th>$N$</th>
<th>Disc Size</th>
<th>DV</th>
<th>Source</th>
<th>Direct</th>
<th>Information about the cue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batson et al. (1997)</td>
<td>3</td>
<td>.67</td>
<td>.32</td>
<td>60</td>
<td>0</td>
<td>1</td>
<td>P</td>
<td>fac</td>
<td>Empathy instruction for exemplar</td>
</tr>
<tr>
<td>Bless &amp; Schwarz (1998)</td>
<td>2</td>
<td>-.16</td>
<td>-.07</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>P</td>
<td>fac</td>
<td>Question linking the exemplar to the group</td>
</tr>
<tr>
<td>Bless et al. (2001)</td>
<td>1</td>
<td>.43</td>
<td>.21</td>
<td>138</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>fac</td>
<td>Cue to view exemplar as typical representative</td>
</tr>
<tr>
<td>Carnaghi &amp; Yzerbyt (2007)</td>
<td>1</td>
<td>.79</td>
<td>.37</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>fac</td>
<td>Request to justify judgments to an equal opportunity organisation</td>
</tr>
<tr>
<td>Carnaghi &amp; Yzerbyt (2007)</td>
<td>2</td>
<td>.61</td>
<td>.29</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>fac</td>
<td>Request to justify judgments to an out-group</td>
</tr>
<tr>
<td>Galinsky &amp; Ku (2004)</td>
<td>1</td>
<td>.33</td>
<td>.16</td>
<td>67</td>
<td>0</td>
<td>1</td>
<td>S</td>
<td>fac</td>
<td>Instructed to adopt exemplar perspective</td>
</tr>
<tr>
<td>Galinsky &amp; Ku (2004)</td>
<td>2</td>
<td>1.15</td>
<td>.5</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>S</td>
<td>fac</td>
<td>Instructed to adopt exemplar perspective</td>
</tr>
<tr>
<td>Galinsky &amp; Ku (2004)</td>
<td>2</td>
<td>-.20</td>
<td>-.10</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>S</td>
<td>inh</td>
<td>Negative mood and perspective-taking</td>
</tr>
<tr>
<td>Guinote et al. (2002)</td>
<td>2</td>
<td>.86</td>
<td>.39</td>
<td>78</td>
<td>0</td>
<td>8</td>
<td>D</td>
<td>internal</td>
<td>Cue flagging sufficient information for judgment</td>
</tr>
<tr>
<td>Kunda &amp; Oleson (1995)</td>
<td>3</td>
<td>-1.02</td>
<td>-.45</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>inh</td>
<td>Cue to exclude exemplar information</td>
</tr>
<tr>
<td>Maurer et al. (1995)</td>
<td>1</td>
<td>.55</td>
<td>.27</td>
<td>65</td>
<td>3</td>
<td>16</td>
<td>S</td>
<td>external</td>
<td>Sort exemplars into groups of like-members</td>
</tr>
<tr>
<td>Maurer et al. (1995)*</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>3</td>
<td>16</td>
<td>D</td>
<td>external</td>
<td>Sort exemplars into groups of like-members</td>
</tr>
<tr>
<td>Nisbett et al. (1983)</td>
<td>2</td>
<td>.38</td>
<td>.19</td>
<td>116</td>
<td>0</td>
<td>3</td>
<td>S</td>
<td>internal</td>
<td>Task to promote out-group dispersion</td>
</tr>
<tr>
<td>Paolini et al. (2009)</td>
<td>1</td>
<td>.82</td>
<td>.38</td>
<td>30</td>
<td>0</td>
<td>1</td>
<td>S</td>
<td>external</td>
<td>Request to justify judgments to experimenter</td>
</tr>
<tr>
<td>Paolini et al. (2009)</td>
<td>2</td>
<td>.43</td>
<td>.21</td>
<td>38</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>external</td>
<td>Request to justify judgments to experimenter</td>
</tr>
<tr>
<td>Paolini et al. (2009)</td>
<td>2</td>
<td>-.47</td>
<td>-.23</td>
<td>40</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>inh</td>
<td>Cued about low perceived validity of information</td>
</tr>
<tr>
<td>Paolini et al. (2009)</td>
<td>3</td>
<td>.8</td>
<td>.37</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>external</td>
<td>Request to justify judgments to experimenter</td>
</tr>
<tr>
<td>Paolini et al. (2009)</td>
<td>3</td>
<td>-.52</td>
<td>-.25</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>inh</td>
<td>Cued about low perceived validity of information</td>
</tr>
<tr>
<td>Paolini et al. (2009)</td>
<td>3</td>
<td>.58</td>
<td>.28</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>D</td>
<td>external</td>
<td>Request to justify judgments to experimenter</td>
</tr>
<tr>
<td>Park et al. (1992)</td>
<td>2</td>
<td>.51</td>
<td>.25</td>
<td>80</td>
<td>9</td>
<td>16</td>
<td>S</td>
<td>external</td>
<td>Sort exemplars into groups of like-members</td>
</tr>
<tr>
<td>Park et al. (2001)</td>
<td>1</td>
<td>.35</td>
<td>.17</td>
<td>80</td>
<td>3</td>
<td>15</td>
<td>S</td>
<td>external</td>
<td>Sort exemplars into groups of like-members</td>
</tr>
<tr>
<td>Park et al. (2001)</td>
<td>1</td>
<td>.72</td>
<td>.34</td>
<td>80</td>
<td>3</td>
<td>15</td>
<td>D</td>
<td>external</td>
<td>Sort exemplars into groups of like-members</td>
</tr>
<tr>
<td>Yzerbyt et al. (1999)</td>
<td>2</td>
<td>1.02</td>
<td>.45</td>
<td>32</td>
<td>3</td>
<td>3</td>
<td>S</td>
<td>external</td>
<td>Cognitive load allowing for generalisation</td>
</tr>
<tr>
<td>Yzerbyt et al. (1999)</td>
<td>3</td>
<td>.12</td>
<td>.06</td>
<td>36</td>
<td>1</td>
<td>1</td>
<td>S</td>
<td>external</td>
<td>Cognitive load allowing for generalisation</td>
</tr>
</tbody>
</table>

*Notes. g = Hedges’ $g$ effect size for each sample; $N$ = Number of participants; Dis = Disconfirming sample size; Size = total out-group sample size; DV = outcome measure (S = Stereotypicality; D = Dispersion; P = Prejudice); Source = Cue source; Direct = Cue Direction (fac = facilitating; inh = inhibiting);* Not included in basic generalisation effect analysis.
meta-information cues and should lend support for Paolini et al. ’s (2009) distinction that both sources of meta-information are important moderators of member-to-group generalisation.

By coding the direction and source of the meta-information, Meta-analysis Study 2 will provide a direct extension of the previous narrative reviews on impression formation research (e.g., Hewstone, 1994; Paolini, 2001) and my earlier meta-analytic review.

Method

Studies were located and coded using the same search strategies identified in Meta-analysis Study 1. There was a minor change in the inclusion criteria for this meta-analysis. The tests still needed to provide participants with out-group member information and then measure out-group judgments, using one or more of the three outcome measures (group stereotypicality, group dispersion or group prejudice). However, a separate individual control condition was not required in this meta-analysis. Instead, all participants needed to receive the same out-group information, and participants in the experimental condition had to receive an additional task cue. A significant difference between the no-cue vs. cue conditions would reflect a clear role of the task cue in the process of generalisation.

The search criteria resulted in 24 included tests, representing a total of 1,601 participants. See Table 3.1 for a summary of the included tests and the coded variables. Of the included tests, 18 measured stereotypicality (75%), four measured dispersion (17%) and two measured out-group prejudice (8%). An effect size was coded as positive when the exemplar information was generalised to the out-group, that is, a positive effect size reflects out-group judgments that were in the direction of the exemplar information content. An effect size was coded as negative when the exemplar information was not generalised to the out-group, that is, a negative sign means the out-group judgment was in the opposite direction to the exemplar content.

Coding for the direction of the task cue was based on the rationale of the cue provided by the authors of the paper. For example, Paolini et al. (2009, Studies 2 & 3) provided some participants with a cue that was expected to flag that the exemplar information was low in perceived validity. Paolini et al. predicted that the low validity cue would inhibit generalisation of the exemplar information to the out-group. Hence, this cue was coded in the meta-analysis as an inhibiting (or exclusion) cue. Paolini et al. (Studies 1 – 3) also made some participants high in motivation (i.e., an accountability cue) that was expected to increase use of the exemplar information in the group judgment by encouraging systematic processing.

Maurer et al. (1995) measured both stereotypicality and dispersion. To avoid over representing the contribution of an individual participant in the overall generalisation effect, only one of these results were included. As their paper focused on stereotypicality the results for dispersion were not included in the general analysis, but the dispersion data are still maintained within the analysis of dispersion. This is the same approach I used in my earlier meta-analysis.
That is, Paolini et al. predicted that the high motivation cue would facilitate generalisation of the valid exemplar to the out-group. This cue was subsequently coded in the meta-analysis as a facilitating (or inclusion) cue. The coding for direction of the meta-information cue revealed that 20 tests (83%) used facilitating task cues and four tests (17%) used inhibiting task cues.

As noted above, all impression formation studies necessarily rely on a cue that is provided by the experimenter. Hence, for the purpose of this meta-analysis the source of the task cue (i.e., internal or external) was based on the way the participant was expected to react or attend to the cue. Cues were coded as internal when participants were encouraged to consider (or think specifically about) the intrinsic or inherent qualities of the exemplar material. In the example above by Paolini et al. (2009), participants provided with the low perceived validity cue were expected to attend to the exemplar information and recognise that it was invalid to include the information in the out-group judgment. Cues were coded as external when participants were asked to complete an additional task during, or after, information presentation. For example, Paolini et al. provided participants with an accountability cue (high level of motivation) where participants expected to justify their group decisions to other individuals (i.e., the experimenter). The key difference between internal and external meta-information cues is that internal cues require the participant to consider the sample information in a way that is somehow different to the non-cue condition, while external cues require the participant to do something extra that is extraneous to the sample information. Using this distinction, coding for the source of the cue revealed that 13 tests (54%) used external cues and 11 tests (46%) used internal cues.

Effect size calculation followed the rationale adopted in Meta-analysis Study 1 (see Chapter 2), Hedges’ $g$ was again used as the effect size and a random effects model was adopted. My use of a random effects model was supported by significant heterogeneity of the fixed effects model ($Q (22) = 41.16, p < .01$).

**Results**

**Basic Generalisation Effects**

A significant facilitating effect of meta-cognitive task cues was detected across studies ($M_g = .42, r = .21, p < .001$). This indicates that the provision of meta-cognitive cues typically leads to the inclusion of the exemplar information in the group judgments. According to Cohen (1977, 1988), this is a medium effect size. Cooper’s (1989) fail-safe $n$ assessment of publication bias found that 267 null tests would be required to nullify the effect. This effect, however, needs to be interpreted with caution because it partly reflects the disproportional representation of inclusive cue studies in the current literature (83% vs. 17%). As a consequence, I will now separately analyse effect size according to the direction of the meta-information cue.
Direction of the Meta-Information Cue

When separating studies on the direction of their meta-information cue, as predicted, I found that the inclusive meta-information cue facilitated generalisation ($M_g = .54$, $r = .26$, $p < .001$, see Table 3.2). That is, inclusive meta-information cues encourage changes in the group judgments that are in line with the content of the exemplar information. Within this variable level, each of the outcome measures was found to have similar effect sizes: stereotypicality ($M_g = .50$, $r = .24$, $p < .001$), dispersion ($M_g = .52$, $r = .25$, $p = .004$) and prejudice ($M_g = .51$, $r = .25$, $p = .03$). The difference between these outcome variables was non-significant (all $Qs < 1$). Due to the small number of tests measuring dispersion and prejudice, however, it is worthwhile to interpret these results with some caution.

As predicted, the tests that provided exclusive meta-information cues inhibited generalisation ($M_g = -.51$, $r = -.25$, $p = .01$, see Table 3.2). That is, in these studies the cue encouraged participants to revise their group judgments in a direction opposite to that implied by the content of the exemplar information. However, as only four tests provided exclusive cues some caution must be exerted in interpreting these results. With due caution, these results do suggest that under certain conditions meta-information cues are able to over-ride the default generalisation process that was identified in Meta-Analytic Study One.

In addition, there was a significant difference between the generalisation outcomes of inclusive vs. exclusive meta-information cues ($Q (1) = 26.93$, $p < .001$). Or put differently, this suggests that meta-information cues encourage people to consider the cognitive inclusion of the exemplar material with some flexibility. These ideas will be explored further below in the concluding comments section.

<table>
<thead>
<tr>
<th>Table 3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of Effect Sizes for Overall Generalisation and the Outcome Measures</td>
</tr>
<tr>
<td>N (N%)</td>
</tr>
<tr>
<td>Overall</td>
</tr>
<tr>
<td>Inclusion</td>
</tr>
<tr>
<td>Stereotypicality</td>
</tr>
<tr>
<td>Dispersion</td>
</tr>
<tr>
<td>Prejudice</td>
</tr>
<tr>
<td>Exclusionb</td>
</tr>
</tbody>
</table>

Notes. N = number of tests; N% = percentage of the 24 total tests; SE = standard error of $g$; 95% CI = 95% confidence interval of $g$; $Z$ = the standard score; $p$ = probability of $g$. a test of difference between each category: $Q (1) = 26.93$, $p < .001$; b all tests measured stereotypicality; * small sample set.

Source of Cue

The effect of the source of the meta-cognitive task cue was investigated separately across the two levels of the cue direction. For inclusive cues, it was found that both internal cues ($M_g = .51$, $r = .25$, $p < .001$) and external cues ($M_g = .56$, $r = .27$, $p < .001$) led to
significant generalisation facilitation (Table 3.3). The difference between internal and external cues was non-significant ($Q < 1$). All the exclusive cues were internal in nature, and so the data remains the same as that presented in Table 3.2 for exclusion cues. Altogether, from this analysis it is apparent that internal and external cues are equally able to facilitate generalisation, with no distinct advantage of either cue source.

Table 3.3

<table>
<thead>
<tr>
<th></th>
<th>N (N%)</th>
<th>$G$</th>
<th>$r$</th>
<th>SE</th>
<th>95% CI</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion (19 tests)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>12 (50)</td>
<td>.51</td>
<td>.25</td>
<td>.09</td>
<td>.32/.69</td>
<td>5.46</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>External</td>
<td>7 (29)</td>
<td>.56</td>
<td>.27</td>
<td>.08</td>
<td>.40/.72</td>
<td>6.82</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Exclusion (4 tests)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>4 (17)</td>
<td>-.51*</td>
<td>-.25</td>
<td>.19</td>
<td>-.89/-1.13</td>
<td>-2.65</td>
<td>= .008</td>
</tr>
</tbody>
</table>

Notes. $N$ = number of tests; $N\%$ = percentage of the 24 total tests; $SE$ = standard error of $g$; 95% CI = 95% confidence interval of $g$; $Z$ = the standard score; $p$ = probability of $g$.

* Small sample set.

Discussion

The overall effect size identified in this meta-analysis testifies that the magnitude of the generalisation effect size depends also on the meta-cognitive cues provided in impression formation paradigms. This meta-analysis builds on the research reviewed in Chapter 2 because it suggests that the specific instructions provided to participants when they receive the exemplar information impacts on the influence of this information on group judgments. Traditionally, social cognitive interventions to reduce stereotyping have used information-based programs, such as interventions of mass communication aimed to alter biased group perceptions by conveying messages loaded with stereotype-incongruent information. This evidence suggests that a new approach to promoting stereotype change may be achieved through programs based on raising the awareness of the ways in which people monitor and carry out member-to-group generalisation (e.g., Paolini et al., 2009; for a similar argument in member-to-member research, see Yzerbyt, Dardenne, & Leyens, 1998).

This meta-analysis found that the meta-information cue can either facilitate or inhibit the use of the exemplar information for the group judgment. In particular, inclusive meta-information cues promote member-to-group generalisation, while exclusion meta-information cues inhibit the process of member-to-group generalisation by encouraging the exclusion of exemplar information from the group judgment. This finding raises two important considerations. Firstly, the idea of meta-information cues either facilitating or inhibiting inclusion of the exemplar material does not reflect a necessarily good or a bad social outcome (respectively). The desirable vs. undesirable nature of the end product for intergroup relations, rather, depends on the nature of the stereotype under investigation and on the nature of the
exemplar information provided. This means that, in the context of interventions aimed at addressing a negative group stereotype, the outcome would be desirable if an inclusive cue is paired with stereotype disconfirming exemplars. However, the outcome would be undesirable (i.e., greater group stereotyping) if an inclusive cue is paired with stereotype confirming exemplars. The pattern is completely reversed in contexts of interventions aimed at improving already positive group views. The important thing from this finding is that, at a more basic level, the meta-cognitive cue can operate in either direction. Secondly, there is limited evidence on the role of exclusive cues. Only four of the 24 tests (16%) used exclusive meta-information cues. This should be an area for further investigation.

The significant difference between facilitating and inhibiting meta-information cues provides support for the inclusion-exclusion model (Schwarz & Bless, 1992; Bless & Schwarz, 1998). At least in their earlier work, Schwarz and Bless did not explicitly frame their inclusion-exclusion model in terms of meta-cognition; however, they do depict stereotypes as fluid cognitive structures that change in response to cues from the environment. In the following chapter, I will build on Schwarz and Bless’ inclusion-exclusion model to more explicitly include meta-cognitive considerations within a meta-cognitive framework for member-to-group generalisation.

In this meta-analysis the source of the meta-information was also assessed. For inclusive task cues, I have shown that external and internal cues equally influence generalisation, in absolute terms, but predictably in the opposite direction due to the intended impact of the cue. At a minimum, this evidence supports Paolini et al.’s (2009) argument that meta-cognition in member-to-generalisation research is “not limited to variables and processes under direct and exclusive control of the experimenter”. This meta-analysis shows that internal cues are also able to influence generalisation. The assessment of source effects among exclusive meta-cognitive cues was, however, limited by a small number of tests. Future research will need to investigate cues that include exclusive meta-cognitive cues.

Altogether, this evidence points towards significant moderating effects of the direction of the meta-cognitive cue and limited moderating effects of the cue source in member-to-group generalisation. However, it should be noted that the impression formation paradigm remains limited as a context to investigate and trigger meta-cognitive considerations. This is because the impression formation paradigm relies heavily on the provision of experimenter controlled external sample information. This makes it a rather artificial framework for the injection of internal meta-cognitive cues. Put differently, even though I coded certain meta-information cues as internal, the actual experimental manipulations behind these cues were also artificially injected by an external source (the experimenter), making any conclusions surrounding the source of meta-information provisional. Later in this thesis, I will argue that a more suitable paradigm to investigate the
role of internal meta-information cue is one where internal meta-information comes directly from the participant, and is not artificially injected by the experimenter. To achieve this, I will develop a ‘self-generation’ paradigm.

**Concluding Remarks**

Three key considerations have emerged from this review of the fluency literature and impression formation literature. Firstly, the review of impression formation research has revealed that meta-information can impact on generalisation to facilitate inclusion or exclusion of the exemplar information. However, within the retrieval fluency literature on social judgments it was noted that the lack of a baseline condition leaves the exact directional impact of retrieval fluency meta-information unclear. It is important that future retrieval fluency research includes a baseline condition to identify the directional influence of retrieval fluency meta-information in out-group judgments (for a baseline condition in hindsight bias, see Sanna et al., 2009). One goal of my research will be to address this gap in the fluency literature. Secondly, the member-to-group generalisation research has also revealed that both internal and external meta-information cues influence generalisation to a similar extent. However, a criticism of this research is that the ‘internal’ cues remain artificially injected within the impression formation paradigm. My research will address this criticism by ensuring that internal cues are truly ‘internal’ to the participant and not an extrinsic part of the research procedure. Finally, little explicit member-to-group generalisation work has investigated the direction and source of meta-information task cues. My work is a serious and direct effort at testing these two meta-information concepts. Specifically, my research will test whether internal meta-information (retrieval fluency; Studies 1 – 5) and external meta-information (processing fluency; Study 5) facilitates inclusion or exclusion of the exemplar material in member-to-group generalisation.
CHAPTER 4

MEMBER-TO-GROUP GENERALISATION AND FLUENCY:
THEORY AND METHOD

My thesis aims to extend member-to-group generalisation research through a serious and explicit effort at understanding the involvement of meta-cognition. In this chapter, I take a first step towards achieving this goal by introducing a dual process account of meta-cognition and member-to-group generalisation. Firstly, I will discuss the theoretical framework to account for the influence of meta-cognition in member-to-group generalisation. Then, I introduce the paradigm that will be used to test this theoretical model.

The Theoretical Model: A Dual Process Account

My dual process model draws from two previous dual process models of member-to-group generalisation (for a review of dual process models in social inference, see Kruglanski & Orehek, 2007). The first model is presented by Bless and Schwarz (1998; see also Schwarz & Bless, 1992) and the second is presented by Paolini et al. (2009). To my knowledge, the first dual process account of member-to-group generalisation was the inclusion-exclusion model presented by Schwarz and Bless (1992; also see Bless & Schwarz, 1998). Bless and Schwarz assume that group judgments depend on how the exemplar information is used (Schwarz & Bless, 1992). Exemplar information that is included in the group judgment should result in assimilation effects, while exemplar information that is excluded from the group judgment should result in contrast effects (see Bless et al., 2001). To take place, these assimilation or contrast effects require information about the exemplar (typically provided by the researcher in standard impression formation studies) and information about the group (typically the out-group stereotype). Importantly, in their model, Bless and Schwarz argue that the “default operation is to include the information that comes to mind” (p. 169). However, this ‘default operation’ can be avoided through contextual cues that trigger additional cognitive processing (i.e., meta-cognitive processes), resulting in the exclusion of material that is perceived as unrepresentative to the category or irrelevant to the judgment at hand.

Bless and Schwarz (1998) tested their model by asking students at a German university to consider a highly respected German politician (Richard von Weizsäcker) before
expressing their group judgment towards the Christian Democratic Party (CDP). Through the use of a cleverly worded question, participants were either encouraged to include or exclude this positive exemplar from the group representation. Specifically, in the inclusion condition, participants were encouraged to consider Richard von Weizsäcker as a member of the CDP. In the exclusion condition, participants were encouraged to consider von Weizsäcker in his role as the above-party representative of the Federal Republic of Germany. In line with Bless and Schwarz’s predictions, when participants were cued to consider von Weizsäcker as a CDP member, the group judgments included the exemplar’s high positivity and were more positive than in the no-exemplar baseline condition. However, when participants were cued to exclude von Weizsäcker from the CDP, the group judgments excluded his positivity and were more negative than in the no-exemplar baseline. These results support Bless and Schwarz’s tenet that the use of information in out-group judgments is not only influenced by ‘what’ information is accessible (e.g., von Weizsäcker), but also ‘how’ it is used during judgment construction (inclusion vs. exclusion cues).

It is important to note that the inclusion-exclusion model presented by Bless and Schwarz (1998) was not framed in meta-cognitive terms. However, in my opinion, the consideration of additional cognitive processes involving the exclusion of exemplar information is fully consistent with a meta-cognitive take on generalisation. It is my intention to extend Bless and Schwarz’s model by providing a more precise articulation of the impact of meta-cognition on member-to-group generalisation.

The second dual process model that I have built upon is by Paolini et al. (2009). Paolini et al. argued that exemplar information will be excluded or included in a group judgment depending on the joint impact of accountability and perceived information validity. Specifically, Paolini et al. argued that under normal circumstances accountability facilitates inclusion of the exemplar material by encouraging information processing. However, when participants are provided with information that undermines the perceived validity of the exemplar information accountability should result in the exclusion of the exemplar information because of the heightened sensitivity. The evidence for this model, reviewed in Chapter 3, supported the dual role of accountability as a meta-cognitive process. Unlike the model provided by Bless and Schwarz (1998), the dual process model by Paolini et al. (2009) is explicitly meta-cognitive. However, it is bound to the specific process of accountability. The dual process model that I propose is broader in applicability than the model presented by Paolini et al.

I propose a dual process model that involves automatic meta-cognitive processes and controlled meta-cognitive processes (for a similar distinction, see Ferreira, Garcia-Marques, Sherman & Sherman, 2006). Broadly, meta-cognitive research can be divided into models that investigate automatic meta-cognitive processes (i.e., meta-cognitive processes occurring
outside of conscious awareness; Lories, Dardenne, & Yzerbyt, 1998), and models that investigate controlled meta-cognitive processes (i.e., people’s meta-beliefs about their cognitive processes; Banaji & Dasgupta, 1998). I hypothesise that (1) the meta-information associated with high fluency is applied to the judgment in an automatic or sub-conscious manner (for a similar point regarding mood, see Jost et al., 1998); and (2) that this automatic process can be overridden by a controlled process under the provision of additional meta-cognitive cues. Below I detail this model which offers a significant extension of the current conceptualisation of fluency.

**Fluency as an automatic process.** Automatic meta-cognitive processes are those in which there is a ‘spontaneous’ use or application of the processes of working memory. Automatic processes were initially conceptualised as occurring outside of conscious control (e.g., Hays-Roth, 1977; Shiffrin & Schneider, 1977). It is now accepted that automaticity can be inferred when a perceiver lacks awareness of the process, the process is efficiently conducted using minimal cognitive resources, has no intention to carry out the process, or cannot control it (Andersen, Moskowitz, Blair & Nosek, 2007). A problem in arguing for automatic meta-cognitive processes lies with the traditional conceptualisation of meta-cognition as an ‘overriding’ controlled process (for a similar point, see Lories et al., 1998). This ‘overriding’ conceptualisation emerged because meta-cognition has been modelled as a system used to control and monitor the processes that occur at a cognitive level (e.g., Ridley, Schutz, Glanz, & Weinstein, 1992). However, this controlled conceptualisation raises the spectrum of homunculi: If the cognitive level needs an overseer, then the meta-cognitive level should also require an overseer, and so on ad infinitum. Nelson (1992) challenges this traditional view of cognitive and meta-cognitive processes as separate structures, with one level overseeing the other. Instead, he suggests that both cognition and meta-cognition should be viewed as operating through the same cognitive architecture. From Nelson’s standpoint, accepting that some cognitive processes are automatic (e.g., memory access, stereotype activation, spreading of semantic activation) necessarily implies accepting that some meta-cognitive processes are also automatic.

According to Andersen and colleagues, automatic processes can be broadly organised around four basic processes (Andersen et al., 2007). Firstly, as a precursor to automaticity, the process must be “available”. In other words, the individual must have the process available as a strategy to adopt. These processes may develop as a learned process that becomes increasingly automatic with practice (Petty, 2006). Evidence of a retrieval fluency effect on the explicit measures will support the psychological availability of the process. Secondly, the process must be “accessible”. That is, in the present context, the fluency heuristic should be readily accessible to participants. I will test the accessibility of retrieval fluency by including in Study 1 an Implicit Association Test (Greenwald et al., 1998; Greenwald et al., 2003).
expect that the IAT displays a retrieval fluency effect supporting the role of accessibility. Thirdly, “applicability” is required, that is, the process must be applicable to the situation at stake. In regards to retrieval fluency, I will test for applicability by including group judgements on retrieval irrelevant traits. If the application of fluency to the subsequent judgment is truly automatic then there should be a retrieval fluency effect on retrieval relevant traits but not on retrieval irrelevant traits. Finally, “self-regulation”, or control, should prevent the application of the automatic process. In Studies 2 – 4, I will test the control aspect of retrieval fluency by introducing a cue that is designed to encourage self-regulation.

Within fluency research, several researchers have argued that people are aware of variations in fluency (Jacoby & Dallas, 1981; Kelley & Jacoby, 1998; Kelley & Lindsay, 1993; Winkielman et al., 2003; Whittlesea, 1993; Whittlesea & Leboe, 2003). With my model, I dispute this – I do not suggest that people are unaware of when it is easy or difficult to recall or process stimuli. Rather, I suggest that they are unaware of the application of the meta-information to the judgment (i.e., that fluency affects judgment; for a similar point on ‘unconscious’ influences, see Dijksterhuis & Smith, 2005). Hence, I suggest that people have learned through experience and feedback to automatically apply meta-cognitive considerations, such as fluency, when building judgments (e.g, Kahneman, 2003; Rieskamp & Otto, 2006; Strack & Deutsch, 2004).

**Invoking a controlled process.** Meta-cognitive processes that require conscious and effortful use of strategies are regarded as controlled. The very idea of meta-cognition evokes a sense of a ‘controlled’ process and controlled meta-cognitive processes have been found to have a direct influence on behaviour and judgment. For example, controlled meta-cognitive processes have been demonstrated to override automatic stereotype activation and application (Devine, 1989; Yzerbyt et al., 1998).

Particularly relevant to the integration between meta-cognition and member-to-group generalisation is the controlled meta-cognitive process of bias correction. Bias correction processes emerge when individuals attempt to avoid judgmental bias by adjusting their judgments against their suspected biasing action (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Martin, 1986; Wegener & Petty, 1995; Wegener, Petty, & Dunn, 1998). These processes may result in an assimilation effect, where the information is considered valid and included in a judgment, or a contrast effect, where the information is considered invalid and is excluded from a judgment.

Individuals have a variety of correction devices and display extreme flexibility when correcting for unwanted bias. According to Wegener and colleagues (Wegener & Petty, 1995; Wegener et al., 1998), corrections are triggered by contextual factors and are guided by naïve theories of how a given factor can influence judgments. In Wegener and colleagues’ view, people correct for unwanted influences on their judgments only when they are aware of those
factors’ influence. As part of my dual process account, I suggest that when people are made aware of the automatic influence of fluency they will engage in a bias correction process to correct for this influence.

**Supporting the automatic vs. controlled distinction.** Retrieval fluency research encourages bias correction by increasing participant’s awareness of otherwise automatic fluency effects, thus supporting my dual process account (for a parallel distinction on unconscious vs. conscious processing, see Dijksterhuis & Nordgren, 2006). For example, Rothman and Hardin (1997, Study 2) demonstrated that bias correction processes can influence the application of retrieval fluency meta-information. Similar to their Study 1 reviewed in Chapter 3, they used a few-versus-many paradigm to investigate the role of retrieval fluency in the social domain. That is, male and female participants were asked to recall three or six behaviours performed by either the same gender (in-group) or different gender (out-group) that were considered to be assertive or non-assertive. In addition, some participants were provided with a task cue that was expected to alert them to the influence of retrieval fluency in their judgments. Specifically, prior to recalling the assertive behaviours, participants were provided with the following meta-information task cue: “because you have been completing a number of tasks during this session, you may find it somewhat difficult (or easy) to recall specific behaviours” (p. 127). This meta-information cue was either congruent (e.g., being told the task would be hard and recalling six behaviours) or incongruent (e.g., being told the task would be hard and recalling three behaviours). It is this kind of explicit cue that should interrupt the automatic application of retrieval meta-information to the subsequent judgment.

Rothman and Hardin (1997) predicted that the provision of a congruent task cue would alert participants to the potential influence of fluency meta-information on their judgment (i.e., stimulating a bias correction process). This occurs because the congruent cue is informative and relevant to the task, encouraging participants to ensure that the meta-information associated with the retrieval experience is (correctly) attributed to the task. As a result, participants should rely on the retrieved content when making judgments. On the other hand, participants in the incongruent cue condition were expected to display a retrieval fluency effect because the incongruent cue is uninformative for the task (or insufficient to stimulate a bias correction process). As a result, participants should continue relying on the meta-information associated with the retrieval experience to help form their judgement.

Partial support of the predicted pattern of results emerged amongst male participants, but not amongst female participants. In the congruent cue condition, male participants tended to consider women more assertive after recalling six female assertive behaviours than after recalling three behaviours. However, this difference did not reach conventional levels of significance, thus showing a null effect of retrieval fluency under congruent task cues, rather
than the predicted ‘reversal’. In the incongruent condition, male participants considered women more assertive after recalling three female assertive behaviours than after recalling six female assertive behaviours (i.e., a classic retrieval fluency effect). The same pattern of results emerged for the male participants recalling female non-assertive behaviours. However, there were no effects of the manipulation of retrieval fluency for female participants when judging men (out-group) or women (in-group). The key point from Rothman and Hardin’s (1997) research is that retrieval fluency effects were found to be controlled for through the provision of an explicit meta-information task cue at least in a subgroup of their sample.

The Research Method: A Self-Generation Paradigm

To test my dual process model, I will introduce a hybrid paradigm that blends the impression formation paradigm with the few-versus-many paradigm. As argued in the previous chapter, both the traditional impression formation paradigm and the few-versus-many paradigm are sub-optimal testing grounds for an extension of retrieval fluency to member-to-group generalisation. Thus, extending the investigation of fluency to member-to-group generalisation seems to require a novel approach that hybridises these two existing paradigms.

The few-versus-many paradigm requires participants to retrieve information from their memory and then judge some aspect of the retrieved information domain. The impression formation paradigm presents participants with exemplar information and then measures out-group judgments. In my new hybrid paradigm, I will ask people to retrieve exemplar information from their memory and then measure for out-group judgments. This hybrid paradigm will be referred to as the self-generation paradigm, and will be detailed more thoroughly in the next chapter.

Overview of the Empirical Research

The five empirical studies presented in this thesis will primarily focus on retrieval fluency. Studies 1 – 4 will manipulate retrieval fluency by systematically varying the number of counterstereotypical individuals that participants are asked to retrieve from their memory. Study 5 will present a novel manipulation of fluency that allows a joint test of retrieval and processing fluency within the one paradigm. A significant extension to the previous fluency research is the inclusion of a no-exemplar control condition in each of my empirical studies. As noted in the previous chapter, retrieval fluency studies have not typically included a control condition (e.g., Dijksterhuis et al., 1999; Hardin & Rothman, 1997; Schwarz et al., 1991; but see, Sanna et al., 2009). The lack of a baseline in social judgments makes it impossible to identify the exact direction of the effects. Participants in the few-exemplar
condition may be using the meta-information about the state of their retrieval process as a cue to include the exemplar information (e.g., Schwarz et al., 1991). Similarly, participants in the many-exemplar condition may be using the meta-information as a cue to exclude the exemplar information (e.g., Sanna et al., 2009). This interpretive uncertainty will be addressed in the subsequent studies by including a no-information control condition that highlights the direction of the manipulated effects.

Measure type was found to have a powerful moderating effect on generalisation in Meta-analysis Study 1. The outcome measures designed to assess generalisation will be detailed more thoroughly in the following chapter. For now, it is important to note that to ascertain the moderating effect of outcome variables generalisation is assessed using explicit measures of group stereotypicality, group dispersion, and prejudice towards the group.

In a first step towards understanding the automaticity of fluency effects, I have included implicit and explicit measures of out-group judgments in Studies 1 and 4. Participants will be asked to judge the out-group on explicit measures and also measures of implicit stereotyping through an Implicit Association Test (IAT; for a review see Greenwald et al., 2002). Greenwald and Banaji (1995, p. 5) argued that it is “theoretically essential” to include an implicit measure when testing for an implicit (or automatic) cognitive process. If fluency is an automatic process, the results on the IAT should mirror the results on the explicit measures. This hypothesis will be addressed in more detail in the following chapter. To my knowledge, my work is the first to test the ‘implicit’ application of meta-information in fluency research.

**Explaining Retrieval Fluency**

Many theorists have investigated fluency and its impact in different judgment domains, but the mediating processes remain largely elusive (but see later for a discussion of Tormala et al., 2002; Tormala, Falces, Petty, & Briñol, 2007). To build our understanding of the psychological underpinning of retrieval fluency effects, I will investigate a number of different mediating processes. These processes include unrequested cognitions, perceived prior contact, cognitive fatigue, and differences in depth of cognitive processing. I will introduce each of these processes in detail in the following empirical chapters. For now, I will outline three mediating processes that I have most systematically investigated: (1) exemplar vividness, (2) the ‘easy-thus-liked’ heuristic, and (3) the ‘easy-thus-typical’ heuristic.

**Exemplar vividness.** A problem associated with the traditional manipulation of retrieval fluency in the few-versus-many paradigm is the unequal number of exemplars retrieved across conditions. Participants in the many condition not only have a feeling of retrieval difficulty they also have more retrieved items cognitively accessible than participants in the few condition. Retrieval fluency effects may be associated with poorer quality of the
latter retrieved exemplars. Most retrieval fluency researchers have ruled out this potential problem by checking the quality, or vividness, of the latter retrieved examples with the earlier retrieved examples. In line with this approach, and to rule out this interpretation, I will compare the vividness of the early retrieved exemplars against the vividness of the latter retrieved exemplars throughout each of my empirical studies.

Stepper and Strack (1993, Experiment 2) addressed the problem of unequal sample sizes by manipulating retrieval fluency while holding the number of items constant (see, also Garcia-Marques & Mackie, 1999). Stepper and Strack asked their participants to generate six examples of assertive or unassertive self-behaviors. During the process of recall the participants were asked to produce a light smile or furrow their brow. These researchers argued that a light smile is an expression commonly associated with a feeling of ease, while a furrowed brow is an expression commonly associated with a feeling of effort. They found that participants in the light smile condition judged themselves more assertively than the participants in the furrowed brow condition, thus demonstrating a retrieval fluency effect without varying the number of retrieved behaviors. Studies 1 – 4 will adopt a standard few-versus-many retrieval fluency paradigm. However, Study 5 will introduce a novel manipulation of fluency that holds the number of items constant across conditions, in line with Stepper and Strack’s approach.

Easy-thus-liked heuristic. The easy-thus-liked heuristic comes from processing fluency research, rather than retrieval fluency research. Processing fluency research shows that high processing fluency is positively marked. According to Winkielman and Cacioppo (2001), high fluency results in a “brief affective reaction that is selectively positive” (p. 995). Put simply, ‘easily’ processed items are subsequently ‘liked’. Through the easy-thus-liked heuristic, I extend this reasoning to retrieval fluency effects and propose that high retrieval fluency may generate a positive affective state that becomes associated with the retrieved content resulting in enhanced positivity of the out-group judgment.

There is some evidence that fluency affects valence judgments, supporting the role of the easy-thus-liked heuristic. For example, Winkielman and Cacioppo (2001) asked participants to view simple line drawings of neutral objects such as a house, bird or an aeroplane on a computer screen. Study 1 used a visual prime to manipulate processing fluency that was either matched (high fluency) or mismatched (low fluency) to the image. For instance, prior to viewing the line drawing participants in the matched condition saw a contour on the screen that matched the drawing, while participants in the mismatched condition saw an un-matched contour. Study 2 used the length of presentation to manipulate processing fluency, with images presented for 900 ms (high fluency) or 300 ms (low fluency). In both studies, self-reported measures of stimulus positivity indicated that participants in the high fluency condition were more positive in their judgments of the images than participants
in the low fluency condition. Importantly, Winkielman and Cacioppo did not find an increase in negative judgments of the stimuli when the processing was difficult, suggesting that fluent processing is positively, rather than negatively, marked.

In my research, support for the easy-thus-liked heuristic should take the form of ease of retrieval being associated with an increase in the perceived likeability of the out-group (i.e., reduced prejudice) and a corresponding increase in generalisation.

**Easy-thus-typical heuristic.** The ‘easy-thus-typical’ heuristic builds on social cognitive research and theorisation. According to Rothbart and John’s (1985) cognitive analysis of generalisation, typicality is important at both encoding and retrieval stages (see also Rothbart, 1996; Rothbart & Lewis, 1988). At the encoding stage typical exemplars are more closely stored with the group stereotype than atypical exemplars. At the retrieval stage, people more easily retrieve typical than atypical exemplars (e.g., Rothbart, Sriram, & Davis-Stitt, 1996). Therefore typical group members are more likely to be used to form a group judgment (see also, Rothbart, 1996).

Tversky and Kahneman (1973) were the first to formalise the idea that people come to a judgment based on the “ease with which instances or associations come to mind” (Tversky & Kahneman, 1973, p. 208). They also pointed toward the possible involvement of perceived exemplar typicality or representation in fluency effects. In their empirical work with the availability heuristic they found that participants would consistently overestimate the number of words in the English dictionary that began with the letter r (i.e., regarded as more typical or more represented in the dictionary) and underestimate the number of words that have r as the third letter (i.e., regarded as less typical or less represented in the dictionary). They suggested that this effect occurred because people have more r words easily available in memory than they do for words that have r as the third letter. Put differently, ease of retrieval should inform judgments of exemplar typicality.

Drawing from Rothbart and John’s (1985) and Tversky and Kahneman (1973) arguments, I suggest that if recalling typical exemplars from a category is easy, then easy retrieval should cue typicality and ultimately contribute to the explicit retrieval fluency effects. On the other hand, when retrieval is difficult participants should become aware that there are limited available exemplars of that kind in memory and, perhaps, in their repertoire of past experience with the group (see Schwarz et al., 1991). As exemplar retrieval becomes difficult people should rely more heavily on the momentary experience of one’s cognitive processes because they regard the exemplars that they have retrieved as highly atypical. The end result will be reduced generalisation of the retrieved exemplars to the out-group judgment.

Despite the plausibility of the easy-thus-typical heuristic, no direct evidence for this mediating process is currently available. In Studies 1 – 3, I will measure exemplar typicality
and use a measurement-of-mediation design (Spencer, Zanna, & Fong, 2005) to test the mediating role of exemplar typicality in explaining retrieval fluency effects. In my final empirical study, I will use a moderation-of-process design (Spencer et al., 2005) and include exemplar typicality as a moderating variable. Evidence from this heuristic should take the form of ease of retrieval being associated with an increase in perceived exemplar typicality and an increase in generalisation.

**Summary of Key Hypotheses**

I anticipate that this research will enrich our understanding of the interplay between meta-cognition and stereotyping (Jost et al., 1998) and the implications of fluency for the process of member-to-group generalisation. To summarise, the present research will test the following key hypotheses that:

(1) Group judgments will normally reflect the nature of exemplar information cognitively accessible at the time of the judgment (i.e., the information) and the subjective experience of accessing the exemplar information (i.e., the meta-information).

(2) People should automatically use meta-information when making group judgments, but only if the informational validity of the meta-information is not undermined by congruent meta-information cues.

(3) When the informational value of the meta-information is undermined by a congruent meta-information cue, participants should rely more extensively on the content of the retrieved material at the time of building up the group judgements.

(4) People should use the meta-informational component of the recall experience to infer likeability and/or typicality of the exemplar and include this in their group judgments; this should mediate the effect on group judgments.

The research presented in this thesis focuses predominately on retrieval fluency effects (Studies 1 – 5), as internal meta-information is the area where further member-to-group generalisation research is needed. However, external meta-information, in the form of processing fluency, will be tackled in Study 5. In the following chapter, I will present my first empirical test of the impact of retrieval fluency on member-to-group generalisation. In this first study, participants will be asked to retrieve either a few or many counterstereotypical group members from their memory and then complete out-group judgments. If internal meta-information plays an influential role in member-to-group generalisation then group judgments will differ between the few and many conditions, with participants in the few-condition demonstrating greater generalisation than participants in the many-condition.
CHAPTER 5

A META-COGNITIVE APPROACH TO GENERALISATION:
TESTING RETRIEVAL FLUENCY EFFECTS (STUDY 1)

In the earlier chapters, I argued that social meta-cognition can provide a new approach to member-to-group generalisation by focusing on the role of specific meta-cognitive processes involved in generalisation. In particular, fluency may facilitate inclusion or exclusion of exemplar material from the out-group judgments. In this chapter, I design a new paradigm that integrates the meta-cognitive process of retrieval fluency within a member-to-group generalisation framework. Using this new paradigm, I will then assess the role of internal meta-information on member-to-group generalisation. Finally, to expand our understanding of retrieval fluency, I will also assess the nature of retrieval fluency effects. Each of these goals is addressed in more detail below.

The Experimental Paradigm

I have defined internal meta-information as information that comes from the individual about the success and/or ease with which mental operations are performed. I will include internal meta-information within my self-generation paradigm by having participants retrieve information about counterstereotypical group members from their memory. Retrieval fluency will be manipulated through the use of the few-versus-many approach (see, Schwarz et al., 1991; for a review, see Schwarz, 1998), where participants are asked to generate either a few counterstereotypical exemplars or many counterstereotypical exemplars. It is important to note that through the use of a self-generation paradigm, my research will not have to externally inject meta-information via cues from the experimenter (see Chapter 3). Generalisation will be measured through changes in group judgments between conditions. These include explicit outcome variables assessing changes in group stereotypicality, dispersion and prejudice as well as in implicit stereotyping.

Asking people to self-generate personally known counterstereotypical exemplars improves the ecology of traditional social-cognitive research by removing the sterile presentation of unfamiliar counterstereotypical exemplars typical of traditional impression formation paradigms. Through media, friends, family, and colleagues we are continually exposed to a variety of individuals that confirm and disconfirm our stereotypes. Fielder and Walther (2004) proposed that stereotypes emerge, and are maintained, due to the biased
nature of this information rather than biased individuals. By ignoring the individual’s prior history with the out-group, impression formation research has typically ignored the exemplar information that contributes to stereotype formation and maintenance. To address this limitation, and improve the ecological validity of impression formation paradigms, my paradigm will account for an individual’s social ecology and focus on the individual’s previous experience with, and exposure to, out-group members.

It is worthwhile noting that Blair, Ma and Lenton (2001) have already made steps to reduce past reliance on presented information in stereotype change research (see also, Stathi & Crisp, 2008). In their research, participants were asked to imagine counterstereotypical exemplars prior to make outgroup judgments. This mental imagery process resulted in reduced implicit stereotyping. Blair et al.’s imagined paradigm, the self-generated paradigm and the impression formation paradigm are all likely to be subjected to fluency effects.

Without diminishing the pragmatic significance of stereotype change through a variety of presentation mode, it remains that, as real outgroup examples are closer to the individual’s own experience and social reality, a self-generated paradigm should have a net advantage in ecological validity over the other two paradigms. Future research should investigate whether these different ways of accessing out-group member information are associated with differences in perceived information ecology in the eyes of the participant and if this ultimately impacts on the magnitude of generalization on out-group judgments.

Asking people to self-generate counterstereotypical exemplars also improves on contact research by removing some of the problems of control over confounding variables. For example, early contact research by Stouffer, Suchman, DeVinney, Star, and Williams (1949) compared the attitudes of White American soldiers towards African American soldiers after a period of intergroup contact during WWII. The research compared a pre-contact baseline with post-contact data, and found that White American soldiers were less opposed to racial desegregation. The problem with this research is that it does not control for the complexities associated with face-to-face contact. Moreover, it is unclear the exact causal sequence operating (MacCullum & Austin, 2000). By asking people to retrieve out-group exemplars from their memory the self-generation paradigm accesses intergroup contact experiences within a controlled methodology. Here the participants are focused on the counterstereotypical traits displayed by the exemplar rather than other peripheral information (e.g., sights, sounds, smells) available at the time of contact.

There are reasons to believe that this paradigm will lead to generalisation. The meta-analyses conducted by Pettigrew and Tropp (2000; 2006; Tropp & Pettigrew, 2004; 2005) on the contact literature and those I carried out in Chapters 2 and 3 demonstrated that generalisation of contact experiences and generalisation after exposure to out-group information occur at similar rates. Therefore, it is appropriate to expect that participants
retrieving past experiences with counterstereotypical out-group members will also generalise this information to the out-group.

As noted in Chapter 3, much of the previous few-versus-many research has failed to include an assessment of the direction of the effects (but see, Sanna et al., 2009). For example, Schwarz et al. (1991, Study 1) asked their participants to retrieve either 6 or 12 examples of assertive behaviour from their memory, and then judge how assertive they considered themselves. The results indicated that in the six-example condition participants judged themselves to be more assertive than the participants in the 12-example condition. Because of the lack of a baseline condition, it is possible that the participants in the 12-example condition were driving the effects and considered themselves to be unassertive. As reviewed in Chapter 3, Sanna et al. (2009) included a control condition and found evidence that meta-information is included in the judgment when retrieval is easy (i.e., assimilation effect), while meta-information is excluded from the judgment when retrieval is difficult (i.e., a contrast effect). Sanna et al.’s research was in hindsight bias, it is therefore important to extend the investigation of the direction effects of meta-information to the domain of out-group judgments. To address the directional impact of internal meta-information in out-group judgments, similar to Sanna et al., I will include a control group that will only complete the dependent measures providing a baseline to detect stereotype change as well as stereotype maintenance.

According to Kunda and Thagard’s (1996) connectionist model, the process of retrieval should activate an initial memory search which then activates a network of interconnected group members. From this perspective, once the network from one category (e.g., family members) is exhausted participants would then need to activate another category (e.g., famous people). Because such information searches are likely to differ in frequency, proximity, accessibility, richness and interconnectedness (Fielder & Walther, 2004), a measure of category shifts and time taken to complete a retrieval search will be taken as an indication of increasing task difficulty.

Retrieval fluency effects have been identified in a vast array of judgment areas including self-judgments, memory assessments, hindsight bias and intergroup judgments (for a review, see Schwarz, 1998). However, application of retrieval fluency to the area of out-group judgments is limited. As noted in Chapter 3, an isolated exception is by Rothman and Hardin (1997). Rothman and Hardin used a classic retrieval fluency paradigm and asked participants to consider behaviours performed by members of the out-group, and then asked participants to judge the out-group. They found that male participants generalised the retrieved sample of behaviours to the out-group but failed to achieve an effect of the retrieval fluency manipulation for female participants. Asking participants to retrieve ‘behaviours’ rather than ‘exemplars’ may have limited generalisation because behaviours are oftentimes
considered a transient example of a person’s values and attitudes (for a similar point, Gawronski, 2003). Also, by asking participants to consider ‘behaviours’ Rothman and Hardin had limited control over the specific number of exemplars that were retrieved. For example, some participants may have retrieved only one extremely atypical exemplar while others may have retrieved many different moderately atypical exemplars. I will address each of these shortfalls by asking participants to consider a specific number of exemplars who display counterstereotypical traits rather than counterstereotypical behaviours (for the importance of trait-based information in impression formation, see Köpetz & Kruglanski, 2008).

**Key Hypotheses**

**Basic Retrieval Fluency Effects**

As identified in Chapter 3, the process of retrieval provides two distinct sources of information: (1) the retrieved content and (2) the meta-information associated with the process of retrieval (for reviews, see Reber et al., 2004; Schwarz, 1998). In general, when the meta-information indicates that retrieval was easy, the retrieved content is generalised to the subsequent judgment (Sanna et al., 2002; Schwarz et al., 1991; for a review see Schwarz, 1998). It is expected that participants in the ‘few’ exemplar condition will demonstrate greater generalisation than participants in the ‘many’ exemplar condition. When the meta-information indicates that retrieval was hard, the retrieved content may not be generalised to the subsequent judgment at all (i.e., exclusion) or generalised to a lesser extent than in the ‘easy’ retrieval condition (i.e., inclusion). The use of the control condition will isolate the direction of the internal meta-information associated with the hard retrieval condition.

**Exploring the Nature of the Meta-Information**

This research will investigate the exact nature of the retrieval meta-information by testing alternative mechanisms responsible for the predicted retrieval fluency effect. I will test the extent to which retrieval fluency effects reflect an affective and/or a cognitive response. I will also test whether the process is automatic or controlled. Finally, I will assess whether retrieval fluency effects are driven by participants in the ‘many’ condition relying on less vivid exemplars retrieved towards the end of the task.

**Affective vs. cognitive.** In the previous chapter, I introduced two heuristics that may explain the retrieval fluency effect: (1) the easy-thus-liked, and (2) the easy-thus-typical heuristics. The easy-thus-liked heuristic links high fluency to a positive affective state. Previous research has recognised that high fluency elicits a positive affective reaction that perceivers misattribute to the specific stimulus (Schwarz et al., 1991; Winkelman et al., 2003; for a review see, Schwarz, 1998). The easy-thus-typical heuristic instead links high fluency to exemplar typicality. Member-to-group generalisation research has found that disconfirming
information is most successful when the exemplar is perceived to be moderately atypical rather than extremely atypical (e.g., Bless et al., 2001; Desforges et al., 1997; Hantzi, 1995; Hewstone & Hamberger, 2000; Johnston & Hewstone, 1992; for models suggesting this mechanism, see Weber & Crocker, 1983; Wilder, 1984; however, see my meta-analysis results). The easy-thus-typical heuristic suggests that the meta-information associated with high fluency cues typicality which enhances the cognitive inclusion of the exemplars in the group judgement. By including measures of liking and typicality I will be able to identify if either of these processes mediate the predicted retrieval fluency effects.

Support for the easy-thus-liked heuristic will emerge if retrieval fluency effects are found to be mediated by measures of liking. Support for the easy-thus-typical heuristic will emerge if retrieval fluency effects are found to be mediated by measures of typicality. It is possible that one, or both, of these measures mediate retrieval fluency effects and, in so doing, this research will clarify the nature of retrieval fluency effects in the context of member-to-group generalisation (for a similar point, see Tormala et al., 2007).

**Exemplar vividness.** It is possible that the predicted effects are driven by poor exemplars generated towards the end of the ‘many’ exemplar task. As the retrieval task becomes increasingly difficult, participants in the many-exemplar condition may generate less vivid exemplars to complete the task. Consequently, the most accessible exemplar (i.e., the most recently retrieved exemplar) for participants in the many-exemplar condition may be less vivid than for participants in the few-exemplar condition (for a similar point, see Haddock et al., 1999; Haddock, Rothman, & Schwarz, 1996; see also Kopetz & Kruglanski, 2008). To assess exemplar vividness, participants will be asked to provide a description of each generated exemplar. These descriptions will be rated for vividness by counting the number of words included in the exemplar description. To test this explanation of retrieval fluency effects, the vividness of the initial exemplars generated by the ‘many’ participants will be compared with the vividness of the latter exemplars. If this explanation is valid, I should detect a decrease in exemplar vividness from the initial exemplar descriptions to the latter exemplar descriptions and this decrease in vividness should explain the predicted key effect.

**Controlled vs. automatic processing.** In Chapter 4, I proposed a dual process model to account for retrieval fluency effects. I argued that the application of internal meta-information occurs automatically. This is a novel account of retrieval fluency as researchers on fluency are predominately silent on the automatic vs. controlled nature of the retrieval fluency effect. In a first step towards testing my dual process model, I will include an implicit measure of stereotype change.

To measure the ‘automatic’ nature of retrieval fluency, I will use the IAT (Greenwald et al., 1998; Greenwald et al., 2003). The IAT is assumed to be an implicit measure of
stereotyping. This is because participants are typically unable to exert control over their responses while they are focused on speeded categorisation decisions and accuracy (however see, Brendl, Markman, & Messner, 2001; Mierke & Klauer, 2001). The IAT has been extensively used in recent research on stereotyping (for a review see, Greenwald et al., 2002). For example, across five experiments, Blair et al. (2001) used the IAT to measure implicit stereotyping and demonstrated a change in implicit stereotyping after having their participants engage in different mental imagery tasks. Dasgupta and Greenwald (2001) also found that exposure to admired elderly exemplars yielded a substantially smaller automatic age bias on the IAT than after exposure to admired young exemplars. It is, therefore, expected that the IAT will be a suitable instrument to measure implicit stereotyping in the context of my new paradigm.

Evidence of an automatic retrieval fluency effect will be supported if there is a significant effect of the retrieval fluency manipulation on the IAT measure that mirrors the effects on the explicit outcome measures (i.e., greater change in the ‘few’ than ‘many’ condition). To my knowledge, this is the first study to incorporate an implicit measure in retrieval fluency research. In so doing, this study not only answers the call to incorporate implicit and explicit measures in social cognitive research (Givens & Monahan, 2005; Maass, Castelli, & Arcuri, 2000) but it also extends meta-cognitive research to implicit measures.

Although an IAT effect can support a claim of automatic process; it does not necessarily imply that the measured phenomenon is automatic. That is, an IAT effect does not preclude the individual from being ‘aware’ of the measured attitude (Fazio & Olson, 2003). The inclusion of the IAT in Study 1 serves as a first step in testing the dual process model. In a second step, I will test Andersen et al.’s (2007) ‘applicability’ of the retrieval process. That is, I will test that easy retrieving of counterstereotypical group members results in generalisation only if this information is judgment relevant. To test the generalisability (or lack thereof) of retrieval fluency effects to out-group judgments on stereotype traits unrelated to the retrieval task, Study 1 will include out-group judgements on unrelated stereotype traits (see also Gross & Hardin, 2008; Hardin & Rothman, 1997). If retrieval fluency is an automatic process, its effects should occur only on the applicable traits and no effects should emerge on the retrieval un-related stereotype measures.

**Summary**

In summary, Study 1 has four primary goals. Firstly, I aim to test the applicability of a fluency manipulation within a member-to-group generalisation framework. This will be achieved using a self-generation paradigm. In line with a retrieval fluency effect, I expect that participants will demonstrate greater generalisation under conditions of high fluency than under conditions of low fluency. Secondly, I will test the underlying mechanism(s)
Chapter Five

responsible for retrieval fluency effects within a member-to-group generalisation framework. In particular, I will test for likeability and typicality as mediating variables as well as exploring the role of exemplar vividness. Finally, I will test my dual process account of retrieval fluency by including automatic and controlled outcome measures. I will also test for the applicability of retrieval fluency (Andersen et al., 2007), where I expect that generalisation will be limited to judgments for retrieval relevant stereotypes.

Method

Participants and Design

Participants were 48 students (28 females and 20 males) from the University of Newcastle who were recruited through posters placed around the campus. The posters explained that the research was investigating different role models and that participants would be required to recall information about different types of people. The elderly were selected as the out-group for this investigation, in order to avoid any potential in-group effects. Participants needed to be 30 years of age or younger to qualify for this research. The same qualifying criteria were used in all studies. The mean age of the participants was 22.54 years ($SD = 3.61$ years). The majority of the participants indicated that their first spoken language was English (65%; $N = 31$) and that their cultural background was Australian (58%; $N = 28$).

Participants were tested individually and given $10 for their help. This study followed a one-way between subjects factor design: number of exemplars (three vs. nine) with an appended control group that completed only the dependent variables. There were 16 participants in each cell of the design.

Setting up the Paradigm

The elderly were selected as the out-group for this investigation for several reasons. Research on social attitudes has identified that many young people hold negative attitudes toward older adults (Allan & Johnson, 2009; Kite & Johnson, 1988; Perdue & Gurtman, 1990; but see Jackson & Sullivan, 1988), and this negative view has been recognised as particularly prevalent amongst young Australians (Davison, 1995). The elderly were also selected as ageism has not been as caught-up in the politically correct movement as racism and sexism, and is one of the remaining ‘isms’ to which people still ascribe negative attributes (Butler, 2005; Duval et al., 2000). Overall, I expected that due to the unique, and marginalised, position of the elderly in Australian society this out-group would be subjected

7 The control condition will be used as a baseline condition to depict change in the group stereotype. Studies 2, 4 and 5 have all included an appended 0-exemplar control condition to also show change in group stereotyping. To maintain consistency in presentation across studies, I have analysed the results here using a number of exemplars (three vs. nine) between subjects ANOVA rather than a number of exemplars (zero vs. three vs. nine) between subjects ANOVA.
to stereotypical attitudes, and due to the numerical size of this out-group the information coming to young people about the elderly would be large enough to allow for the implementation of a paradigm that relies on the retrieval of counterstereotypical exemplars from memory. Throughout the present research an elderly person was described as being over 65 years of age. This was selected as it is the traditional age of retirement in Australia.

Although there are many stereotypes associated with the elderly, such as ‘conservative’, ‘close-minded’ and ‘stuck in the past’, the stereotype of the elderly as ‘inactive’ was selected. The ‘inactive elderly’ stereotype has prevailed throughout Western history. During the late 19th Century social Darwinism, industrialisation, and medical and scientific observations led to an increasing assessment of the elderly being associated with reduced activity and productivity (Haber, 1983; Stearns, 1980). Early empirical support for the inactive elderly stereotype emerged from Tuckman and Lorge’s (1953) survey of psychology students’ attitudes towards the elderly. Tuckman and Lorge found that their participants, who were attending a psychology course on the ageing process, perceived elderly people to be inactive. More recently Palacios, Torres and Mena (2009) found that older people also subscribe to the stereotype of the elderly as inactive. For all of these reasons, the wide-spread stereotype of the elderly being inactive will be investigated in the subsequent studies.

Procedure

Stimulus material. Pilot Study 2 was conducted to identify the specific number of exemplars to be used in the few and many conditions and to establish the traits to be included in the dependent measures. Participants (N = 24; 21 females and 3 males) were first year psychology students at the University of Newcastle. They were sent the pilot study questionnaire via email and received course credit when they returned the completed study. The pilot study was completed in three sections. The first section identified the number of exemplars that would be used to manipulate retrieval fluency. Here, participants were asked to write the names of as many active elderly people as possible (up to a maximum of 12) and provide a brief description justifying the inclusion of each person as an active elderly exemplar. Specifically, active people were defined as being:

Mentally alert and/or physically on-the-go. For example, someone would be considered mentally active if they were quick witted and enjoyed mental challenges like completing crossword puzzles; someone would be considered physically active if they were to enjoy occasional exercise like walking.

After retrieving as many active elderly people as possible, pilot study participants were asked to identify the specific exemplar that represented the point at which the retrieval
task became quite difficult for them to continue. This response could range from 1 (meaning the first exemplar was difficult to retrieve) to 12 (meaning the last exemplar was difficult to retrieve). The participants indicated that the mean exemplar difficulty was 6.78 ($SD = 1.98$). There were no differences in the retrieval rates between male and female participants.

The cut-off point of three exemplars was used for the few-condition. This number of exemplars is two standard deviations below the mean number of exemplars identified as difficult, and should ensure that the few-condition is easy to complete for the vast majority of participants. The cut-off point of nine exemplars was used for the many-condition. As participants in the pilot study were required to identify when the task became difficult, the many-condition was one standard deviation above the value where, on average, participants found the retrieval to be difficult.

Next, pilot study participants completed a trait rating task to assist in preparing the explicit and implicit outcome measures. For this task, a list of 77 words were compiled from previous stereotyping research investigating the elderly (e.g. Dijksterhuis, Spears & Lepinasse, 2001; Duval et al., 2000; Klauer, Wegener, & Ehrenberger, 2002; Redman & Snape, 2002). The Collin’s Thesaurus (Collins & Hands, 2002) was used to generate synonyms and antonymous of the words. This resulted in 30 words which were associated with the ‘active-inactive’ dimension (e.g. intelligent, energetic, brainless, and idle). The remaining 47 words were associated with other stereotypes of the elderly (e.g. conservative, wise) but not relevant to the to-be-retrieved inactive elderly stereotype. These were included to assess the applicability of the process of retrieval fluency. Pilot study participants were asked to rate how relevant each trait was for describing the elderly and how favourably they regarded each trait. All ratings were made on 7-point scales (1 = Not at all, 7 = Extremely). Selected stereotypical traits were rated as high in relevance (mean scores between 5 and 7); selected counterstereotypical traits were rated as low in relevance (mean scores between 1 and 3). This resulted in the following ‘active-inactive’ trait dimension words being selected for use as explicit rating scales: ‘frail’, ‘forgetful’, ‘slow’, ‘weak’, ‘fast’ and ‘strong’. The traits ‘active’ and ‘inactive’ were also included in the rating scales. The same relevance criterion (i.e., high and low mean scores) was used to select the retrieval-irrelevant stereotyping measures. This resulted in the following traits being used: ‘progressive’, ‘flexible’, ‘stubborn’, and ‘stupid’. The words ‘kind’, ‘likeable’ and ‘pleasant’ were used to compile the prejudice rating scale. These items were high in relevance (mean scores between 5 and 7) and high in favourability (mean scores between 5 and 7).

The IAT requires participants to categorise words related to the stereotypical concept of interest, in this case the active-inactive dimension. Six active-related words (‘dynamic’, ‘energetic’, ‘fast’, ‘lively’, ‘quick’ and ‘strong’) were rated by the pilot study participants as being poorly associated with the elderly (mean relevance scores below 3.5). Six inactive-
related words (‘forgetful’, ‘lethargic’, ‘slow’, ‘tired’, ‘weak’ and ‘weary’) were rated by the pilot study participants as being highly associated with the elderly (mean relevance scores above 5). This resulted in 12 stereotype-relevant adjectives (6 counterstereotypical and 6 stereotypical) that were included in the IAT.

The final section of Pilot Study 2 aimed at identifying 12 target pictures (6 young people and 6 elderly people) to be used in the IAT. I photographed a sample of 35 individuals (17 = young people; 18 = elderly people) who volunteered to have their image used for the purpose of this research. All volunteers were over 18 years of age. Photographs were taken indoors against a plain white wall. The image was taken two metres from the person using a digital camera with 1.4 mega pixels. All volunteers were asked to have a neutral expression, with their mouth closed (e.g., Brebner & Macrae, 2008). Images were cropped to show only head and neck and were transformed to a black and white image to increase standardisation.

Pilot study participants viewed each of the pictures and were asked to estimate the age of the person. The mean age estimates were checked to ensure that the ‘young’ pictures were perceived to be under 30 years of age and the ‘elderly’ pictures were perceived to be over 65 years of age. Eight pictures were removed from the sample because they had a mean age higher than 30 (for the young pictures) or lower than 65 (for the elderly pictures; see also, Nosek, Greenwald, & Banaji, 2005). After the age estimate, pilot study participants were asked to rate the attractiveness of each individual (1 = Not at all; 7 = Extremely). Attractiveness was measured to ensure none of the images were perceived to be extremely attractive or unattractive, and potentially distracting during the IAT. Four pictures were removed from the sample because they were not considered to be ‘average’ in attractiveness (i.e., mean attractive score was not between 2.5 and 4.5). The final consideration for picture inclusion was the person’s perceived typicality with reference to their age group. Previous research on categorisation has found that when individuals are perceived as typical they are more likely to be included in the cognitive representation of the group (e.g., Desforges et al., 1997) and the validity of the IAT relies on the prototypicality of its images. In addition, Locke, Macrae and Eaton (2005) found that exemplar typicality influences the time taken to verify category membership, with typical images facilitating faster categorisation than atypical images. To ensure the photos included in the IAT were perceived as typical, pilot study participants were asked to rate the photographs on typicality (1 = Not at all typical of old/young people; 7 = Extremely typical of old/young people). From the remaining photographs the three most typical pictures for each gender and age group (i.e. male, female, young, and old) were selected. For the final sample of selected images, see Appendix II.

**Exemplar generation task.** Study 1 participants were tested individually by a female experimenter and were randomly allocated to conditions. The experimental session began by seating participants in front of a computer and they were told that they would be required to
think about three (or nine) different active elderly people. When the computer programme was launched participants saw a screen that asked: “Please type the name, or other identifier, of an elderly person that you consider to be active”. At the bottom of the screen there was an instruction to press ‘Enter’ to continue with the next screen. This screen was displayed three or nine times according to the participant’s condition.

The time taken to enter each exemplar’s information was recorded. This latency measurement will serve as a manipulation check of retrieval fluency (for a review of the relationship between processing speed and fluency, see Winkielman et al., 2003). To avoid interfering with the way participants approached the task, participants were not informed of this recording until debriefing.

**Dependent measures.** Participants in the appended control condition did not complete the exemplar generation task and began their experimental session with the dependent measures. At this point, all participants were asked to complete explicit measures of group stereotyping on a paper-and-pen questionnaire. Participants rated elderly people on the eight traits identified in Pilot Study 2 using an 8 cm line (0 = Not at all, 8 = Extremely). Participants were asked to mark on the line where the ‘average’ elderly person would fall on each of the adjectives. The active-inactive traits identified in Pilot Study 2 were arranged in a single fixed order. A stereotypicality index was created by reverse scoring the counterstereotypical traits (i.e., ‘active’, ‘fast’, and ‘strong’) and then averaging these responses with ‘frail’ and ‘slow’ (‘forgetful’, ‘inactive’ and ‘weak’ were weakly correlated with the other traits and were therefore excluded). Cronbach’s alpha for the index was .78. High scores on this inactive stereotypicality index represented more stereotypical responses. The inactive stereotypicality index was then subtracted from the mean of the control group (dependent measures only; M = 5.63, SD = .86) to create a differential score representing stereotype change from a baseline. After this data manipulation, positive scores indicate greater stereotype change towards a view of the elderly as active while negative scores indicate an increased reliance on the stereotype of inactive elderly people.

The retrieval-irrelevant stereotyping traits were internally consistent (Cronbach’s α = .71) and, thus, averaged in an appropriate retrieval-irrelevant stereotypicality index. As with the inactive stereotypicality index, this index was subtracted from the mean of the control group (M = 4.41, SD = 1.08) to create a measure of stereotype change.

Following the group stereotyping ratings, participants made dispersion ratings on the same 8 cm scales used for the stereotype change index. Participants were asked to mark two slashes to indicate where they believed the two most extreme group members would fall on each of the traits (e.g., Paolini et al., 2004b). The distance between the two slashes (range) provides an indication of group dispersion (i.e., the greater the range the greater the perceived dispersion of the group, hence, the weaker the stereotype; Park & Judd, 1990). The dispersion
index comprised of the same five traits included for the inactive stereotypicality index. Cronbach’s alpha for the resulting index was .80. A differential score was then created by subtracting the mean of the control group (dependent measures only; $M = 3.73$, $SD = 1.44$) from the scores of the experimental conditions to represent change in dispersion from a baseline. After this data manipulation, positive scores on the dispersion change index will indicate greater perceived dispersion while negative scores will indicate a decrease in perceived dispersion.

Prejudice ratings were completed by asking participants to rate how characteristic each of the words ‘kind’, ‘likeable’ and ‘pleasant’ were of the elderly on 8cm lines ($0 = \text{Not at all}, 8 = \text{Extremely}$). These positive traits were reverse scored because prejudice is a negative affective state (Paolini et al., 2004a). Responses to these items were internally consistent (Cronbach’s $\alpha = .71$). A differential score was again created by subtracting the mean of the control group (dependent measures only; $M = 2.19$, $SD = 1.20$) from these scores to represent change in prejudice from a baseline. After this data manipulation, positive scores indicate a decrease in group prejudice (or ‘ageism’), while negative scores indicate an increase in prejudice. The prejudice measure will also be used to assess the easy-thus-liked heuristic.

After completing the explicit measures, participants completed the IAT. For this task, participants returned to the computer and were informed that they would be required to sort pictures and words into different categories as quickly and accurately as possible. As in standard IAT tasks (e.g., Greenwald et al., 1998), participants completed the IAT in 5 blocks of trials. Detailed instructions were provided on the computer screen at the start of each block. In the first block (initial picture discrimination task) participants were asked to assign the 6 pictures of old people and the 6 pictures of young people to the categories “Old” (labelled in the top left hand corner of the screen) or “Young” (labelled in the top right hand corner of the screen). Participants were asked to press a left-hand key “X” when a picture of an old person appeared on the screen and a right hand key “N” when a picture of a young person appeared. In Block 2 (inactive-active trait task) participants were asked to assign 6 inactive traits and 6 active traits to the categories “inactive” (key “X”) and “active” (key “N”). In Block 3 (initial picture-trait combined task), pictures and traits were combined in a stereotype-consistent manner. Participants were asked to press the left hand key when either an old picture or an inactive trait was presented and the right hand key when either a young picture or an active trait was presented. In Block 4 (reversed picture task), participants were asked to press the left hand key when a young picture appeared and press the right hand key when an old picture appeared. In the final block (reversed picture-trait combination task) picture and traits were again combined, but this time in a stereotype-inconsistent manner. Participants were asked to press the left hand key when either a young picture or an inactive trait was presented and the right hand key when either an old picture or an active trait was presented (for a similar IAT
procedure, see Gawronski, Gescke, & Banse, 2003). A positive difference (adjusted by pooled standard deviations) in reaction times between the blocks of stereotypical (Block 3) and counterstereotypical (Block 5) pairings indicates the existence of an implicit association between the opposing out-groups old-young and the inactive-active dimension (for a similar scoring algorithm, see Greenwald et al., 2003). In this case, a positive difference will reflect an implicit association with the old as inactive and the young as active.

After the IAT task, the final computerised task was loaded, which displayed the name of each exemplar the participant had generated during the retrieval task. Participants were asked to reconsider each exemplar they generated and answer some additional questions aimed at assessing alternative mediational accounts of retrieval fluency effects. Firstly, participants were asked to provide a brief description of the generated exemplars. The purpose of the exemplar description was to assess whether the exemplars became less vivid as the task became progressively more difficult. The number of words that make up the description of each exemplar were counted as a proxy of vividness. After each exemplar description, participants rated the exemplar typicality. Participants were asked to indicate how typical they considered the exemplar to be of elderly people in general (1 = Not at all, 7 = Extremely). Responses to the exemplar typicality items were averaged across exemplars to form a typicality index (Cronbach’s α = .87) to be used to assess the easy-thus-typical heuristic. Next, participants were asked if the exemplar was famous or known to them personally. This question allowed for an assessment of category shifts. Category shifts was calculated by counting the number of times a participant switched categories (i.e., if the first exemplar was known personally and the second exemplar was famous, this represented one category shift). The individual exemplar questions were repeated for each of the generated exemplars. Following the individual exemplar questions, participants completed demographic measures, were debriefed, and thanked for their participation.

Results

Preliminary Analyses

Preliminary analyses revealed no effects of gender and so this variable was no longer considered. The responses on the IAT were screened following a procedure similar to Greenwald et al. (2003). Firstly, Greenwald et al. suggested that responses greater than 10,000 ms be removed. As their data was based on a large web sample, I used a more conservative approach and replaced them (N = 6) with a score of 10,000 ms. Greenwald et al. also eliminated cases with more than 10% of the trials having a response latency of less than 300 ms, as this might reflect participants guessing the response. No participants were eliminated using this criterion. Greenwald et al.’s modified IAT task uses 7 blocks (the five blocks that I used plus two practice trials for Blocks 3 and 5). As I used the original five block version of
the IAT (Greenwald et al., 1998), I adapted the scoring algorithm recommended by
Greenwald et al. (2003) by removing the step of the algorithm involving the practice
responses. To identify the implicit association score, the mean response latency for Block 3
was subtracted from the mean response latency for Block 5. To adjust for the variability of the
data from which the means are computed, the difference score was divided by the pooled
standard deviation of reaction times on the congruent and incongruent blocks in line with
Greenwald et al. (2003). This resulted in the $D$ statistic which is similar to Cohen’s $d$ measure
used for power estimates (Greenwald et al., 2003, p. 208). On this scale $d = 0.8$ is considered
a large effect size (Greenwald et al., 1998).

IAT effects detected in the control group was tested using a one-sample $t$-test against
zero and found to be significant, $t(15) = 8.59, p < .001$. As expected a basic IAT effect was
found in the baseline (no counterstereotypical exemplar) condition. To maintain consistency
with the explicit dependent measures an implicit stereotype change index was created by
subtracting the experimental participants’ $D$ values from the mean of the control group ($M =
.75, SD = .35$). After this data manipulation, positive scores indicate a change in implicit
stereotyping relative to the baseline and negative scores indicate an increase in implicit
stereotyping.

**Manipulation Check: Retrieval Fluency**

It was expected that participants in the nine-exemplar condition would take longer to
generate the required number of active elderly exemplars than participants in the three-
exemplar condition. The total amount of time taken to generate the required number of exemplars was submitted to a one-way ANOVA as an indirect manipulation check of task ease. The time difference between the three-exemplar and nine-exemplar conditions was highly significant, $F(1, 30) = 27.07, p < .001$. The three-exemplar condition took a mean time of 136 seconds ($SD = 64.82$ seconds) to complete the task and the nine-exemplar condition took a mean time of 268 seconds ($SD = 78.00$ seconds).

Because the difference in response latency could be interpreted as assessing task
‘length’ rather than task difficulty, the mean number of category shifts was used as an
additional manipulation check of task ease. This was identified by counting the number of
times a participant swapped between personally known and famous exemplars. The mean
number of category shifts was expected to increase as the task became more difficult (for a
similar point, see Kunda & Thagard, 1996). Category shifts for participants in the three-
exemplar condition were significantly less frequent ($M = .31, SD = .60$) than for participants
in the nine-exemplar condition ($M = 1.69, SD = 1.50$), $F(1, 30) = 11.67, p = .002$. This result
indicates that participants in the nine-exemplar condition needed to change category more
often than participants in the three-exemplar condition to complete the task.
Overall, the retrieval fluency manipulation was successful. Participants in the three-exemplar condition completed the exemplar generation task more quickly and required less category shifts than participants in the nine-exemplar condition.

**Testing Basic Effects: Group Judgments**

Means and standard deviations for all outcome measures are shown in the Table 5.1. The predicted retrieval fluency effect emerged on the explicit stereotype change index, \( F(1, 30) = 9.18, p < .005 \). Participants in the three-exemplar condition changed their group stereotype more (\( M = 1.77, SD = 1.23 \)) than participants in the nine-exemplar condition (\( M = .54, SD = 1.07 \)).

A test of change between the three-exemplar and nine-exemplar conditions against the control condition was conducted using a one-sample \( t \)-test against the baseline of 0 on the explicit stereotype measure. There was a significant change in the three-exemplar condition, \( t(15) = 5.75, p < .001 \), and a marginally significant change in the nine-exemplar condition \( t(15) = 2.02, p = .06 \). To my knowledge, this is the first retrieval fluency study to demonstrate that participants from both the few-condition and many-condition included the retrieved content in their judgment but participants in the few-condition did so to a greater extent.

### Table 5.1

*Change in Group Measures Relative to Baseline as a Function of Number of Exemplars (SDs shown in parenthesis)*

<table>
<thead>
<tr>
<th>Group Measure</th>
<th>Number of Exemplars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Three</td>
</tr>
<tr>
<td>Explicit stereotype</td>
<td>( 1.77_a^{**} (1.23) )</td>
</tr>
<tr>
<td>Implicit stereotype</td>
<td>( .31_a^{**} (.43) )</td>
</tr>
<tr>
<td>Retrieval-irrelevant stereotype</td>
<td>( .66^* (1.31) )</td>
</tr>
<tr>
<td>Dispersion</td>
<td>( .35 (1.03) )</td>
</tr>
<tr>
<td>Prejudice</td>
<td>( .08 (1.14) )</td>
</tr>
<tr>
<td></td>
<td>Nine</td>
</tr>
<tr>
<td>Explicit stereotype</td>
<td>( .54_b^* (1.07) )</td>
</tr>
<tr>
<td>Implicit stereotype</td>
<td>( .03_b (.34) )</td>
</tr>
<tr>
<td>Retrieval-irrelevant stereotype</td>
<td>( .32 (1.07) )</td>
</tr>
<tr>
<td>Dispersion</td>
<td>( .02 (1.48) )</td>
</tr>
<tr>
<td>Prejudice</td>
<td>( .14 (1.35) )</td>
</tr>
</tbody>
</table>

*Notes.* Subscripts within rows indicate statistically different means. **Significantly different from zero at \( p < .01 \); */p = .06.*

To test retrieval fluency on implicit group stereotyping, the implicit stereotype change index was subjected to a number of exemplars (three vs. nine) one-way ANOVA. A significant main effect once again emerged, \( F(1, 30) = 3.99, p = .05 \). Participants in the three-exemplar condition displayed a larger change in implicit stereotyping (\( M = .31, SD = .43 \)) than participants in the nine-exemplar condition (\( M = .03, SD = .34 \)), suggesting that retrieval fluency results in implicit stereotype change. As the results on the implicit and explicit group
stereotyping measures follow similar patterns (Pearson’s $r = .22$, $p = .07$, one-tailed), this pattern of results suggests that retrieval fluency effects are automatic in nature.

A test of change on the three-exemplar and nine-exemplar conditions was conducted using a one-sample $t$-test against the baseline of 0 on the implicit stereotype change index. There was a significant change in the three-exemplar condition, $t(15) = 2.85$, $p = .01$, but no significant change from the baseline in the nine-exemplar condition, $t(15) < 1$.

There were no significant effects of number of exemplars on the outcome measures of change in dispersion, prejudice, or retrieval-irrelevant stereotyping, all $F$s < 1. This suggests that the effect is limited to judgments of group stereotyping that are strictly related to the retrieved content and does not extend to perceptions of group dispersion and liking.

**Testing Mediating Processes**

Prior to carrying out a full mediation analysis on the easy-thus-liked and the easy-thus-typical heuristics indices, three preliminary steps were taken to gauge the necessity of such analysis (for a similar approach, see Rubin et al., 2010). Firstly, the continuous mediating variable (i.e. likeability or typicality) should show a similar pattern as the outcome variable (i.e., explicit stereotype change; Aiken & West, 1991). Secondly, when the potential mediating variable is entered as a covariate on the stereotype change analysis, there should be a sizeable decrease in the significance of the effect. Thirdly, there should be a correlation between the continuous variables, and the outcome variable. If these three preliminary criteria all hold, then a full mediation analysis is warranted to explain some of the variation in the dependent variable (Aiken & West, 1991). The results from these preliminary steps are reported for the easy-thus-liked and easy-thus-typical heuristics.

**The easy-thus-liked heuristic.** The easy-thus-liked heuristic was tested using the change in group prejudice index. The first preliminary test compares the results from the one-way ANOVA with the stereotype change index against the results from the one-way ANOVA using the prejudice index. As noted above, the effect detected on group stereotyping ($F(1, 30) = 9.18, p < .005$) was not detected on the prejudice index ($F < 1$). Hence, it is evident that these measures are not operating in the same direction. The second preliminary test used group prejudice as a covariate in the analysis on explicit stereotype change. There was no sizeable decrease in significance, from $F(1, 30) = 9.18, p < .005$ to $F(1, 29) = 8.65, p = .006$. The third step revealed that there was no correlation between the prejudice index and stereotypicality, $r = -.11$, $ns$. From this preliminary assessment it is apparent that there is an insufficient basis to test a full mediational model (Baron & Kenny, 1986). Hence, there was no indication of a mediating role of group prejudice on the explicit retrieval fluency effect.

**The easy-thus-typical heuristic.** To assess whether the easy-thus-typical heuristic was operating as a mediating variable, the typicality index was tested. The number of exemplars
(three vs. nine) ANOVA was non-significant on this index, $F < 1$. Secondly, when typicality was entered as a covariate in the one-way ANOVA on the measure of stereotypicality, no sizeable change in the effect of exemplars was detected, from $F(1, 30) = 9.18, p < .005$, to $F(1, 29) = 8.87, p = .006$. There was no correlation between typicality and stereotype change, Pearson $r = -.07, ns$. This preliminary analysis precludes the need for further investigation and suggests that changes in typicality were unable to capture the underlying processes responsible for the explicit retrieval fluency effect.

**Exemplar vividness.** Vividness was assessed to ensure that the exemplars retrieved towards the end of the task were not significantly different from those retrieved at the beginning of the task. The length of the descriptions varied from 1 to 27 words ($M = 5.61, SD = 3.53$). Amongst participants in the nine-exemplar condition, the mean vividness of the first three exemplars was compared with the mean vividness of the final three exemplars (for a similar procedure, see Haddock et al., 1996; Haddock et al., 1999). A paired samples $t$-test revealed a non-significant difference between the vividness of the first three exemplars ($M = 6.56, SD = 3.96$) and the last three exemplars ($M = 5.81, SD = 4.90$), $t(15) = 1.20, ns$. This finding indicates that as the task became progressively more difficult it did not cause a significant deterioration in the vividness of the recalled exemplars.

**Discussion**

Study 1 had three primary goals: (1) to develop and use the self-generation paradigm as a new test of member-to-group generalisation; (2) to extend the theoretical understanding of retrieval fluency by investigating the automatic-controlled nature of the meta-information; and (3) to test the relevance of affective and/or cognitive explanations for the retrieval fluency effects. The findings relevant to each of these goals will now be discussed.

**The Self-Generation Paradigm: Meta-Information and Member-to-Group Generalisation**

The first goal of this study was to design and test a paradigm that included internal meta-information within a member-to-group framework. This was achieved by modifying the standard impression formation paradigm so that participants were asked to retrieve, instead of being presented with, out-group sample information prior to expressing out-group judgments. The self-generation paradigm successfully adopted the few-versus-many approach to manipulating retrieval fluency (e.g., Schwarz et al., 1991; for a review, see Schwarz, 1998). The detailed piloting procedure ensured that the self-generation paradigm successfully manipulated retrieval fluency and measured for generalisation.

I argued that out-group judgments are built from two sources of information: (1) the actual content and (2) the meta-information about the content. A problem with relying on an impression formation paradigm to investigate meta-information is the need to artificially
inject the meta-information through the provision of experimenter-controlled cues (see Chapter 3). The use of a self-generation paradigm avoids this problem by requiring participants to retrieve out-group sample information from their memory, thus creating a situation where the information (i.e., number of exemplars) and the meta-information (i.e., retrieval fluency) are potentially, but naturally, in direct opposition as the difficulty of the task increases.

The manipulation of internal meta-information was expected to influence the size of generalisation, whereby participants experiencing low retrieval fluency (nine-exemplars) would generalise the sample information less than the participants experiencing high retrieval fluency (three-exemplars) even though the disconfirming sample size was objectively larger in that condition. Consistent with the prediction that meta-information plays an important role in out-group judgments, a significant retrieval fluency effect emerged on the explicit stereotype change index. This is consistent with traditional findings from the retrieval fluency literature where participants experiencing high fluency include more of the retrieved content in their judgments than participants experiencing low fluency (e.g., Broemer, 2001; Dijksterhuis et al., 1998; Haddock, et al., 1999; Rothman & Hardin, 1997; Schwarz et al., 1991; Wänke et al., 1996). This finding also extends research by Rothman and Hardin (1997) by demonstrating the importance of considering the impact of meta-information in out-group judgments.

The parallel results on the explicit and implicit measures of stereotype change point towards an automatic effect of retrieval fluency. This represents a significant extension of previous fluency research. This finding is consistent with a dual process model of retrieval fluency effects. Dual process models (e.g., Chaiken, Liberman, & Eagly, 1989; Cacioppo, Petty, Kao, & Rodriguez, 1986) describe two parallel, yet interactive, approaches to information processing: (1) heuristic processing, typified by a ‘quick and easy’ reliance on surface features (i.e., subjective experience), and (2) systematic processing, typified by attention to both quality and quantity of the retrieved information (i.e., the retrieved content). From this perspective, people should use heuristic processing and rely on retrieval fluency when they require a rapid ‘automatic’ assessment of judgment-relevant material (Ruscher, Hammer, & Hammer, 1996), but should use systematic processing and rely on the retrieved content when a more accurate or ‘controlled’ assessment of the material is triggered (Fiske & Neuberg, 1990; Verplanken, Jetten, & van Knippenberg, 1996). Hence, according to a dual process account, people may use heuristic processing strategies (e.g., easy-thus-liked and/or easy-thus-typical) in an automatic fashion, but may undergo a systematic processing strategy when appropriate cues are available in the environment.

The findings of Study 1 suggest that retrieval fluency is an automatic process and, according to my dual-process account, this automatic process should be able to be interrupted
through the provision of specific cues. To test this dual-process account directly, in Study 2, some participants will be provided with an external meta-information cue to encourage systematic processing. Specifically, some participants will be subtly alerted to the specific nature of the task by being provided with a cue that warns them that the retrieval task is easy or difficult. The provision of this cue is expected to encourage a systematic processing strategy whereby the participants will use controlled processing and rely on the retrieved content when completing their out-group judgments.

In order to test the applicability of retrieval fluency as an automatic process (Anderson et al., 2007), a measure of change on retrieval-irrelevant stereotypes was also included. As predicted, there were no retrieval fluency effects on the retrieval-irrelevant stereotype measure (see also, Gross & Hardin, 2008; Hardin & Rothman, 1997). That is, Study 1 demonstrated that the retrieved material needs to be relevant to the subsequent judgment for meta-cognitive information to produce an effect. Subsequent studies will focus solely on retrieval relevant group stereotypes.

In the following study, I will introduce a third test for my dual-process model that includes a cue for controllability (Andersen et al., 2007). For now, parallel results on the explicit and implicit measures as well as evidence supporting the applicability of retrieval fluency would suggest that retrieval fluency is an automatic process.

The index measuring change in group dispersion was expected to mimic the findings on the explicit stereotype change measures, but, these predictions were not supported. The use of a range measure, however, may have been unsuitable to measure change in dispersion within this paradigm. Asking participants to retrieve counterstereotypical exemplars may have encouraged all participants to access extreme members of the out-group, resulting in no appreciable difference between experimental conditions. To address this criticism, Study 2 will use a similarity measure to assess perceived group variability.

The outcome measure of prejudice change did not show any effects of retrieval fluency, suggesting that prejudice towards the elderly is somehow unrelated to the perception of the elderly as inactive. Although dissociations between prejudice and stereotyping are not uncommon in the literature (for a review, see Mackie & Smith, 1998), this specific dissociation might also be attributed to the role of the matching hypothesis presented in Chapter 2 (also, see Paolini et al., 2007). The findings from Meta-analysis Study 1 revealed a general advantage in the impression formation literature for generalisation effects to emerge on cognitive outcome measures. In this study, the ‘cognitive’ nature of the retrieval task might limit generalisation effects to ‘cognitive’ outcome measures (such as stereotype change) and not extend to affective measures (such as prejudice). The stability of this result will be assessed with more research using the same out-group.
**Explaining the Nature of the Meta-Information**

Tormala et al. (2002) argued that several variables may mediate the retrieval fluency effect. I now summarise the results from the analysis of the mediating variables. The assessment of exemplar vividness revealed that the retrieval fluency effect detected in the group judgments was not due to the nine-exemplar participants relying on less vivid exemplars towards the end of the retrieval task. This means that the nine-exemplar participants were not experiencing higher stereotyping because of the different number of retrieved exemplars. It is important to note, however, that the length of the description task was limited in this study by a cap in the computer programming. As a result, participants were unable to write descriptions that exceeded approximately three sentences. This may have limited sensitivity in the assessment of exemplar vividness. The test of exemplar vividness will be improved in the following study by removing this programming limitation.

In this study, I included two mediating variables: likeability and typicality. Support for the easy-thus-liked heuristic was expected to reveal that retrieval fluency operates along an affective dimension while support for the easy-thus-typical heuristic was expected to demonstrate that retrieval fluency operates along a cognitive dimension. There was no mediating support for either heuristic in this study. In regards to the proposed easy-thus-liked heuristic, the lack of mediating evidence may be tied to the specific out-group. In this study, the elderly were perceived as being low on the ‘negativity’ dimension. Recall that the no-exemplar control participants indicated a mean prejudice score of 2.19 (SD = 1.20) on an 8 cm scale. The low baseline score may not allow sufficient room for the easy-thus-liked heuristic to emerge. In regards to the proposed easy-thus-typical heuristic, it may be necessary to extend the typicality index to an assessment of the perceived frequency (or amount) of active elderly exemplars. By definition, a typical exemplar is also perceived to be frequent; therefore including a measure of perceived frequency may extend the power of this measure to capture variations in typicality within the self-generation paradigm.

**Strengths and Limitations**

In this study, retrieval fluency affected member-to-group generalisation in a way consistent with the hypotheses. On measures of explicit and implicit group stereotyping participants retrieving three counterstereotypical exemplars demonstrated greater stereotype change than participants retrieving nine counterstereotypical exemplars.

There are a number of important merits of this research. Firstly, by including a no-exemplar control group, this was the first retrieval fluency study to demonstrate that the few and many-exemplar conditions both produce judgment change or generalisation. That is, low fluency does not result in cognitive exclusion of the exemplar information under conditions of low retrieval fluency, generalisation just occurs to a lesser extent than under conditions of
high retrieval fluency. This finding is different to the research by Sanna et al. (2009), where the difficult retrieval condition showed a contrast effect compared to the baseline condition. It is important to note that Sanna et al.’s research is in the area of hindsight bias which is a more abstract judgment domain than out-group judgments. Therefore, the ambiguity of hindsight-based judgment tasks may be more heavily influenced by automatic meta-cognitive cues than out-group judgments.

Secondly, change on the IAT demonstrates that implicit stereotypes are amenable to change meta-cognitively, and adds to the growing body of research that suggests implicit attitudes are malleable (e.g., Blair et al., 2001; Sinclair, Hardin, Lowery, & Colangelo, 2005; for a review, see Blair, 2002). This finding also challenges traditional models of stereotyping, such as the prototype model, which view stereotypes as permanent cognitive structures impervious to temporary thoughts and goals such as those activated by my manipulation of retrieval fluency (see also, Bargh, 1999; Devine & Monteith, 1999; Dunton & Fazio, 1997; Fazio, Jackson, Dunton & Williams, 1995; Fiske, 1989; Hamilton & Sherman, 1994; Plant & Devine, 1998). In Chapter 2, I argued that Bless and Schwarz’s (1998) inclusion-exclusion model was potentially able to incorporate the findings and conclusions drawn from the meta-analytic review. The change on implicit stereotyping found in this study adds further weight to the validity of this model, which views stereotypes as fluid cognitive structures that change on-line according to the nature of the information immediately available (for further evidence of the fluidity of stereotypes, see Garcia-Marques et al., 2006).

Despite the success in demonstrating generalisation and meta-cognitive effects through the use of the self-generation paradigm, I was unable to explain the retrieval fluency effect. Hence, I was unable to isolate the psychological process responsible for the effect. Study 2 addresses this limitation by implementing key changes in the measurement potential mediators.

Conclusions

Study 1 shows a sizeable influence of meta-cognitive processes on member-to-group generalisation. It was demonstrated that people use internal meta-information to inform the extent of exemplar inclusion in subsequent group judgments. This is important for generalisation research because it steps away from the idea of permanent stereotype change, and suggests that people use transient internal meta-information when forming a group judgment. In the following study, I will investigate the competing predictions from the traditional disconfirming sample size effect and the retrieval fluency effect. This will be achieved through the incorporation of a presented information condition in Study 2.
CHAPTER 6

CONTRASTING APPROACHES TO MEMBER-TO-GROUP GENERALISATION:
SELF-GENERATED VS. PRESENTED INFORMATION (STUDY 2)

In Study 1, retrieval fluency was found to influence member-to-group generalisation on implicit and explicit stereotype change measures. Specifically, generalisation was greater when participants retrieved three counterstereotypical exemplars than when participants retrieved nine counterstereotypical exemplars from their memory. This retrieval fluency effect is in direct contrast with the bookkeeping effect from impression formation research (Johnston & Hewstone, 1992; Hewstone et al., 1994; Weber & Crocker, 1983). In Chapter 2, I provided meta-analytic evidence for the bookkeeping effect whereby generalisation was shown to increase as the sample of disconfirming exemplars increased (e.g., Nisbett, et al., 1983; Tversky & Kahneman, 1973; see also Fielder & Walther, 2004; Chapter 2). To demonstrate that both the retrieval fluency effect and the bookkeeping effect are valid, in Study 2 I will contrast the traditional impression formation approach to stereotype change with the new self-generation approach to stereotype change. Study 2 will also extend the investigation of ‘automatic’ retrieval fluency effects by including a cue designed to facilitate controlled cognitive processing. The cue is expected to interrupt, or challenge, the automatic retrieval fluency effect. Finally, this study will continue to explore the mediating processes that may explain retrieval fluency effects.

Impact of the Meta-Information

The competing bookkeeping effect and retrieval fluency effect will be tested by manipulating the information source of the counterstereotypical exemplars. Participants will be required to either self-generate or read about active elderly individuals before judging elderly people in general. The procedure for the self-generated condition will follow that of Study 1. While participants in the presented information condition will be told that previous participants had provided the information about the active elderly people. The number of exemplars will again be manipulated by exposing some participants to three counterstereotypical exemplars and others to nine counterstereotypical exemplars. Change in group stereotyping will be the primary outcome measure. Dispersion was measured in Study 1 using a range measure and no effects of retrieval fluency emerged. To improve on this, I will use a cognitively less demanding similarity measure, where participants indicate how
similar they perceive the elderly to be on each active-related trait (for a similar measure, see Park et al., 2001). Change in group prejudice will again be measured.

Wänke et al. (1996, Study 1) investigated information source by manipulating retrieval versus presented information (see also Sanna et al., 2009). However, Wänke et al. investigated attitude judgments, not member-to-group generalisation per se. Participants were asked to self-generate or read arguments either supporting or opposing the use of public transport. The number of arguments was also manipulated, with participants being exposed to either three or seven arguments. In the self-generation condition, Wänke et al. found a retrieval fluency effect. Participants generating three arguments were more extreme in their judgment towards the use of public transport than those generating seven arguments (i.e., a standard retrieval fluency effect). On the other hand, in the presented condition there was support for the law of large numbers or a disconfirming sample size effect (Nisbett et al., 1983). Participants presented with seven arguments were more extreme in their judgment towards the use of public transport than those in the three argument condition. Wänke et al. explained that these effects reflected the subjective experience that accompanies thoughts associated with attitude construction.

Building on Wänke et al.’s (1996) findings, and the contrasting patterns identified in Meta-Analysis Study One and the self-generation Study 1, I expected to detect a significant interaction between information source and number of exemplars on group judgments. Participants in the self-generation condition were expected to replicate the findings of Study 1 and demonstrate a retrieval fluency effect, with those self-generating three disconfirming exemplars demonstrating greater stereotype change than those self-generating nine disconfirming exemplars. On the other hand, participants in the presented condition are expected to demonstrate a bookkeeping effect, or disconfirming sample size, with those presented with three disconfirming exemplars demonstrating weaker stereotype change than those presented with nine disconfirming exemplars.

The predictions from these two opposing research traditions (i.e., presented information vs. self-generated information), in my view, really differ in terms of a single critical factor: the meta-information associated with the exemplar information. I expect people to have a naïve appreciation of the law of large numbers. Under presented conditions, this meta-cognitive strategy should encourage greater inclusion of the exemplar information (Nisbett et al., 1983; also see Chapter 3). On the other hand, in the self-generation condition the meta-information available during the retrieval of many exemplars should restrict inclusion of the exemplar information probably because subjective retrieval difficulty flags sample atypicality or dislike.
Challenging Automatic Generalisation Effects through a Congruent Task Cue

There is evidence to suggest that both the retrieval fluency and bookkeeping effects occur as a default, or automatic, generalisation response to available group exemplars. In Chapter 5, I presented evidence of an implicit change in group stereotyping to support the automatic inclusion of retrieval fluency meta-information in out-group judgments. In Chapter 2, I presented meta-analytic evidence suggesting that exemplar information is included in out-group judgments as a default, or automatically (also see Schwarz & Bless, 1992). In an effort to challenge these automatic effects, I will provide some participants with a meta-information cue that alerts these participants to the biasing effects of meta-information (for a similar point, see Fielder & Walther, 2004; Martin, 1985; Martin, 1986; Martin & Stapel, 1998).

Research from a variety of fields, such as memory (Schooler & Engstler-Schooler, 1990), perception (Wyatt & Campbell, 1951), attractiveness (Sigall & Ostrove, 1975), all indicate that external meta-cognitive cues can reduce unwanted automatic effects. For example, Schwarz and Clore (1983) found that an external meta-information cue influenced subsequent judgments. In a well-known study, participants living in Chicago were interviewed over the phone about their general life satisfaction. Participants were found to have higher life satisfaction on a sunny day compared to an overcast day, in line with basic mood-congruent effects. However, if the interviewer cued participants about the ‘biasing’ impact of the weather on their judgment by asking in the opening sentences of the interview “how is the weather in Chicago?”, the effect of the weather on life satisfaction judgments was wiped off. According to Schwarz and Clore when people become aware of a biasing factor they will try to correct for it. Put differently, a subtle meta-information cue may serve to undermine the perceived validity of the automatic meta-information effects. In the present study, the presentation of an external meta-information cue was expected to follow a similar pattern to Schwarz and Clore’s and wipe off the automatic effects of the meta-information in both presented and self-generated conditions.

In this study, I will present a meta-information cue to some participants prior to, and immediately after, exposure to the counterstereotypical exemplars. The cue will be congruent with the experience of retrieval. For example, participants self-generating nine-exemplars presented with a meta-information cue will read: “This task is (was) quite difficult. Most people find (have found) this hard to complete” (for similar manipulations, see Rothman & Hardin 1997, Study 2; Schwarz et al., 1991, Study 2). Participants self-generating three-exemplars presented with a meta-information cue will read that the task is (was) easy. Presented information participants who receive a meta-information cue will learn that previous participants had found the retrieval task difficult in the nine-exemplar condition or easy in the three-exemplar condition (recall the cover story for presented information participants indicated that the exemplar information was generated by previous participants).
Self-generation research by Schwarz et al. (1991, Study 2) supports the prediction that the presentation of a congruent meta-information cue will wipe off the retrieval fluency effect. In their study, participants in the easy (hard) retrieval condition were told that previous participants had found the task easy (difficult) to complete. Participants in the meta-information cue condition did not demonstrate a retrieval fluency effect. Schwarz et al. argued that the meta-information cue undermined the participants’ subjective experience of recall encouraging them to disregard the feelings associated with the retrieval process (for reviews see, Schwarz et al., 1998; Wänke et al., 1995). Building on this research, I expect that self-generation participants will demonstrate a null effect of number of exemplars when they are exposed to an external meta-information cue.

Impression formation research by Paolini et al., (2009, for a detailed description, see Chapter 3) also supports the nullifying influence of a meta-information cue. In their research, Paolini et al. provided participants with a meta-judgmental cue questioning the validity of the exemplar information and found that the generalisation of the presented counterstereotypical information to out-group stereotype was also removed. These authors argued that nullification of generalisation occurred (at least partly) because the meta-judgmental cue encouraged the participants to cognitively exclude the counterstereotypical information from their group judgment. Building on this research, I expect that presented information participants will also demonstrate a null effect of number of exemplars when they are provided with an external meta-information cue.

The demonstration of a nullification of the retrieval fluency and bookkeeping effects following the presentation of an external meta-information cue will ultimately support my dual process account of member-to-group generalisation. According to my dual process account, people use information in a heuristic fashion unless cues in the environment encourage a controlled processing approach (for a similar information processing model, see Chaiken et al., 1989; Cacioppo et al., 1986). In Study 1, the evidence of an implicit retrieval fluency effect supports the role of a heuristic or automatic processing approach. In this study, providing participants with a meta-information cue should undermine the perceived validity associated with (retrieval and bookkeeping) meta-information. As a result, participants should demonstrate a controlled processing strategy that accounts for the biasing nature of these automatic effects.

**Testing Multiple Mediating Processes**

In an effort to develop a clear understanding of retrieval fluency, I will again investigate the affective easy-thus-liked heuristic and the cognitive easy-thus-typical heuristic as mediating variables. I modified the measure of likeability from Study 1 by using a feeling thermometer (e.g., Dasgupta & Greenwald, 2001; Gross & Hardin, 2007). I modified the
measure of typicality by including a separate measure of the perceived frequency of active elderly people (for evidence supporting the link between frequency and typicality, see Collins & Loftus, 1975; Kunda & Thagard, 1996). Exemplar vividness will again be measured. As an improvement from Study 1, there will be no limitation on the length of the descriptions provided by the participants. In addition to these modifications, I will investigate the role of unrequested cognitions and perceived prior contact as explanatory processes for the impact of meta-information across the self-generated and presented information conditions.

**Unrequested cognitions.** Tormala et al. (2007) investigated the role of unrequested cognitions as an explanation for retrieval fluency effects. In their Study 1, Tormala et al. informed participants at Ohio State University that the University was considering a policy that required all seniors to pass comprehensive exams in order to graduate. To ensure that the task was highly relevant, the participants were led to believe that they were part of a select group to inform the policy decision makers. Furthermore, participants were told that if the exams were endorsed all students would sit these new exams the following year. Following this cover story, participants were asked to generate either 2 or 10 arguments in support of the alleged proposal. Following the supportive argument generation, the participants indicated on a 9-point scale the number of unsupportive arguments, or unrequested cognitions, which came to mind during the argument generation task (0 = 0 negative thoughts; 1 = 1 negative thought ... 8 = 8 or more negative thoughts). The capping of data collection at ‘8 or more’ unrequested cognitions is conservative, especially considering that participants in the difficult retrieval condition were asked to generate 10 supportive arguments. Despite this possible ceiling effect, Tormala et al. demonstrated that unrequested cognitions were significantly greater in the many condition and that unrequested cognitions mediated a basic retrieval fluency effect on the judgment of the proposed examination policy.

The importance of Tormala et al.’s (2007) finding for my self-generation paradigm is easy to appreciate: It is possible that retrieval fluency effects detected in Study 1 are due to an increased accessibility of unrequested cognitions (i.e., stereotypical exemplars) amongst participants in the nine-exemplar condition. Importantly, this mediating variable may also be involved in the effect expected in the presented condition. If people are motivated to defend their stereotypical views (Moreno & Bodenhausen, 1999), participants might react to the presentation of counterstereotypical information by self-generating stereotypical exemplars. Hence, unrequested cognition might play a role also among presented participants.

**Perceived prior contact.** The self-generation paradigm draws on the previous contact experiences of an individual, but does not expose participants to any immediate contact. Contact research has found that increased contact with out-group members results in reduced negative out-group judgments (for a review, see Pettigrew & Tropp, 2000; 2006). Asking some participants to ‘think’ about their experience with elderly people (vs. other participants
being presented with active elderly people) may enhance the accessibility of prior contact experiences with members of the out-group which, may contribute to enhanced generalisation (e.g., Schwartz & Simmons, 2001). Similarly, presenting some participants with more disconfirming exemplars (i.e., nine-exemplars) may also impact on generalisation by making extra information potentially accessible at the time of building up the group judgment. To investigate the potential impact of perceived prior contact on out-group judgments, participants will be asked to report their (perceived) prior contact with the elderly.

**Summary**

I predict a three-way interaction between the number of exemplars (three vs. nine), information source (presented vs. retrieval) and meta-information cue (absent vs. present) on each of the outcome measures. In the absence of a meta-information cue, a bookkeeping effect should be evident for participants in the presented information condition and a retrieval fluency effect for participants in the self-generation condition. This will demonstrate that people use both meta-information and information when making group judgments. The provision of an external meta-information cue should result in a nullification of these basic effects, because the automatic process of applying all the accessible information to the subsequent judgment has been corrected.

**Method**

**Participants and Design**

Participants were first year psychology students at the University of Newcastle (N = 260; 209 females and 51 males) who completed the research for 2% course credit. The mean age of the participants was 20.18 years (SD = 3.58 years). The majority of participants indicated that their first spoken language was English (96%; N = 250) and that their cultural background was Australian (93%; N = 243).

The design was a three-factor between subjects design: 2 (information source: self-generation vs. presented) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) with an appended no-exemplar control group. There were between 20 and 35 participants in each cell of the design, except in the self-generation/nine-exemplar/present cue condition where there were 51 participants. This was due to a computer programming error that resulted in double the random allocation to that condition.

**Procedure**

The procedure adopted was similar to that in Study 1, with some minor changes that I now detail. The experiment was conducted online using a secure website, rather than a one-to-one laboratory session. This change meant that the IAT was not included in this study.
because time sensitive responses could not be obtained (Birnbaum, 2004). Upon login, participants were randomly allocated to one of the 9 conditions and provided with the same cover story from Study 1 (i.e., the research was investigating different role models requiring them to think about different kinds of people). As in Study 1, number of exemplars was manipulated by asking participants to consider three or nine active elderly people. As this study manipulated information source, some participants were asked to retrieve the exemplars from their memory (self-generation) while others were asked to read a list of counterstereotypical examples (presented). Participants in the self-generation condition received the same retrieval instructions given in Study 1. Participants in the presented condition were asked to read through descriptions of elderly individuals that “previous participants had considered active”.

The nine active elderly descriptions were derived from a qualitative analysis of the exemplar descriptions provided by participants in Study 1 (for a similar strategy, see Corley & Pollack, 1996). The qualitative analysis revealed that previous participants had generated about twice as many personally known exemplars as famous exemplars, with an even balance of male to female exemplars. Thus, the provided list for this study was arranged so that it included two non-famous exemplars for each famous exemplar and was balanced for gender. To avoid any order effects the exemplars were presented in a completely randomised manner for each participant. Each exemplar description used phrases commonly included by the previous participants (e.g., “always very busy”, “frequently on the go”) and were all between 69 – 74 words in length (for the full list of experimental profiles, see Appendix III). An example follows:

“Frank Wilkinson, 78 years of age:
Frank is a very energetic man. He walks to the store each morning to buy the paper. After reading through the news he then tries to complete the crosswords in the back section of the paper. After lunch he walks to the local primary school where he volunteers to help students with reading difficulties. He makes sure that he gets about 30 minutes of exercise by walking every day.”

To ensure that participants processed the presented information they were asked to type a description detailing their impression of each elderly person immediately after exposure to the exemplar. The length of this description will serve as the vividness measure. To maintain consistency between conditions, the self-generation participants were also asked to complete the exemplar description task immediately after generating each exemplar.

---

8 At the time of data collection, I was not aware that there was a paper-and-pencil version of the IAT (e.g., Lowery, Hardin & Sinclair, 2001). Future research could include this version to maintain an implicit measure in retrieval fluency research during on-line and small group administration.
Participants allocated to the meta-information cue: present conditions were provided with a cue that alerted them to the ease or difficulty of the task. Before (and after) exposure to the counterstereotypical information, participants in the three-exemplar self-generation conditions read: “This task is (was) quite simple. Most people find (have found) this easy to complete”. Participants in the nine-exemplar self-generation conditions read that previous participants had found the task difficult. To maintain consistency with the cover story, the presented information participants were provided with the cue in the past tense. Specifically, presented information participants were told that other participants had found it easy (or difficult) to retrieve the information from their memory. Participants in the absent meta-information cue conditions did not receive any additional meta-information cue.

**Dependent measures.** As in Study 1, participants allocated to the no-exemplar appended control started the research at the dependent measures section of the study. At this point, all participants were asked to complete explicit measures of group stereotyping. In this study, participants were asked to indicate how well each adjective from Pilot Study 2 “described the elderly in general” on a 6-point Likert scale (1 = *Not at all*; 6 = *Extremely*). The stereotype change index (α = .65) was created by averaging responses to ‘active’, ‘strong’ (both reverse scored), ‘frail’ and ‘slow’. A differential score was created by subtracting from the control group mean (M = 4.49, SD = .78). After these data manipulations, positive scores depicted stereotype change while negative scores depicted an increase in stereotyping.

Dispersion was measured by asking participants to indicate how similar they considered the elderly to be on each of the adjectives from Pilot Study 2 (1 = *Not at all similar*; 6 = *Extremely similar*; e.g., see Stroessner et al., 2005). The responses were reverse scored so that higher scores depicted greater perceived dispersion and were then combined to create a dispersion index (Cronbach’s α = .70). A differential score was created by subtracting the participants mean dispersion score from the control group (M = 3.32, SD = .88). After this manipulation, positive scores indicate greater perceived dispersion while negative scores indicate a decrease in dispersion.

Prejudice was measured by asking participants to indicate their feelings towards the elderly as a group, on the traits ‘kind’, ‘likeable’ and ‘pleasant’ (1 = *Not at all*; 6 = *Extremely*). A differential score was created by subtracting the prejudice index (Cronbach’s α = .77) from the mean of the control group (M = 2.47, SD = 1.02). After this manipulation, positive scores indicate a less prejudiced attitude.

---

9 The traits ‘fast’, ‘forgetful’, ‘inactive’ and ‘weak’ did not contribute to the reliability score and were excluded.

10 The traits ‘fast’, ‘forgetful’ and ‘strong’ did not contribute to the reliability score and were excluded.
After the group judgments, both the presented and self-generated participants were asked to complete the individual exemplar typicality and frequency measures. Here, the name of each exemplar and the participant’s description for that exemplar was displayed at the top of the screen. Participants were asked to indicate how “typical they considered this person to be of elderly people in general” (1 = Not at all; 6 = Extremely). The responses were combined across exemplars to create a mean typicality index ($\alpha = .73$) that ranged from 1 to 6, with higher scores depicting greater perceived typicality. Participants were also asked to estimate “how many other elderly people, amongst those they knew, would be as active as this exemplar”. The participants could enter any number. Responses ranged from 1 (indicating that only one other person was as active as the exemplar, i.e., infrequent) to 30 (indicating that 30 other elderly people were as active as the exemplar, i.e., very frequent). The frequency responses were combined across exemplars to create a mean frequency index ($\alpha = .85$).

The manipulation check for number of exemplars followed. In this study I used a self-reported task ease measure. Participants in the self-generation conditions responded to the question: “How easy was it to recall all the active elderly people?” (1 = Not at all; 6 = Extremely). Participants in the presented conditions responded to the question: “How easy was it to form an impression of all the active elderly people?” (1 = Not at all; 6 = Extremely; for a similar measure, see Dijksterhuis et al., 1999; Haddock et al., 1999).

Next, all participants answered a final series of questions. This included a feeling thermometer to assess the easy-thus-liked heuristic. Here, participants saw a picture of a thermometer that was labelled at 10° increments, with 0° labelled as ‘Extremely cool’ and 100° labelled as ‘Extremely warm’. The other labels moved from very, quite, fairly, slightly cool/warm. The midpoint was labelled ‘neither cool nor warm’. Higher ‘temperatures’ represent more warmth or positive feelings towards the elderly (for a similar measure with the elderly, see Dasgupta & Greenwald, 2001). Participants then indicated the number of unrequested cognitions by indicating the number of inactive elderly people that came to mind during the research. Participants then answered a series of questions to assess their perceived prior contact with active elderly people. These questions asked participants to consider three categories: (1) family, (2) friends, and (3) media. For each category they were asked to estimate the total known number of elderly people (i.e., total prior contact) and then provide a separate estimate of the number of active elderly in each category (i.e., counterstereotypical prior contact). For each category (i.e., family, friends, and media), I created a ratio score by dividing the number of active elderly by the number of known elderly. These ratio scores were then combined to establish a mean prior counterstereotypical contact index. This ranged
from 0 to 1; with higher scores indicating that a greater proportion of the known elderly people were considered active\textsuperscript{11}. Finally, participants completed their demographic measures.

**Results**

**Manipulation Checks: Number of Exemplars**

Responses to the self-reported task ease measure were used to check the fluency manipulation. As this measure was worded differently for participants in the self-generation and presented conditions the analysis was conducted separately for each level of information source. It was expected that both the self-generation and presented information participants would display a main effect of number of exemplars in the 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) between subjects ANOVA. These predictions were supported: participants in the self-generation conditions found it easier to retrieve the exemplars in the three-exemplar condition ($M = 3.59$, $SD = 1.56$) than in the nine-exemplar condition ($M = 3.01$, $SD = 1.54$), $F(1, 133) = 4.01$, $p = .04$. This demonstrates that retrieval fluency was successfully manipulated in the self-generation condition. Participants in the presented conditions also found it easier to form an impression in the three-exemplar condition ($M = 4.23$, $SD = 1.25$) than in the nine-exemplar condition ($M = 3.68$, $SD = 1.30$), $F(1, 89) = 4.58$, $p = .04$.

**Testing Basic Effects: Group Judgments**

Descriptive values for change in group stereotyping, group dispersion and prejudice as a function of meta-information cue, information source and number of exemplars are reported in Table 6.1. Preliminary analyses revealed no effects of gender, and so this variable was not considered further.

**Change in group stereotyping.** The predicted three-way interaction in the 2 (information source: self-generation vs. presented) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) between subjects ANOVA was significant for the change in (explicit) group stereotyping index, $F(1, 223) = 4.72$, $p = .03$. In order to isolate the predicted retrieval fluency and bookkeeping effects, this interaction was followed up separately by investigating each level of the meta-information cue.

In the meta-information cue absent condition, participants were expected to demonstrate an interaction between the presented and self-generation participants, with a bookkeeping effect emerging for the presented participants and a retrieval fluency effect

\textsuperscript{11} Three participants provided scores that were higher than one; this meant that the participant estimated more active elderly people than they had indicated were in the category. These three instances were set to the highest score within their respective condition that was within 2 standard deviations of the group mean.
Table 6.1
Outcome Measures as a Function of Meta-Information Cue, Information Source and Number of Exemplars (SD shown in parenthesis)

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Number of Exemplars</th>
<th>Meta-Information Cue</th>
<th>Absent Information Source</th>
<th>Present Information Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotypicality</td>
<td>Three</td>
<td></td>
<td>.39* (.70)</td>
<td>Three</td>
</tr>
<tr>
<td></td>
<td>Nine</td>
<td></td>
<td>.06 (.65)</td>
<td></td>
</tr>
<tr>
<td>Dispersion</td>
<td>Three</td>
<td></td>
<td>.01 (.85)</td>
<td>Three</td>
</tr>
<tr>
<td></td>
<td>Nine</td>
<td></td>
<td>-.15 (.67)</td>
<td>Nine</td>
</tr>
<tr>
<td>Prejudice</td>
<td>Three</td>
<td></td>
<td>.33 (.84)</td>
<td>Three</td>
</tr>
<tr>
<td></td>
<td>Nine</td>
<td></td>
<td>.24* (.62)</td>
<td>Nine</td>
</tr>
</tbody>
</table>

Notes. Positive scores depict greater change from the baseline.
* Significantly different from zero at $p \leq .05$. 
emerging for the self-generation participants. This prediction was supported with a significant two-way interaction between information source and number of exemplars, $F(1, 100) = 8.74, p = .004$. This interaction is depicted in Figure 6.1.

![Figure 6.1: Change in group stereotyping for meta-information cue: absent participants as a function of information source and number of exemplars (error bars: 95 % CI). * Significantly different from zero at $p \leq .05$.](image)

Following up on the two-way interaction between information source and number of exemplars amongst participants in the meta-information cue-absent condition revealed the predicted bookkeeping effect amongst presented information participants. Participants presented with nine exemplars changed their stereotype more ($M = .63, SD = .85$) than participants presented with three exemplars ($M = .13, SD = .61$), $F(1, 46) = 5.51, p = .02$. One sample $t$-tests revealed that the nine-exemplar/presented participants were significantly different from zero ($t(24) = 3.73, p = .001$), while the three-exemplar participants were not significantly different from zero ($t < 1$). Amongst participants in the self-generation condition, there was marginal support for a retrieval fluency effect, $F(1, 54) = 3.16, p = .08$. Here, participants who retrieved three exemplars changed their stereotype more ($M = .39, SD = .70$) than participants who retrieved nine exemplars ($M = .06, SD = .65$). This marginal effect is consistent with the retrieval fluency effect identified in Study 1. One sample $t$-tests revealed that the three-exemplar/self-generation participants were significantly different from zero ($t(20) = 2.58, p = .02$), while the nine-exemplar participants were not significantly different from zero ($t < 1$).
In the meta-information cue present condition participants were expected to demonstrate no retrieval fluency or bookkeeping effects. This prediction was supported by a non-significant two-way interaction between information source and number of exemplars, $F < 1$. This null effect is depicted in Figure 6.2. In both the presented information condition and self-generation conditions, there was no significant difference between the three-exemplar and nine-exemplar means, both $F s < 1$. This shows that the provision of an external meta-information cue wiped off both the bookkeeping effect and the (marginal) retrieval fluency effect detected in the cue-absent conditions. One sample $t$-tests revealed that participants in the self-generated/nine-exemplar condition were significantly different from zero, $t (50) = 3.32, p = .002$. Similarly, a one-sample $t$-test for the self generated/three-exemplar condition approached significance, $t (29) = 1.83, p = .08$. This suggests that self-generation participants changed their negative out-group stereotype, while the lack of difference between these two conditions indicates that the meta-information cue had wiped-off the retrieval fluency effect. In the presented information conditions there was no change in group stereotyping from the baseline, both $t s < 1$.

**Figure 6.2:** Change in group stereotyping for meta-information cue: present participants as a function of information source and number of exemplars (error bars: 95 % CI). * Significantly different from zero at $p < .05$.

**Change in group dispersion and prejudice.** The 2 (information source: self-generation vs. presented) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) between subjects ANOVA was repeated to test change in group dispersion and group prejudice. On group dispersion, the three-way interaction was non-
significant, $F(1, 223) = 2.21, p = .14$. The two-way interactions were all non-significant ($F < 1$). There were no main effects, all $F$s $\leq 2.86$, $ps \geq .09$. The lack of effects replicates the null findings on change in group dispersion identified in Study 1. On group prejudice, all interactions and main effects were non-significant, all $F$s $\leq 1.89$, $ns$. These non-significant effects are consistent with the findings of Study 1.

Testing Mediating Processes

The easy-thus-liked heuristic. To assess the easy-thus-liked heuristic, I used the ‘feeling thermometer’ as a measure of global likeability. As this heuristic is relevant only to the basic retrieval fluency effect, I focused on participants in the self-generated/meta-information cue-absent conditions. To identify a mediating variable, I followed the three preliminary tests used in Study One. There was no mediating evidence for the easy-thus-liked heuristic. A full comparison using all participants was also conducted and the results followed a similar pattern as with the reduced design (for results of the restricted design and the full analysis, see Appendix IV).

The easy-thus-typical heuristic. Two measures of typicality assessed the easy-thus-typical heuristic: (1) the typicality index, and (2) the frequency index. For each of these measures, I carried out the three preliminary tests to identify a mediating variable. As in the analysis for the easy-thus-liked heuristic, I focused this investigation on the results of the participants in the self-generation task absent conditions. There was no mediating evidence for exemplar typicality or frequency. The results for the restricted design and the full analysis can be found in Appendix V.

Unrequested cognitions. Participants could respond to the unrequested cognitions item with any number of stereotypical elderly people and 90% of participants indicated they considered 10 or less inactive elderly people. There were 8 participants 2.5 SDs or more away from their condition mean. These were replaced with their group mean.

The three steps to test for a mediating variable (ANOVA, ANCOVA and correlation analysis) were carried out using unrequested cognitions. The three-way interaction involving information source, number of exemplars, and meta-information cue on the number of unrequested cognitions (i.e., stereotypical exemplars) was non-significant, $F < 1$. The two-way interactions involving number of exemplars were all non-significant, both $F$s $< 1$. There was a trend for significance on the interaction between meta-information cue and information source, $F(1, 222) = 3.04, p = .08$. Following up on this marginal two-way interaction, I found that participants in the presented information condition generated more unrequested cognitions under no meta-information cue ($M = 6.18, SD = 7.43$) than when a meta-information cue was provided ($M = 2.80, SD = 2.83$), $F(1, 91) = 8.37, p = .005$. Similarly, participants in the self-generation condition generated more unrequested cognitions under no
meta-information cue ($M = 4.66, SD = 4.03$) than when a meta-information cue was provided ($M = 3.50, SD = 2.22$), $F (1, 135) = 3.46, p = .07$, but this difference was noticeably smaller. This general trend for more unrequested cognitions amongst meta-information cue absent participants was also evident in a marginal main effect of meta-information cue, $F (1, 133) = 3.53, p = .06$. Participants in the meta-information cue: absent condition reported thinking of more inactive elderly ($M = 4.66, SD = 4.03$) than participants in meta-information cue: present condition ($M = 3.50, SD = 3.21$).

When the number of unrequested cognitions was entered as a covariate in the three-way ANOVA on change in group stereotyping, no observable change was detected in the effect of information source, number of exemplars and meta-information cue from $F (1, 223) = 4.72, p = .03$ to $F (1, 222) = 5.25, p = .02$. Also, there was no correlation between unrequested cognitions and stereotype change, $r = .06, p = ns$. Overall, unrequested cognitions did not mediate the impact of meta-information within a member-to-group generalisation framework. However, the marginal effects of the manipulated meta-information cue on unrequested cognitions suggest that participants may use the cue to monitor whether their experience is in line with others and to regulate their task engagement.

**Perceived prior contact.** The prior contact index was entered as the dependent variable on the three-way ANOVA. All interactions and main effects were non-significant, all $Fs \leq 1.04$, all $ps \geq .31$. When prior contact was entered as a covariate in the three-way ANOVA on change in group stereotyping, there was no statistically observable change of the effect of information source, number of exemplars and meta-information cue from $F (1, 223) = 4.72, p = .03$ to $F (1, 222) = 5.72, p = .02$. There was a significant correlation between prior contact and stereotype change, $r = .23, p < .001$ (two-tailed), indicating that the more contact with active elderly the participants reported, the larger the reduction in stereotyping after the disconfirming stereotype retrieval task. Overall, these analyses suggest that the perceived level of prior contact with active elderly people is not a mediating variable in this study.

**Exemplar vividness.** As in Study 1, exemplar vividness was measured by counting the number of words in the exemplar description task (overall word range from 1 – 173 words; $M = 26.19; SD = 18.26$). To check exemplar vividness amongst participants in the nine-exemplar conditions, the mean number of words from the first three descriptions was compared to the mean number of words from the last three descriptions. As manipulation of information source and meta-information cue varied for the nine-exemplar participants, this analysis was done separately across conditions. With the exception of the participants presented with nine exemplars in the meta-information cue: absent condition ($t (23) = 2.26, p = .03$), all nine-exemplar conditions did not significantly differ when vividness of the first

\footnote{Two exemplars were found to be over 4 standard deviations outside their category mean, and were replaced with their condition mean.}
three exemplars was compared with the vividness of the last three exemplars (all $t_s \leq 1.36, p_s \geq .18$)\(^\text{13}\). Importantly, this indicates that participants in the self-generation nine-exemplar conditions were not generating less vivid exemplars towards the end of their retrieval task.

**Discussion**

The goal of this research was to investigate two contrasting approaches to member-to-group generalisation: (1) presented information, and (2) self-generated information. The predicted three-way interaction between information source, number of exemplars, and meta-information cue was significant. Below, I discuss the details of the effects that emerged from this interaction, beginning with a discussion of the basic effects and then addressing the findings from the analysis of the mediating variables.

**Impact of the Meta-Information**

Amongst the presented information participants who did not receive a meta-information cue, the meta-information associated with the presentation of a larger sample size was expected to facilitate the inclusion of the exemplar information (see Nisbett et al., 1983). This prediction was supported with the emergence of a main effect for number of exemplars, whereby participants in the nine-exemplar condition demonstrated greater change in group stereotyping compared to participants in the three-exemplar condition. In line with the default approach to generalisation that was identified in Chapter 2, this finding suggests that, under normal or default conditions, people include presented disconfirming information in their group judgments and they do so to a greater extent as the sample size increases. In other words, the impact of the meta-information associated with a large number of presented disconfirming exemplars facilitates inclusion of the exemplar material in out-group judgments.

Amongst the self-generation participants who did not receive a meta-information cue, the meta-information associated with the retrieval experience was expected to facilitate inclusion of the exemplar information as the size of retrieved exemplars decreased. This prediction was also supported with the emergence of a marginal main effect for number of exemplars, whereby participants in the three-exemplar condition demonstrated greater change in group stereotyping compared to participants in the nine-exemplar condition. This demonstrates that the impact of the meta-information associated with retrieving a large number of disconfirming exemplars restricts the inclusion of the exemplars in out-group judgments.

\(^{13}\) In the nine-exemplar presented meta-information cue: absent condition the mean task length for the first three descriptions was $M = 28.25$ ($SD = 20.00$) and the mean of the length of the last three words was $M = 23.18$ ($SD = 12.63$).
Taken together, this study demonstrates that the exact impact of the meta-information depends, at least in part, on the source of the actual information. On one hand, when the exemplar information is presented the internal meta-information associated with the law of large numbers informs out-group judgments by enhancing generalisation as the sample size increases (i.e., the more-the-merrier). On the other hand, when the information is self-generated the internal meta-information associated with the experience of retrieval informs out-group judgments by enhancing inclusion when the sample is smaller (i.e., the less-the-merrier). It is important, therefore, for future member-to-group generalisation research to better predict the impact of the exemplar information to consider both the meta-information and the information source.

**Challenging Automatic Effects**

The presentation of an explicit meta-information cue was expected to engage a more controlled approach to generalisation and override the ‘automatic’ bookkeeping and retrieval fluency effects identified in the meta-information cue: absent condition. This was demonstrated through a null effect of information source and number of exemplars in the meta-information cue-present condition. Social cognitive research by Paolini et al. (2009, Study 1) supports this finding. Paolini et al. found that the provision of a meta-information cue flagging the potential unreliability of the member information resulted in a null effect of the counterstereotypical information. It seems that when participants are presented with additional information about the nature of a reading or retrieval task, this information may be sufficient to engage bias control that undermines the effect of the meta-information.

The demonstration that an explicit meta-information cue nullified the impact of the meta-information in the self-generated and presented information conditions also supports the operation of a dual process (e.g., Chaiken et al., 1989; Cacioppo et al., 1986). In the absence of a meta-information cue, participants rely on heuristic meta-cognitive strategies, such as fluency or the law of large numbers. However, when participants are encouraged to use systematic processing, these automatic heuristic strategies are disrupted and controlled for. The ability of individuals to use either automatic or controlled processes demonstrates dual processing routes are available to individuals and that either route may be activated through contextual cues available at the time of judgment. In the next chapter, I follow up on the dual process model that I have presented to account for the impact of retrieval fluency on member-to-group generalisation.

**Outcome Measures of Dispersion and Prejudice**

As in Study 1, there was no effect of the manipulations on the measure of dispersion and prejudice. It is possible that the elderly are already perceived to be dispersed on the
stereotypical trait ‘inactive’. This suggestion is supported by the lack of difference between the no-information control group and the experimental conditions. On the measure of prejudice, it is possible that the null effects are due to concerns of political correctness. The prejudice measure blatantly asked participants how much they ‘like’ the elderly, and this may encourage the participants to put forward socially desirable responses. With two studies producing null effects on these outcome measures, it is likely that the following studies will fail to identify any effects on these two outcome measures.

Explaining the Nature of Meta-Information

Study 2 also aimed to explain the underlying influences driving the retrieval fluency effects. In addition, I also included measures to assess the potential role of unrequested cognitions, perceived prior contact and exemplar vividness. In this section I summarise these results.

Affective vs. cognitive. The easy-thus-liked and the easy-thus-typical heuristics were not supported. The inability to isolate these cognitive and affective explanations as the underlying mechanism/s may be due to, at least, two reasons: (1) The process variables used are poor and unable to capture the process; or (2) the process responsible for the effect is different from the two tested heuristics. Both of these issues will be addressed in the following chapter. Firstly, a change to the measurement of the typicality process variable will be made by measuring for overall, or global, typicality. Secondly, it is possible that there is another mechanism driving the retrieval fluency effects, specifically, increased tiredness amongst participants in the nine-exemplar condition. To my knowledge, no previous research has addressed the idea that participants are adopting a heuristic processing strategy because of cognitive fatigue. Study 3, presented in the following chapter, will address this gap in the literature. This will be achieved by including a condition that parallels the retrieval experience of participants in the nine-exemplar condition by asking some participants to generate stereotype-unrelated exemplars.

Unrequested cognitions. Tormala et al. (2007) found that retrieval fluency effects were mediated by unrequested cognitions. I found no evidence that unrequested stereotypical exemplars impacted on the obtained results. This is important, as it furthers the validity of this work as a strategy for changing negative group stereotypes, through increasing our understanding and awareness of internal meta-cognitive cues.

There are a number of possible reasons for the lack of support for the role of unrequested cognitions. Firstly, Tormala et al. (2007) demonstrated the mediating role of unrequested cognitions by having participants generate previously unconsidered cognitions, such as hypothetical arguments in support of comprehensive university exams. Retrieving counterstereotypical exemplars from memory is a different cognitive process to the
construction of ‘novel’ arguments, and it is therefore reasonable that the mediating processes will also be different. A second possibility is that measuring unrequested cognitions in the present research may have evoked concerns of political ‘correctedness’ by asking participants to admit that they considered ‘stereotypical’ exemplars. However, the cover story led participants to believe I was investigating role models, so this explanation seems unlikely.

Finally, Tormala et al. (2007) asked their participants to generate 2 vs. 10 arguments in their research, making the ‘many’ condition five times more difficult than the ‘few’ condition. In the present research participants were asked to generate three vs. nine exemplars, making the ‘many’ condition three times more difficult than the ‘few’ condition. Therefore, the retrieval task may not have have been sufficiently different in the two conditions to produce enough variance along unrequested cognitions. While this is a possibility, the important point is that unrequested cognitions do not explain the retrieval fluency effects identified in this study.

What did emerge from the testing of unrequested cognitions, however, was a marginal main effect of the meta-information cue. This demonstrated that participants without a meta-information cue thought of more stereotypical exemplars. These results point towards the meta-information cue acting as a de-motivator or task engagement suppressor in the retrieval and impression formation tasks. As suggested by Festinger’s (1957) classic theory of cognitive dissonance, people use others’ perceptions as additional sources of information to help formulate uncertain attitudes. The meta-information cue provides this additional information by clarifying how participants had experienced the task. Hence, the cue would limit the ambiguity of the task (i.e., ‘I am doing like other participants, there is no need to try harder’) and reduce participants’ effort, consequently limiting the unrequested content. On the other hand, in the cue absent condition the task remains ambiguous (i.e., ‘Am I doing this right? I don’t know…perhaps I should try harder’) and, as a consequence, participants would more thoroughly engage in the task, ultimately finding more unrequested content.

**Perceived prior contact.** Contact with an out-group has been found to generate more positive out-group judgments (Pettigrew & Tropp, 2000). Perceived prior contact was included as a measure in this study to check if the retrieval fluency and bookkeeping effects were driven by perceived increased contact with the active elderly. There was no evidence that prior contact was mediating the obtained results on change in group stereotyping.

**Exemplar vividness.** The measure of exemplar vividness was included to ensure that participants in the nine-exemplar were not retrieving poorer quality exemplars towards the end of the retrieval task. Importantly, amongst participants in the self-generated nine-exemplar condition the first three exemplars were not significantly more vivid than the final three exemplars. Participants presented with nine exemplars in the meta-information cue: absent condition did, however, show a significant decrease in exemplar vividness from the first three exemplars to the last three exemplars. This may suggest that these participants had
disengaged from the presentation of the exemplar information. However, since these participants had demonstrated the predicted bookkeeping effect this explanation seems unlikely. It is possible that this reduction in exemplar vividness may be attributable to the task assessing vividness. A problem with using length of the description task to assess exemplar vividness is that participants may become briefer in their description in order to complete the task rather than brevity reflecting less vivid exemplars. To address this concern, a different measure of vividness will be used in the following study.

**Implications and Limitations**

It is important to note that this study did not include a direct manipulation of processing fluency. Recall that processing fluency is tested by manipulating processing ease. For example, Reber, Stark and Squire (1998) tested processing fluency by manipulating figure-ground contrast and found that participants judged the stimulus more positively when it was presented with a higher figure-ground contrast, allowing for high fluency. Participants in the presented information condition all experienced high processing fluency. Processing fluency will be manipulated in Study 5.

By demonstrating that people can control for the biasing effect of fluency highlights the value in programs targeting increased awareness of meta-cognitive processes. Making people more aware of the automatic influences on their judgments is likely to reduce the negative impact of stereotyped based judgments. The individual who is unaware of the biasing influence of retrieval fluency and bookkeeping effects may be unduly stereotypical without knowing the specific meta-cognitive process that is intruding on their judgment. This study has shown that individuals can override the bias associated with fluency when they are presented with external meta-information cues. Ultimately, this research suggests that it may be necessary for social policy to educate individuals on general strategies that reduce the influence of these automatic meta-cognitive effects.

**Conclusions**

This chapter has demonstrated that both the impression formation paradigm and the self-generation paradigm are valid approaches to stereotype change. It is only recently that meta-cognitive processes have been considered as a strategy to reduce negative out-group stereotypes and prejudices (Blair et al., 2001; Paolini et al., 2009; Rothman & Hardin, 1997; Yzerbyt, Leyens, & Schadron, 1997). This study builds on this research base by demonstrating the important role that meta-cognitions play in stereotype change. The next step is to further investigate the explanatory power of my dual process account of retrieval fluency.
CHAPTER 7

TESTING THE DUAL PROCESS ACCOUNT OF RETRIEVAL FLUENCY:
DEPTH OF PROCESSING, FATIGUE, AND BIAS CORRECTION
(STUDIES 3 AND 4)

Retrieval of information provides a perceiver with more than just the retrieved information; it also provides meta-information about the actual retrieval process. Both sources of information (information and meta-information) have been shown to influence subsequent judgements (e.g., Aarts & Dijksterhuis, 1999; Benjamin, Bjork, & Schwartz, 1998; Caruso, 2008; Dijksterhuis, Macrae, & Haddock, 1999; Haddock, 2000; Haddock et al., 1999; Sanna & Schwarz, 2003; Winkielman & Schwarz, 2001). In this thesis, Studies 1 and 2 have shown that retrieval fluency effects extend to the process of member-to-group generalisation. I used this evidence to argue for a dual process account of retrieval fluency whereby people automatically use fluency meta-information to help build their out-group judgments unless they are alerted to the biasing nature of this metacognitive process.

In this chapter, I will report two studies that continue the investigation of retrieval fluency within member-to-group generalisation. Study 3 will explore alternative explanations that may account for the application of retrieval meta-information within member-to-group generalisation. Study 4 will explore the nature of the cue required to engage bias correction. I expect that these two studies will extend the dual process model of retrieval fluency within member-to-group generalisation.

Ruling out Alternative Explanations

Throughout Studies 1 and 2, I assessed the easy-thus-liked and the easy-thus-typical heuristics as two possible explanations for the retrieval fluency effects. So far, these variables have not received empirical support, which may be attributed to, at least, two possible reasons. Firstly, the mediating variables are unable to capture the process. To address this concern, I will make changes to the assessment of these variables. Secondly, it is likely that these heuristics are not responsible for the retrieval fluency effects; rather some other variable is at play. To address this concern, I will introduce the assessment of depth of processing and cognitive fatigue as two new variables that may explain retrieval fluency effects within the process of member-to-group generalisation.
Depth of Processing

Psychological research from different areas has demonstrated that the more thoroughly individuals engage in cognitive activities the more inclined they are to cognitively elaborate on issue-relevant material (e.g., Cacioppo & Petty, 1982; Cacioppo et al., 1986). For example, cognitive research has demonstrated that when people are encouraged to process information in a deeper and more meaningful manner they are more likely to accurately use the information in judgments (Craik & Lockhart, 1972; Craik & Tulving, 1975, Hyde & Jenkins, 1969). Depth of processing might be involved in retrieval fluency effects.

In the original retrieval fluency research by Schwarz et al. (1991, Study 1) participants were asked to describe past examples of their past behaviours. Schwarz et al. found a marginal retrieval fluency effect, which was dismissed in a footnote as being “of little concern” (p. 196). Research by Dijksterhuis et al. (1999) asked participants to list traits on which “women and men are, on average, different” (p. 768). Dijksterhuis et al. found a significant, rather than marginal, retrieval fluency effect amongst people high in prejudice. The difference in significance between Schwarz et al. and Dijksterhuis et al. (i.e., marginal results vs. significant results) may reflect differences in the task instructions (i.e., describe vs. list). A description task requires more cognitive engagement than a listing task. Thus, rather than the meta-information influencing subsequent judgments, and the differences between the marginal and significant effects being “of little concern”, it is possible that asking people to describe (rather than list) the retrieved content is promoting greater involvement in the content. Similarly, in Study 1, I identified a significant retrieval fluency effect when participants were asked to list counterstereotypical exemplars. In Study 2, I identified a marginal retrieval fluency effect when participants were asked to describe counterstereotypical exemplars at the time of retrieval. It is possible, that participants in Study 2 were (unintentionally) encouraged to process the retrieved information in a more thorough manner, thus accounting for the differences between the studies.

If depth of processing, rather than retrieval ease-difficulty per se, is responsible for retrieval fluency effects, then the conclusion that meta-information is used in formulating judgments may be flawed. This depth of processing explanation will be tested in Study 3 by manipulating the position of the exemplar description task. As in my earlier studies, all participants will be asked to describe each retrieved exemplar. However, this time, half of the participants will be asked to process their retrieved exemplar more deeply by providing a detailed description immediately after generating each person and before the critical dependent measures (this replicates Study 2). The other participants will provide a detailed description for each exemplar following completion of the out-group measures (this replicates Study 1). Through this manipulation of the description task, I hope to demonstrate that differences in cognitive processing are “of little concern”, and rule out the alternative
suggestion that differences in cognitive processing account for retrieval fluency effects. If shallow depth of processing is critical to a significant retrieval fluency effect, then participants listing their exemplars should display stronger retrieval fluency effects than participants immediately describing their retrieved exemplars. If depth of processing is not critical to retrieval fluency then there will be no impact of this variable on generalisation.

**Cognitive Fatigue**

As people deliberate on a difficult task, such as retrieving nine counterstereotypical exemplars, it is reasonable to expect that they may become increasingly mentally fatigued. It is possible that cognitive fatigue, rather than the influence of meta-information, may explain the retrieval fluency effects. Stereotype change research has demonstrated that cognitive fatigue leads to enhanced reliance on a group stereotype (e.g., Bodenhausen & Lichtenstein, 1987; Garcia-Marques & Mackie, 1999; Gilbert & Hixon, 1991; Macrae, Hewstone, & Griffiths, 1993; for a review see Fiske, 1998). For example, Gilbert and Hixon (1991, Study 2) found that, when participants lacked cognitive resources during a brief contact with an Asian who was presented as stereotype-neutral, they were more likely to rely on the group stereotype when judging the target than participants who had sufficient cognitive resources. Recalling nine counterstereotypical exemplars is more demanding than recalling three counterstereotypical exemplars and this difference in cognitive demand may explain the retrieval fluency effect. Specifically, participants in the nine-exemplar condition (i.e., the cognitively demanding condition) may be reporting higher group stereotyping simply because they were cognitively fatigued and thus relied more on dominant (stereotypical) responses.

Study 3 will provide an important test of cognitive fatigue which, to my knowledge, is the first time this variable has been included in retrieval fluency research. This novel account of the retrieval fluency effect will be tested by allocating participants to different exemplar group conditions. Half of the participants will be asked to retrieve active elderly exemplars while the other half of the participants will be asked to retrieve exemplars from a stereotype-irrelevant group. The stereotype-irrelevant group will be pilot tested to ensure that the task is similar in retrieval difficulty to the standard (stereotype-relevant) retrieval fluency condition. In addition, the stereotype-irrelevant group will be similar in valence to control for mood (for mood congruent research, see Esses & Zanna, 1995; Forgas, 1994; Salovey & Birnbaum, 1989). If cognitive fatigue is critical to retrieval fluency effects, then a retrieval fluency effect should emerge in both exemplar group conditions, whereby all participants retrieving nine-exemplars display greater reliance on the stereotype. If cognitive fatigue, on the other hand, is not critical to retrieval fluency, then there should be a difference amongst participants retrieving relevant vs. irrelevant exemplars, with only stereotype-related participants displaying a retrieval fluency effect.
Testing Bias Control

If differences in depth of processing and cognitive fatigue do not account for retrieval fluency effects, this will strengthen the dual process account of retrieval fluency effects. In which case, it is necessary to further explore the bias control mechanism that I have included as part of my dual process account.

Research suggests that when people become aware of a bias, such as retrieval fluency, they attempt to correct for this influence (Lories et al., 1998; Martin, Seta, & Crelia, 1990; Schwarz & Bless, 1992; Wegener & Petty, 1995). Several meta-cognitive models also propose that people correct for a bias once they have recognised this may influence their judgment (e.g., Convey et al., 2005; Martin, 1985; 1986; Wegener & Petty, 1995). Although there are differences in predictions about the direction and extent of the correction process across models, it is generally agreed that once perceivers are aware of biasing factors the correction process is initiated (for reviews see, Kruglanski & Orehek, 2007; Wegener et al., 1998).

I have argued that bias correction forms part of the dual process model to account for the ‘correction’ of retrieval fluency effects. In Study 2, I demonstrated that participants presented with a congruent meta-information cue did not display retrieval fluency effects. In Study 4, I will further investigate the role of the meta-information cue by manipulating the direction of this information. This will be done by presenting some participants with a congruent meta-information cue (replicating Study 2), which will correctly alert participants to the nature of the retrieval task (e.g., participants in the nine-exemplar conditions will be told that the retrieval task is hard). Other participants will be presented with an incongruent meta-information cue, which will incorrectly alert participants to the nature of the retrieval task (e.g., participants in the nine-exemplar conditions will be told that the retrieval task is easy).

Previous retrieval fluency research using an incongruent meta-information cue has revealed contradictory results. On one hand, there is evidence that the provision of both congruent and incongruent meta-information cues can result in a null effect of retrieval fluency. For example, Schwarz et al. (Study 2, 1991) asked participants to recall either six or 12 instances of past assertive behaviours and then judge their own level of assertiveness. Prior to retrieval participants were provided with a cue about the nature of the task that was either congruent or incongruent to the task ease. The provision of the congruent and incongruent meta-information cues resulted in a null effect of retrieval fluency (see Chapter 3). In line with the dual process account of retrieval fluency, it would seem that a simple cue about the retrieval task is sufficient to engage in a bias correction process, and it is not the direction of the cue that is important.
On the other hand, there is evidence that the direction of the meta-information cue is important. For example, Haddock et al. (1999, Study 2) found that when their participants were provided with an incongruent cue (e.g., being told that background music would inhibit their recall in the easy condition) prior to the recall of pro-euthanasia arguments, a marginal retrieval fluency effect was detected. That is, participants’ attitude strength towards euthanasia was higher in the easy (three pro-arguments) than in the hard condition (seven pro-arguments). Haddock et al. argued that the provision of an incongruent cue rendered the individuals’ experience of retrieval diagnostic. As a result they relied on the meta-information associated with retrieval when deriving their judgment. However, Haddock et al. also found that the provision of a congruent cue resulted in a trend for sample size effect. Participants’ attitude strength towards euthanasia was higher in the hard than in the easy retrieval conditions when they were told that the background music would influence their retrieval process in a congruent manner. Haddock et al. argued that this sample size effect occurred because the participants had been induced to discount their subjective experience and instead rely on their retrieved content. In line with my dual process model, it would seem that Haddock et al.’s results suggest that only congruent meta-information cues are able to engage a bias correction process.

In summary, the nature of the meta-information cue required to facilitate bias correction is unclear. Schwarz et al.’s (1991) results suggest that the provision of a meta-information cue is sufficient to engage in a correction process, while Haddock et al.’s (1999) results suggest that the direction of the meta-information cue is an important factor in engaging bias correction. The second study presented in this chapter, Study 4, will investigate these competing findings by manipulating the meta-information cue.

Summary

Through this chapter, I will report research investigating the automatic application of retrieval meta-information to member-to-group generalisation to demonstrate that retrieval fluency effects are not the product of methodological artefacts. I will also investigate the bias correction process further, with the aim to identify the exact nature of the meta-information cue required to engage bias correction. Ultimately, through the two studies contained in this chapter, I aim to support the dual process account for the retrieval fluency effects in member-to-group generalisation.

Study 3

Study 3 will focus on investigating alternative explanations that may account for the effects of exemplar number on member-to-group generalisation. This will be achieved by testing two novel explanations for the retrieval fluency effect: (1) depth of processing and, (2)
cognitive fatigue. As in Study 2, I will use the same manipulation of number of exemplars and meta-information cue. In addition, I will manipulate the position of the description task and the type of exemplar group that is being retrieved.

To test for differences in depth of processing of the counterstereotypical exemplars at the time of the key group judgments, I will manipulate the position of the exemplar description task. Some participants will complete the exemplar descriptions during the retrieval task and *before* the collection of the primary dependent variables (i.e., deeper processing) and other participants will complete the exemplar descriptions *after* the collection of the primary dependent variables (i.e., shallow processing). To rule out depth of processing, the four-way interaction involving position of description task, exemplar group, number of exemplars and meta-information cue should be non-significant and no other effects involving the depth of processing variable (i.e., position of the description task) should be statistically reliable. A lack of difference between the description task conditions will indicate that differences in depth of processing does not explain retrieval fluency, and instead support the role of meta-information in retrieval fluency effects.

To test for a cognitive fatigue explanation, I will include a stereotype-irrelevant exemplar group. Some participants will be asked to retrieve active elderly exemplars and other participants will be asked to retrieve stereotype-irrelevant exemplars. To rule out cognitive fatigue, the three-way interaction between exemplar group, number of exemplars and meta-information cue should be significant. Importantly, this three-way interaction should reflect a non-significant interaction between number of exemplars by meta-information cue and non-significant main effects for participants in the irrelevant-stereotype group. On the other hand, the three-way interaction should reflect a significant two-way interaction between number of exemplars and meta-information cue for participants in the active elderly group. Specifically, this two-way interaction should replicate the interaction on stereotype change detected in Study 2. Participants in the meta-information cue-absent condition should display greater stereotype change after retrieving three exemplars than nine exemplars and participants in the meta-information cue-present condition display a null effect of number of exemplars.

Finally, I will again investigate the role of exemplar vividness. In Studies 1 and 2, I used the exemplar description task to show that amongst participants in the nine-exemplar conditions the exemplars generated at the start of the task were not significantly different from the exemplars generated at the end of the task. A problem with this vividness measure was that it tapped the length of the description task, rather than the quality of each exemplar. It is possible that towards the end of the retrieval task, participants in the nine-exemplar condition were still describing exemplars in the same length as the earlier retrieved exemplars, but actually retrieved less active exemplars. This would result in these participants
including poorer quality exemplars in order to complete the task. If less active exemplars are incorporated in the group judgment, then stereotype maintenance may simply reflect the less active exemplars diluting the impact of the more active ones (cf. Nisbett, Zukier, & Lemley, 1981). To account for this criticism, in this study I will check the quality of exemplars, rather than length of description. It is expected that if retrieval fluency is driving the effects then, consistent with previous research (e.g., Caruso, 2008; Schwarz et al. 1991), there will be no differences in the quality of the last exemplars compared with the first exemplars.

Method

Participants and Design

Participants were second year psychology students from the University of Newcastle (N = 127; 94 females and 33 males) who completed the research as part of a course requirement. The participants’ mean age was 22.17 years (SD = 4.99 years). The majority of participants indicated English was their first spoken language (97%; N = 123) and their cultural background was Australian (85%; N = 108).

The study was a four-factor between subjects design: 2 (position of description task: before dependent measures vs. after dependent measures) x 2 (exemplar group: active elderly vs. dress conscious males) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent). There were between 6 and 10 participants in each cell.

Pilot Study 3

A pilot study was used to identify the stereotype irrelevant exemplar group. Participants were 8 students (7 females and 1 male; age: M = 21.13 years; SD = 2.03) from the general student population at the University of Newcastle who were recruited for another study and completed the pilot study at the conclusion of tasks unrelated to this research.

I identified three Australian stereotyped groups that were similar in valence to ‘active elderly people’, but completely unrelated to the trait ‘active’ and the out-group ‘elderly’. These groups were: dress-conscious males; fashionable teachers, and laid-back Asians. Pilot study participants were asked to identify and describe as many exemplars that were representative of each group (up to a maximum of 12). The order of presentation for the three social groups was counterbalanced. The results indicated that, on average, participants were able to generate more dress-conscious males (M = 7.22; SD = 1.78), than fashionable teachers (M = 3.88; SD = 1.54), and laid-back Asians (M = 3.50; SD = .50). Data from Pilot Study 2 presented in Chapter 5 indicated that participants could generate, on average, eight active elderly exemplars (M = 8.08; SD = 2.47), hence ‘dress-conscious males’ was selected as the group most suitable to produce similar retrieval demands as the active elderly group.
Procedure
Participants completed the research during class sessions. There were six class groups with the number of participants in each class ranging from 14 to 25. The procedure followed the same approach as Study 2. A few changes to account for the introduction of the exemplar group manipulation are noted below. At the beginning of the testing session participants were introduced to the research and given the opportunity not to complete the research; no-one took this option. Following the introduction, one of three researchers (one male; two female) distributed a questionnaire pack to each participant containing three booklets. Participants were instructed to work through the booklets consecutively and to contact the researcher with any questions.

The first booklet provided a pen-and-paper replica of the retrieval task, whereby participants were asked to retrieve either three or nine exemplars from their memory. Each of the six class groups were randomly assigned to one level of the number of exemplars manipulation. That is, all participants in the room were randomly assigned to either the three-exemplar or nine-exemplar condition. This was to ensure that students were finishing the research at similar times, avoiding cues about different conditions.

Participants were individually assigned to the remaining three variables (exemplar group; meta-information cue; and position of the description task) on a random basis. The meta-information cue was manipulated following the procedure used in Study 2. That is, prior to and immediately after, the exemplar generation task participants in the meta-information cue present condition were provided with a congruent cue that alerted them to the task ease or difficulty. Exemplar group was manipulated by requesting participants to retrieve either ‘active elderly people’ or ‘dress conscious males’ from their memory. Participants who were in the deeper processing condition were asked to complete a description of each exemplar at this retrieval stage (i.e., description task before dependent variables).

Dependent Measures. The second booklet contained the group measures. As in Study 1, participants were asked to mark an 8cm line to indicate their group judgments (0 = Not at all; 8 = Extremely). To simplify data collection for the class context only the traits ‘active’ and ‘inactive’ were used to measure for group stereotyping and dispersion. A stereotypicality index was created by reverse scoring the responses to the trait ‘active’ and averaging this with the responses to the trait ‘inactive’ (Cronbach’s $\alpha = .71$, Pearson’s $r = .55$, $p < .01$). The scale ranged from 0 to 8 cm, with higher scores indicating greater group stereotyping. Dispersion was collected using the range measure adopted in Study 1 on the traits ‘active’ and ‘inactive’. The scale ranged from 0 to 8 cm, with higher scores depicting greater dispersion (Cronbach’s $\alpha = .83$, Pearson’s $r = .71$, $p < .01$). The traits ‘likeable’ and ‘pleasant’ were used to measure prejudice. These traits were reverse scored and averaged to create a reliable prejudice index.
(Cronbach’s $\alpha = .76$, Pearson’s $r = .61$, $p < .01$). Higher scores on this index indicate more prejudiced attitudes. As this study used a pen-and-paper format, the IAT was not conducted.

The third booklet contained individual exemplar measures and manipulation checks. Participants were asked to return to their first booklet to view their retrieved exemplars. Typicality was measured by asking participants to indicate how typical they considered each exemplar to be of the elderly (or males) on a 7-point Likert scale ($1 = \text{Not at all}; 7 = \text{Extremely}$). Individual typicality ratings were combined to form an index of typicality (Cronbach’s $\alpha = .73$). To measure exemplar quality, the participants were also asked to indicate how active (or dress conscious) they perceived each retrieved exemplar ($1 = \text{Not at all}; 7 = \text{Extremely}$). Participants who completed the description task after the dependent measures (i.e., shallow processing condition) were also asked to complete a description of each exemplar at this stage.

The ‘easy-thus-liked’ heuristic was assessed by asking participants to complete a feeling thermometer by drawing a line to indicate the level of warmth felt towards the elderly ($0^\circ = \text{Extremely cool}; 100^\circ = \text{Extremely warm}$). Finally, participants completed the self-reported task ease question, used in Study 2 as a retrieval fluency check, where they indicated how easy it was to complete the retrieval task on a 7-point Likert scale ($1 = \text{Not at all}; 7 = \text{Extremely}$).

**Results**

*Ruling out Alternative Explanations*

**Depth of processing.** To test if depth of processing was driving the retrieval fluency effects a four-way ANOVA: 2 (position of description task: before dependent measures vs. after dependent measures) x 2 (exemplar group: active elderly vs. dress conscious males) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) was conducted on the stereotypicality index. There was no significant four-way interaction, $F(1, 111) = 1.14, p = .29$. There were no significant three-way interactions ($ps \geq .14$) or two-way interactions involving position of the description task ($ps \geq .15$). The main effect for position of the description task was also non-significant, $F < 1$. These results indicate that depth of processing was not critical to the key retrieval effect. Subsequent analyses are collapsed across the position of description task variable.

**Gender.** There was a possibility of in-group effects in the dress-conscious male condition, whereby the male (vs. female) participants may have found it easier to generate relevant exemplars. To test for gender differences in ease of retrieval, a 2 (participant gender: male vs. female) x 2 (exemplar group: active elderly vs. dress conscious males) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) ANOVA was run
with self-reported task ease as the dependent variable. The four-way interaction was non-significant, $F < 1$. There were no significant three-way interactions (all $ps \geq .18$) or two-way interactions involving gender (all $ps \geq .16$). There was no main effect for gender, $F < 1$. As it is also possible that in-group effects influence the key dependent variable, the four-way ANOVA was run also on the group stereotypicality index. The four-way interaction was once again non-significant, $F < 1$. There were no significant three-way interactions (all $ps \geq .27$) or two-way interactions (all $ps \geq .22$) involving gender. Finally, there was no main effect for gender, $F < 1$. In light of these null effects, gender will no longer be considered.

**Laboratory group effects.** Because classes, rather than individuals, were allocated to the two levels of the number of exemplars factor (three vs. nine), it was important to ensure that there were no laboratory group effects. As classes were tested in the three-exemplar vs. nine-exemplar factor, number of exemplars was excluded from this analysis. A three-way ANOVA carried out on the group stereotypicality index including class group, meta-information cue and exemplar group was non-significant ($F < 1$). The two-way interactions involving class group were also non-significant ($Fs < 1$), there was no main effect of class group, $F < 1$. In light of these null effects, laboratory group effects were considered negligible.

**Cognitive fatigue.** To rule out a cognitive fatigue interpretation, a significant three-way interaction between exemplar group, number of exemplars, and meta-information cue was expected to emerge on the group stereotyping index. The three-way interaction was expected to reflect no effects involving number of exemplars and meta-information cue amongst the dress conscious male participants. In line with these predictions, the three-way interaction was found to be significant, $F(1, 119) = 2.73, p = .05$, one-tailed$^{14}$. The interaction between number of exemplars and meta-information cue for participants in the dress-conscious male condition was non-significant, $F < 1$. The main effects of number of exemplars and meta-information cue amongst dress-conscious male participants were also non-significant, both $Fs < 1$. This suggests that the retrieval fluency finding in the previous studies is unlikely to be due to cognitive fatigue or tiredness. The results for participants in the active elderly condition will be reported below in the section summarising the key effects.

As a result of this analysis, participants asked to retrieve dress-conscious male exemplars were dropped from the main analysis and treated as a control condition. Consistent with the approach taken in Studies 1 and 2, the means of the dress-conscious male conditions on the group measures were used to create a differential score. That is, the mean index scores for each of the outcome measures provided by the dress-conscious male participants was used as a baseline to create a differential score reflecting change in group stereotyping ($M = 4.71$,

$^{14}$ As the direction of the means was predicted, I used Howell (1997, p. 683) to transfer the $F$ value to a $t$-value and identify the one-tailed significance levels.
Chapter Seven 129

SD = 1.31), change in group dispersion \((M = 5.19, SD = 1.93)\) and change in group prejudice \((M = 2.12, SD = 1.03)\).

**Exemplar quality.** The exemplar quality questions were used to check that participants in the active elderly/ nine-exemplar condition were not drawing on poorer quality exemplars as the task progressed. A \(t\)-test compared the mean responses to the exemplar quality measure for the first three exemplars with the mean responses for the last three exemplars. The quality of the first three exemplars was not significantly different \((M = 5.36, SD = .69)\) to the quality of the last three exemplars \((M = 5.10, SD = 1.12), t (22) = 1.26, p = .22.\) This demonstrates that participants in the nine-exemplar condition have not included less active exemplars at the end of retrieval task.

**Manipulation Checks**

**Number of exemplars.** Self-reported task ease was expected to show that participants in the three-exemplar condition found it easier to generate the exemplars than participants in the nine-exemplar condition. Using self-reported task ease as the dependent variable, the ANOVA including exemplar group, number of exemplars and meta-information cue was non-significant on the three-way interaction, \(F < 1.\) The only fully significant effect to emerge was a main effect of number of exemplars, \(F (1, 119) = 11.13, p = .001.\) The means supported the success of the retrieval fluency manipulation: participants in the three-exemplar conditions indicated that the generation task was easier to complete \((M = 4.78, SD = 1.79)\) than participants in the nine-exemplar conditions \((M = 3.64, SD = 1.99)\)\(^{15}\). No other main effects or interactions emerged (all \(p s \geq .23).\)

**Testing Basic Effects: Group Judgments**

**Change in group stereotyping.** Means and standard deviations for each outcome measure among active elderly participants are shown in Table 7.1. In the active elderly condition, participants were expected to demonstrate an interaction between the number of exemplars and the meta-information cue. This prediction was supported with a significant two-way interaction between number of exemplars and the meta-information cue, \(F (1, 57) = 5.55, p = .02.\) The interaction is depicted in Figure 7.1.

\(^{15}\) A marginally significant main effect of exemplar group was also found, \(F (1, 119) = 3.35, p = .07.\) Despite pilot testing alternative out-groups to equate cognitive fatigue, the means indicated that participants found it slightly easier to retrieve active elderly exemplars \((M = 4.61, SD = 1.86)\) than to retrieve dress-conscious males \((M = 3.92, SD = 2.01)\). This marginal effect is not problematic because if cognitive fatigue is an attributing factor, then participants in the dress-conscious male condition should display even greater retrieval effects in their subsequent group judgments.
Table 7.1

Outcome Measures Among Active Elderly Participants as a Function of Meta-Information Cue and Number of Exemplars (SDs shown in parenthesis)

<table>
<thead>
<tr>
<th>Meta-Information Cue</th>
<th>Absent</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Exemplars</td>
<td>Number of Exemplars</td>
</tr>
<tr>
<td>Stereotypicality</td>
<td>Three (1.17)</td>
<td>Nine (1.31)</td>
</tr>
<tr>
<td></td>
<td>Three (1.12)</td>
<td>Nine (1.35)</td>
</tr>
<tr>
<td>Dispersion</td>
<td>.08 (1.81)</td>
<td>.82 (2.15)</td>
</tr>
<tr>
<td></td>
<td>.08 (2.13)</td>
<td>-.24 (1.64)</td>
</tr>
<tr>
<td>Prejudice</td>
<td>.37 (.82)</td>
<td>.01 (1.09)</td>
</tr>
<tr>
<td></td>
<td>-.39 (1.12)</td>
<td>.02 (1.03)</td>
</tr>
</tbody>
</table>

Notes. ** Significantly different from zero at, \( p \leq .05 \), * \( p = .08 \).

Meta-information cue: absent participants demonstrated more stereotype change in the three-exemplar condition (\( M = 1.56, SD = 1.17 \)) than in the nine-exemplar condition (\( M = .68, SD = 1.31 \)), thus displaying a classic retrieval fluency effect. The difference between the three and the nine exemplar conditions was marginally significant, \( F(1, 28) = 3.70, p = .07 \). This effect replicates the effect detected in Studies 1 and 2. It is worthwhile to note that both participants retrieving three exemplars (\( t(16) = 5.47, p < .001 \)) and participants retrieving nine exemplars (\( t(12) = 1.89, p = .08 \)) displayed a sizeable change from the baseline condition. This result replicates Study 1.

Figure 7.1: Change in group stereotyping for participants retrieving active elderly exemplars as a function of meta-information cue and number of exemplars (error bars: 95 % CI). * Significantly different from zero at \( p < .05 \).
In the meta-information cue: present condition, the retrieval fluency effect was nullified as signaled by a non-significant effect of exemplar number, $F(1, 29) = 1.96, p = .17$. Here, if anything, participants in the nine-exemplar condition showed a tendency to change their stereotype more ($M = 1.30, SD = 1.35$) than participants in the three-exemplar condition ($M = .68, SD = 1.12$), thus displaying a tendency for a bookkeeping effect. This nullification of the retrieval fluency effect replicates Study 2. Participants provided with a meta-information cue also demonstrated change from the baseline, with three-exemplar participants ($t(16) = 2.52, p = .02$) and nine-exemplar ($t(13) = 3.61, p = .003$) participants both displaying a significant change from zero. This null effect is consistent with the results of Study 2.

**Change in group dispersion and prejudice.** The 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) between subjects ANOVA was used to test for differences in group dispersion and group prejudice. Replicating the results in the previous two studies, there were no significant main effects or interactions on these two outcome measures, all $ps \geq .14$.

**Testing Mediating Processes**

I followed the three-step preliminary test presented in Study 1 to test the easy-thus-liked and the easy-thus-typical heuristics. Below are the results of these analyses. As in Study 2, I followed the same principle of focusing on the participants retrieving three or nine active elderly exemplars in the absence of a meta-information cue because the tested variables are relevant to basic retrieval fluency effects. Refer to Appendices VI and VII for results across all conditions.

**The easy-thus-liked heuristic.** On the change in group stereotyping measure number of exemplars produced a marginally significant effect ($F(1, 28) = 3.70, p = .07$), but a non-significant effect on the feeling thermometer $F(1, 28) = 1.67, p = .21$. In the second preliminary step, I used the feeling thermometer as a covariate, and ran the number of exemplars (three vs. nine) ANCOVA on the stereotype change index. There was no decrease in significance of the effect of number of exemplars on the stereotype change index, from $F(1, 28) = 3.70, p = .07$, to $F(1, 27) = 8.02, p = .01$ on the ANCOVA. If anything, the covariate bought about a clearer retrieval fluency effect. In the third preliminary step, the feeling thermometer was found to co-vary with the group stereotyping index, $r = .38, p = .04$, suggesting that as change in group stereotyping increased there was also an increase in warmth towards the elderly. The lack of effects of exemplar number on group likeability and the inability of group likeability to reduce the effect of exemplar number on group stereotyping confirms that the pattern of group stereotyping produced by the number of exemplars manipulation was not explained by variation in likeability.
The easy-thus-typical heuristic. Number of exemplars produced a non-significant effect on the exemplar typicality index ($F(1, 28) = 1.75, p = .20$), which did not mimic the marginally significant effect of number of exemplars on the change in group stereotyping index ($F(1, 28) = 3.70, p = .07$). The mean exemplar typicality index revealed that participants retrieving three counterstereotypical exemplars perceived the exemplars to be more typical ($M = 3.81, SD = 1.11$) than the participants retrieving nine counterstereotypical exemplars ($M = 3.33, SD = 1.08$). In the second preliminary step, I used the typicality index as a covariate, and ran the number of exemplars (three vs. nine) ANCOVA on the stereotype change index. There was a reduction in the size of the effect of number of exemplars from $F(1, 28) = 3.70, p = .07$ to $F(1, 27) = 2.46, p = .13$. Running standard Sobel’s test revealed that this difference was non-significant ($Z = .96, p = .33$). The third preliminary test looked at correlation analyses. Exemplar typicality entertained a meaningful co-variation with the change in group stereotyping index, $r = .31, p = .02$. This suggests that as typicality of the retrieved counterstereotypical exemplars increased, change in group stereotyping also increased.

Although number of exemplars is partly explained by variations in exemplar typicality, the lack of significant effects of the exemplar typicality index indicate that the full mediational model is not significant (Baron & Kenny, 1986, p. 1177).

Discussion
The first goal of Study 3 was to investigate alternative explanations for the role of meta-information in retrieval fluency effects. In particular, I investigated depth of processing and cognitive fatigue. A second goal of Study 3 was to replicate the findings from Studies 1 and 2 by providing additional support for the bias correction portion of my dual process model.

Ruling out Alternative Explanations
Earlier in this chapter, I demonstrated that previous retrieval fluency work has varied the amount of detail required by participants to complete the retrieval task. For example, Dijksterhuis et al. (1999) asked participants to ‘list’ retrieved material, while Schwarz et al. (1991) asked people to ‘describe’ retrieved material. I proposed that retrieval fluency effects may reflect differences in depth of processing, rather than the influence of meta-information per se. This alternative consideration was addressed by varying the position of the exemplar description task. Some participants were encouraged to more thoroughly process the exemplar information by being asked to describe each exemplar at the time of retrieval and before the key group measures. While other participants were encouraged to superficially process the
exemplar information by being asked to list each exemplar (and provide the detailed description later).

As expected, the four-way interaction between number of exemplars, meta-information cue, exemplar group and position of the description task was non-significant. Importantly, all interactions involving the position of the description task were non-significant. These findings indicate that increasing the depth of processing did not account for the differences in magnitude of retrieval effects across member-to-group generalisation studies. These findings, however, do not rule out all involvement of depth of processing in retrieval fluency effects. Depth of processing has been shown to influence retrieval fluency. For example, Rothman and Schwarz (1998) demonstrated that increasing personal relevance reduced the impact of retrieval fluency. What these findings suggest is that, within the context of member-to-group generalisation retrieval fluency effects are not moderated by changes in depth of processing of the exemplar information.

I proposed that cognitive fatigue may be responsible for retrieval fluency effects, and tested this proposal by including a stereotype irrelevant-retrieval group. If cognitive fatigue was responsible for the retrieval fluency effect, participants in the irrelevant-retrieval group (dress-conscious males) should have displayed similar effects as participants in the relevant retrieval group (active elderly). This prediction was not supported. The three-way interaction between exemplar group, number of exemplars and meta-information cue revealed that participants in the dress-conscious male conditions were not significantly different from each other. This finding extends retrieval fluency research by successfully ruling out the alternative explanation of cognitive fatigue.

The Impact of Meta-Information

Evidence of a marginal retrieval fluency effect was found amongst participants in the meta-information cue absent condition where generating less active elderly exemplars resulted in more stereotype change. The robustness of this retrieval fluency effect confirms the important role played by meta-information in the process of member-to-group generalisation. Interestingly, and in line with Study 1, the evidence of stereotype change from the baseline in both the nine-exemplar and three-exemplar conditions demonstrates that retrieved counterstereotypical information is included in the subsequent out-group judgment. The generalisation advantage of retrieving three exemplars over nine exemplars suggests that the meta-information associated with exemplar retrieval influences the extent of generalisation.

Providing participants with an explicit meta-information cue nullified the marginal effect of retrieval fluency, replicating Study 2. This finding further supports the dual process account of retrieval fluency effects, whereby participants demonstrate the ability to correct for
the biasing influence of the retrieval meta-information. The nullification of retrieval fluency effects following the presentation of a meta-information cue is in line with previous retrieval fluency research. For example, Haddock et al. (1999, Study 2) found that participants did not display a retrieval fluency effect when they were presented with a congruent meta-information cue. Interestingly, Haddock et al. also included a meta-information cue that was incongruent. They found that the incongruent cue did not interrupt the retrieval fluency effect, and argued that this occurred because the cue was non-informative in regards to the participants’ actual retrieval experience. Study 4 will investigate the role of congruent and incongruent meta-information cues in member-to-group generalisation.

As in the previous studies, the retrieval fluency effects did not generalise to the outcome measures of perceived group dispersion and out-group prejudice. These consistent null findings support the call from the meta-analysis presented in Chapter 2 for a continued need to investigate these outcome measures using a variety of affective out-groups.

Also, and consistent with the previous studies, the identified effects were not explained by differences in the quality of retrieved exemplars. Amongst participants in the nine-exemplar condition there was no significant change in quality of exemplars retrieved at the beginning of the task compared to those retrieved at the end of the task. The measure used in this study (i.e., assessing for the level of exemplar activity) was different from the exemplar vividness measure used in Studies 1 and 2. This modification addressed the criticism that the measure of exemplar vividness was tapping description length and not closely matched to the retrieval task. Consistent with the review by Schwarz (1998), this demonstrates that the potential impact of exemplars decreasing in quality as the task continues does not account for retrieval fluency effects.

The investigation of affective (easy-thus-liked) and cognitive (easy-thus-typical) explanations of the retrieval fluency effects was continued in this study. There was no support for the easy-thus-liked heuristic. There was evidence that the easy-thus-typical heuristic operates as a covariate in the retrieval fluency effect, however, there was insufficient data to investigate a full mediation model.

Conclusion

The results from Study 3 indicated that the influence of meta-information in member-to-group generalisation is not explained by differences in depth of processing or cognitive fatigue. Rather, retrieval meta-information is included (automatically) with the retrieved content to help formulate out-group judgments unless congruent task cues are provided to stimulate a process of bias correction.
Study 4

The aim of Study 4 was to systematically vary the nature of the meta-information cue, in an effort to extend our understanding of the ‘controlled’ aspect of the dual process model. In Studies 2 and 3, I provided a meta-information cue that was congruent with the participant’s retrieval experience. Hence, participants in the three-exemplar (nine-exemplar) condition were cued that the task was easy (difficult). Across both studies, the meta-information cue resulted in a null effect of the retrieval fluency manipulation. I have argued that the meta-information cue encourages participants to engage in a bias correction process that avoids the influence of automatic retrieval fluency effects. The present study will further examine the nature of the meta-information cue by providing some participants with a congruent meta-information cue and others with an incongruent meta-information cue. It is possible that simply presenting participants with a cue is sufficient to engage a bias correction process, however it is also possible that the cue needs to be presented as congruent to the retrieval task to engage a bias correction process.

This is not the first retrieval fluency study to manipulate the nature of the meta-information cue. As reviewed in the introduction above, previous research differently informs my predictions about the impact of this manipulation. For example, Schwarz et al. (1991) found that an incongruent meta-information cue resulted in a null effect of retrieval fluency. While Haddock et al. (1999) found that an incongruent meta-information cue had no impact on bias correction, and the retrieval fluency effect remained.

There are, at least, two reasons that might explain the mixed effects of an incongruent meta-information cue. Firstly, the presentation of the meta-information cue might affect the bias correction process. In the research by Schwarz et al. (1991, Study 2), where the incongruent cue resulted in a null effect of retrieval fluency (i.e., engaged a bias correction process), the participants were given a subtle verbal cue about the nature of the task indicating how other participants had experienced the task. In the research by Haddock et al. (1999), where the incongruent cue did not influence the retrieval fluency effect, participants were provided with an explicit event (i.e., background music) that was made ‘responsible’ for their experience. It is possible that the more subtle meta-information cue (i.e., being told about ‘other people’s’ experience) is sufficient to engage a bias correction and the direction of the cue is not important. On the other hand, when the cue is more explicit the congruency of the cue is also important. I will use a subtle cue to manipulate the meta-information cue. This manipulation is in line with Schwarz et al., so I expect that participants provided with the congruent or incongruent meta-information cue will all demonstrate a null effect of the retrieval fluency manipulation.

A second reason that may explain the mixed results from the manipulation of incongruent meta-information cues is the anticipated experience associated with the retrieval
task. In previous retrieval fluency research, participants have been alerted to the real nature of the task prior to its completion. For example, Schwarz et al. (1991) began the research by informing the participants that they would be required to provide either 6 or 12 examples of their own assertive behaviour. This direct instruction should invoke an immediate anticipated state of task difficulty. This immediate anticipated experience may undermine the impact of the incongruent meta-information cue, resulting in null effects. That is, many people would expect to find it difficult to recall 12 examples of assertive behaviour and being told that this task would be ‘easy’ may simply be ignored or dismissed in lieu of the anticipated experience. This study will control for the potential confounding influence of anticipated experience by removing all verbal and visual cues about the real nature of the task. The only information available at the start of the task will be the meta-information cue, thus accounting for the criticism that anticipated experience may be diluting the impact of the meta-information cue.

In this study, participants will be tested individually. The one-to-one nature of Study 4’s administration makes the re-introduction of the IAT feasible. With the inclusion of the IAT, the impact of congruent vs. incongruent meta-information cues will be assessed using both explicit and implicit outcome measures. It is expected that the presentation of non-explicit meta-information cues will be sufficient to engage bias correction, resulting in a null effect of retrieval fluency on both the explicit and implicit stereotyping measures.

As this study does not involve a basic retrieval condition (i.e., no meta-information cue) no investigation of the cognitive and affective mediating variables will be carried out.

Method

Participants and Design

Participants were 84 first year psychology students (29 males, 55 females; mean age 21.48 years; SD = 3.71 years) at the University of Newcastle who received 1% course credit for their participation. The majority of the participants indicated that their first spoken language was English (97%; N = 82) and that their cultural background was Australian (94%; N = 80).

The experiment followed a 2 (number of exemplars: four vs. nine) x 2 (quality of meta-information cue: congruent vs. incongruent) between subjects factorial design\(^{16}\). An appended no-exemplar control group completed the dependent measures only. There were 17 participants in each experimental condition and 16 participants in the control condition.  

\(^{16}\) Due to a computer programming error the number of retrieved exemplars in the ‘few’ condition was increased from three-exemplars to four-exemplars, changing the easy-hard ratio from 1:3 to 1:2.25. It is not expected that this change will influence the results. Previous fluency research has used a smaller ratio, for example Rothman et al. (1997) used a ratio of 1:2, with participants in the easy condition listing three examples and participants in the hard condition listing six examples. Haddock et al. (1999) used a ratio of 1:2.3, with participants in the easy condition listing three examples and participants in the hard condition listing seven examples (but see, Tormala et al., 2007).
Preliminary analyses revealed no effects of gender, so this variable will no longer be considered.

**Procedure**

Study 4 closely followed the procedure used in Study 1, with participants being tested individually by a female experimenter who randomly assigned participants to each condition. As in Study 1, participants completed some tasks on the computer and some tasks using pen and paper. Below I focus on the changes unique to Study 4.

Participants were required to retrieve either nine exemplars or four exemplars. To remove anticipated cues, participants were all asked to identify individual elderly people that they considered to be active. The computer displayed a request to generate an active elderly person, and the participant was required to type the details of one exemplar and press ‘enter’. This screen was presented four or nine times, according to condition. If a participant asked how many people they would be required to recall they were told: “You will be asked to think of a number of different people. The computer is programmed to stop when you have reached the end”.

The meta-information cue was again presented prior to, and immediately after, the exemplar generation task. As in Studies 2 and 3, participants allocated to the congruent meta-information cue condition were presented with a cue that was congruent with their retrieval experience. For example, participants in the four-exemplar/congruent condition read: “This task is (was) quite simple. Most people find (found) this easy to complete”. Participants in the nine-exemplar/congruent condition read that the task was difficult. Participants allocated to the incongruent meta-information cue condition were presented with a cue that was incongruent with their retrieval experience. For example, participants in the four-exemplar/incongruent condition read: “This task is (was) quite hard. Most people find (found) this difficult to complete”. Participants in the nine-exemplar/incongruent condition read that the task was simple.

At the completion of the exemplar generation task, the participants were required to provide a brief description of each generated exemplar. This description was limited to approximately three sentences in length. The computer was programmed to record the total time taken to complete the exemplar generation and description tasks. This response latency was expected to reveal that participants generating four-exemplars completed the task in a shorter time than participants generating nine-exemplars. As in Study 1, participants were not informed that the computer was recording their response latency as it was not necessary for participants to feel rushed to complete the task.

**Dependent measures.** The dependent measures closely mirrored Study 1, with participants first completing the explicit trait ratings on paper and then completing the IAT on the computer. The explicit stereotypicality index was created by averaging the responses to
traits: active (reverse scored), fast (reverse scored), strong (reverse scored), frail and slow\textsuperscript{17}. Responses were internally consistent (Cronbach’s $\alpha = .64$). The stereotypicality index was subtracted from the mean of the control group ($M = 5.47$, $SD = .23$) to create a differential score, depicting change in group stereotyping. Higher scores indicate greater change in explicit group stereotyping. Two participants were found to be more than 2.5 standard deviations away from their group mean and were replaced with their group mean for the subsequent analysis.

A measure of change in group dispersion was created by averaging the same traits from the stereotype change index (active, fast, frail, slow and strong; Cronbach’s $\alpha = .77$). This index was subtracted from the mean of the control group ($M = 3.79$, $SD = 1.62$). After this manipulation higher scores depict greater change in group dispersion. A measure of change in prejudice was created by reverse scoring and averaging the responses to the traits ‘likeable’ and ‘pleasant’ (Cronbach’s $\alpha = .65$, Pearson’s $r = .67$, $p < .01$). A differential score was created by subtracting the mean of the control group ($M = 3.07$, $SD = 1.15$) from the prejudice index to represent change in prejudice from a baseline. Following this data manipulation, positive scores on the prejudice index indicate greater change in prejudice.

After the explicit trait ratings, the participants completed the IAT. An implicit stereotypicality index was created following the same procedure detailed in Study 1. First, a D-index was created to indicate implicit stereotyping. Using the D-index, a significant IAT effect was detected in the control group (no counterstereotypical exemplars) when tested using a one-sample $t$-test against zero, $t (15) = 10.82$, $p < .001$. A differential score was then created by subtracting the mean of the control group ($M = .94$, $SD = .35$) from the implicit index of the experimental participants. After this manipulation, positive scores indicate a reduction in implicit group stereotyping and negative scores indicate an increased reliance on the implicit group stereotype.

Following the IAT, the computer was programmed to display the name and description of each exemplar so participants could complete the exemplar quality measure. This measure followed Study 3, whereby participants indicated how active they perceived each exemplar (1 = Not at all; 7 = Extremely). Participants then completed the final section of the research on paper. This section also included any guesses about the hypothesis and demographic measures. Finally, the experimenter queried each participant about his or her recollection of the meta-information cue (for a similar approach, see Kunda & Oleson 1995, Study 3). Participants were then fully debriefed and thanked for their participation.

\textsuperscript{17}The responses for ‘forgetful’, ‘inactive’ and ‘weak’ detracted from the reliability of the scale and were excluded from this index.
Chapter Seven

Results

Manipulation Checks

Number of exemplars. The total time taken to retrieve the exemplars (including the description) was used as the manipulation check for number of exemplars. A 2 (number of exemplars: four vs. nine) x 2 (quality of meta-information cue: congruent vs. incongruent) between subjects ANOVA was run with total time entered as the dependent variable. As expected, the only significant effect was a main effect of number of exemplars, $F(1, 64) = 61.32, p < .001$. Participants in the four-exemplar condition were significantly faster ($M = 149$ seconds, $SD = 54$ seconds) in retrieving and describing their exemplars than participants in the nine-exemplar condition ($M = 264$ seconds, $SD = 67$ seconds). No other effects emerged, all $p$s $\geq .66$. This measure shows that participants retrieving four exemplars were successfully completing the retrieval task faster than participants retrieving nine exemplars.

Meta-information cue. To check the effectiveness of the meta-information cue manipulation, during debriefing participants were probed to see if they had noticed the meta-information cue. All participants indicated that they noticed the cue. Furthermore, the only participants to ask how many more exemplars they were required to generate during the retrieval task were in the nine-exemplar/incongruent condition ($N = 6; 35\%$). This observation suggests that the quality of the cue manipulation had been successful at producing meta-cognitive considerations of incongruency, as these participants were the only participants being asked to generate many exemplars but being told that the task was easy.

Exemplar quality. The exemplar quality measure was used to ensure that participants in the nine-exemplar condition were not retrieving poorer quality exemplars towards the end of the task. The responses for each exemplar were collated to create a mean perceived active level for the first four exemplars and another mean perceived active level for the last five exemplars. These scores could range from 1 – 7 with higher scores indicating that the exemplars were perceived as more active (i.e., higher quality exemplars for the retrieval category). Using these two mean scores, I hoped to demonstrate that exemplars 1 – 4 and exemplars 5 – 9 were not perceived as being statistically different in quality. A paired sample $t$-test demonstrated that there was no difference in quality between exemplars 1 – 4 ($M = 5.21, SD = .96$) and exemplars 5 – 9 ($M = 5.00, SD = .94$), $t(33) = 1.47, p = .15$. This analysis indicates that changes in exemplar quality are unlikely to explain the results.

Testing Basic Effects: Group Judgments

Means and standard deviations for all outcome measures are presented in Table 7.2.

Change in explicit group stereotyping. The measure of change in explicit group stereotyping was submitted to a 2 (number of exemplars: four vs. nine) x 2 (quality of meta-information cue: congruent vs. incongruent) between subjects ANOVA. There was a main
effect of number of exemplars, $F(1, 64) = 6.93, p = .01$, with participants in the nine-exemplar condition changing their explicit stereotype more ($M = .82, SD = .92$) than participants in the four-exemplar condition ($M = .27, SD = .84$). This was qualified by a significant two-way interaction, $F(1, 64) = 4.38, p = .04$. These results are depicted below in Figure 7.3.

Table 7.2

Outcome Measures as a Function of Meta-Information Cue and Number of Exemplars (SDs shown in parenthesis)

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Incongruent</th>
<th>Congruent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Exemplars</td>
<td>Number of Exemplars</td>
</tr>
<tr>
<td>Explicit stereotype change</td>
<td>.44** (.84)</td>
<td>.55** (.91)</td>
</tr>
<tr>
<td>Implicit stereotype change</td>
<td>.28** (.35)</td>
<td>.19** (.27)</td>
</tr>
<tr>
<td>Dispersion</td>
<td>.80* (1.31)</td>
<td>.54 (1.27)</td>
</tr>
<tr>
<td>Prejudice</td>
<td>1.31* (.86)</td>
<td>1.40* (.77)</td>
</tr>
</tbody>
</table>

Notes. Positive scores depict greater change from the baseline.
** Significantly different from zero at $p \leq .01$, * $p \leq .05$.

Figure 7.3: Explicit stereotype change as a function of meta-information cue and number of exemplars (error bars: 95 % CI).
* Significantly different from zero at $p \leq .05$.

The interaction was investigated separately along the levels of the quality of meta-information cue variable. Amongst meta-information cue: congruent participants, there was evidence of ‘overcorrection’ with participants in the nine-exemplar condition demonstrating
greater stereotype change \((M = 1.09, SD = .86)\) than participants in the four-exemplar condition \((M = .10, SD = .44)\), \(F (1, 32) = 11.48, p = .002\). In the meta-information cue: incongruent condition, there was no difference between nine-exemplar participants \((M = .55, SD = .91)\) and four-exemplar participants \((M = .44, SD = .84)\), \(F < 1\).

**Change in implicit group stereotyping.** The interaction between quality of the meta-information cue and number of exemplars was expected to be flat on implicit group stereotyping. To test for this prediction the differential scores created from the IAT D-index was submitted to a 2 (number of exemplars: four vs. nine) x 2 (quality of meta-information cue: congruent vs. incongruent) between subjects ANOVA. There was no interaction and there were no main effects (all \(Fs < 1\)). This analysis supports the role that the meta-information cue (either congruent or incongruent) has in over-riding the automatic retrieval fluency effects that were identified on the IAT in Study 1.

**Change in group dispersion and group prejudice.** There were no effects of the experimental manipulations on change in group dispersion or change in group prejudice (interactions \(Fs < 1\); main effects all \(ps \geq .11\)). This is consistent with the previous studies. However, as noted in Table 7.2, the dispersion ratings provided by participants in the four-exemplar conditions and prejudice ratings across all of the experimental conditions revealed a significant change from the baseline.

**Discussion**

The present study demonstrated that the mere presentation of a meta-information cue is sufficient to wipe off the retrieval fluency effect. The generalisation advantage amongst participants in the nine-exemplar condition indicates that it is not the direction of the cue (congruent or incongruent) that is important for member-to-group generalisation, rather it is the provision of a cue. Simply warning participants about other people’s experience of retrieval was sufficient to engage a bias correction process that removes the automatic effects of retrieval fluency on group judgments.

The results in the meta-information cue: congruent condition is inconsistent with my results in Studies 2 and 3. In Studies 2 and 3 the provision of the meta-information cue resulted in a null effect of the retrieval fluency effects. In this study, the presentation of the congruent meta-information cue resulted in significantly greater change in group stereotyping for participants in the nine-exemplar condition compared to those in the four-exemplar conditions. Minor changes in the retrieval task for participants in the ‘easy’ condition may explain these findings. In particular, participants in this study were required to retrieve four active elderly exemplars rather than three active elderly exemplars. Perhaps the retrieval of three exemplars in Studies 2 and 3 was comparatively easier, making the meta-information
cue insufficient to ‘disrupt’ the automatic process. Either way, the presentation of a congruent task cue removes the effect of retrieval fluency.

The *incongruent* cue demonstrated a null effect of the retrieval fluency manipulation. As noted earlier, previous research providing participants with an incongruent cue has demonstrated contradictory results. In some studies, the incongruent cue was unable to initiate a bias correction process when the cue was attributed to an explicit object such as background music, leaving the retrieval fluency process intact (Haddock et al., 1999; Schwarz et al., 1991, Study 3). In other studies, the incongruent cue was able to initiate a bias correction process resulting in a null effect of retrieval fluency when the cue was subtle and attributed to previous participants experience of the retrieval task (Rothman & Hardin, Study 2, 1997; Schwarz et al., 1991, Study 2). In this study the meta-information cue was subtle, informing participants how ‘other’ people have experienced the retrieval task; this cue seems to operate in the same direction as the congruent cue. That is, being informed about others' subjective experience was sufficient to focus the participants’ attention on their own phenomenal experience, thus initiating the bias correction process and reducing the influence of the meta-information irrespective of the congruency of the cue.

Previous researchers using subtle cues attributed the null effects of incongruent meta-information cues to idiosyncrasies within their research (e.g., Rothman & Hardin, 1997; Schwarz et al., 1991). For example, Rothman and Hardin argued that the null effect was because of the specific trait (assertiveness) that was recalled by their participants; while Schwarz et al. argued that the null effect was because of the nature of the recall task (self-memory). However, the combined weight of these studies demonstrating a null effect of subtle incongruent cues suggests that the explanation lies, at least partly, in the nature of the cue, rather than solely in different designs. It would seem that the use of a subtle meta-information cue is sufficient to initiate a bias-correction process in the same direction as a congruent cue. Future research would be able to help reconcile this position by jointly investigating the role of explicit and subtle incongruent meta-information cues.

This study has demonstrated that the impact of the incongruent meta-information cue is not explained through anticipated difficulty of retrieval experience. By avoiding all cues about the nature of the actual retrieval task any influence of anticipated experience was removed. Participants were not informed about the required number of exemplars, and were unable to look ahead due to the nature of the computer program. Thus the only information

---

18 Hansen and Wänke (2008) further distinguished between implicit and explicit cues. An implicit cue induces a feeling of ease that is unrelated to the subsequent retrieval task. An explicit cue is directly related to the subsequent retrieval task. In Hansen and Wänke’s research, participants were implicitly cued (via a priming task) to induce an expected state of ease vs. difficulty. In my research, participants were subtly cued about the actual nature of the task. According to Hansen and Wänke’s distinction, this subtle meta-information cue remains an explicit cue, and is therefore different from the primed-expectancy research conducted by Hansen and Wänke on retrieval fluency.
available to participants during the retrieval task was the meta-information cue. This procedure ensured that the incongruent cue was not disregarded by participants as irrelevant to their anticipated retrieval experience.

I expected the results of the IAT to mimic the results of the explicit stereotype change index. That is, in this study, I was expecting that both the explicit and implicit measures would show a null effect of retrieval fluency after the presentation of congruent and incongruent meta-information cues. This finding was supported, with the IAT showing no influence of the joint action of the number of exemplar and the meta-information cue manipulations. The inclusion of an implicit outcome measure is novel in retrieval fluency research and rare in member-to-group generalisation research (e.g., Blair et al., 2001). It would be valuable for further retrieval fluency and member-to-group generalisation research to include implicit measures to expand our understanding of automatic meta-cognitive processes.

Altogether, this study demonstrates that the provision of a meta-information cue facilitates a controlled bias-correction process that cancels out any influence of retrieval meta-information on the process of member-to-group generalisation. That is, raising conscious awareness through a non-explicit meta-information cue, that is either congruent or incongruent with the retrieval task, effectively eliminates the automatic effect of retrieval fluency on group judgments.

General Discussion

The purpose of this chapter was to further test the dual process model that I have advanced to explain the influence of retrieval meta-information on member-to-group generalisation. This dual process model accounts for the automatic inclusion of retrieval meta-information in out-group judgments under default conditions, and suggests that such automatic effects will be controlled through a process of bias correction. Study 3 explored alternative explanations for the role of retrieval meta-information and Study 4 explored the nature of meta-information cues that may stimulate bias correction processes.

Ruling out Alternative Accounts for the Effects of Meta-Information

In this chapter, I pointed out that retrieval fluency effects may reflect differences in depth of processing, rather than the influence of meta-information. The evidence regarding the manipulation of the description task in Study 3 demonstrates that depth of processing did not influence retrieval fluency effects. This finding is important as it indicates that subtle methodological differences in the retrieval request do not account for minor variations in the magnitude of retrieval fluency effects. I also suggested that retrieval fluency effects may be attributed to cognitive fatigue. The evidence presented from the manipulation of the exemplar
group in Study 3 demonstrates that cognitive fatigue does not influence these results. This novel extension of the retrieval fluency literature has important implications. It demonstrates that retrieval fluency effects do not reflect the fact that people are just making tired choices after extensive retrieval. Rather, it is consistent with the argument that people are relying on internal meta-information associated with retrieval when judging the out-group.

Instead of relying on the meta-information, it is also possible that participants in the nine-exemplar condition were basing their out-group judgments on less ‘active’ exemplars. The check of exemplar quality in Studies 3 and 4 addressed this alternative interpretation, and found that change in group stereotyping is not attributable to change in the quality of the exemplars in the latter parts of the nine-exemplar retrieval task. This is in line with the results of Studies 1 and 2 on exemplar vividness, and is consistent with other retrieval fluency literature (for a review, see Schwarz, 1998).

Altogether, my evidence supports and extends Schwarz’s (1998) comprehensive review of the retrieval fluency literature. Schwarz concluded that people build judgments to include information from the retrieved content and their meta-cognitive experience of retrieval. My work extends this conclusion by indicating that people use meta-information also when drawing from specific group members to judge the out-group as a whole. In addition, my work in retrieval fluency effects extends Schwarz’s review by addressing the involvement of depth of processing and cognitive fatigue.

**Exploring Bias Correction**

The finding that retrieval fluency effects can be nullified through the provision of an explicit meta-cognitive cue supports the dual process argument. Under normal circumstances, people rely on retrieval meta-information to help form an out-group judgment. However, when environmental cues encourage detailed processing of the retrieval products then participants rely more thoroughly on the retrieved content and ‘switch’ to a more detailed and systematic processing approach to correct for bias. Importantly, the manipulation of the meta-information cue in Study 4 revealed that the simple provision of a cue (irrespective of its congruency with the task) was sufficient to alert participants to the biasing nature of retrieval fluency effects.

**Limitations**

A limitation of Studies 1 – 4, and of much of the retrieval research, is in the actual manipulation of retrieval fluency (e.g. Caruso, 2008; Greifeneder & Bless, 2008; Rothman & Schwarz, 1998; Schwarz et al., 1991; von Helversen, Gendolla, Winkielman & Schmidt, 2008; Winkielman et al., 1998; but see, Hansen & Wänke, 2008). Retrieval fluency has been predominately manipulated by asking participants in one condition to retrieve fewer
exemplars than the participants in the other condition. It is important to ascertain whether the retrieval fluency effect holds when making retrieval difficult and while keeping the number of exemplars constant between conditions. This would ensure that retrieval fluency effects are not due to the different number of available exemplars between participants in the few and many conditions. Therefore, in Study 5, I will ask all participants to retrieve the same number of exemplars but make the task easier for some participants and more difficult for others (also, see Stepper & Strack, 1993). Importantly, this manipulation will enable me to provide a clear test for the different impact of internal and external meta-information in member-to-group generalisation.

As in Studies 1 and 2, there was no support for the affective easy-thus-liked or the easy-thus-typical heuristic in Study 3. It is possible that the position of the mediating variable may be influencing the lack of findings. In particular, because the predictor variable has been measured after the dependent variable it is impossible to identify if the predictor variable is influenced by the fluency effects or attributed to additional processing associated with measuring the independent variable. Since prejudice measures have not been involved in the retrieval fluency effects, it is unlikely that the easy-thus-liked heuristic will explain the effects. However, it remains likely that the cognitive easy-thus-typical heuristic may explain retrieval fluency. To unequivocally address the applicability of the easy-thus-typical heuristic, Study 5 will manipulate exemplar typicality by asking participants to generate either stereotypical exemplars, moderately atypical exemplars or extremely atypical exemplars. The manipulation of these three levels of typicality will clarify the role typicality plays in retrieval fluency research.

Conclusions

It seems that people have little awareness of the cognitive processes influencing judgments (Fielder & Walther, 2004), and the area of member-to-group generalisation is no exception. I have shown that asking people to retrieve counterstereotypical exemplars results in the inclusion of both the retrieved content and the retrieval meta-information, and this effect is not explained by variations in depth of processing, cognitive fatigue or exemplar quality. Furthermore, the evidence presented in this chapter has shown that when individuals are made aware of the biasing influence of retrieval fluency they discount that experience and rely on the available content for building group judgments (see also, Caruso, 2008). These findings support the dual process account of the role of meta-information in member-to-group generalisation.
CHAPTER 8

INVESTIGATING THE SOURCE OF META-INFORMATION AND THE INFLUENCE OF TYPICALITY (STUDY 5)

The overarching goal of this research has been to identify the role of meta-cognition in member-to-group generalisation. To achieve this, I developed the self-generation paradigm to test the impact of fluency. I have argued that retrieval fluency is an internal source of meta-information that is automatically used to build out-group judgments, unless contextual cues engage a bias correction process. Within the broader framework of social meta-cognition and member-to-group generalisation, there are three issues that remain unresolved. Firstly, the differences between external and internal sources of meta-information have not been empirically investigated. Secondly, it is possible that the retrieval fluency effects identified throughout this thesis may be an artefact of the number of exemplars manipulation. Thirdly, the exact role of exemplar typicality in fluency has not been clarified. In this final study, I aim to address these unresolved issues.

Source of the Meta-Information

In Chapter 3, I argued that meta-information comes from external or internal sources (see also Paolini et al., 2009). In that chapter, I meta-analytically tested the distinction between external and internal meta-information in traditional impression formation research and revealed that both sources equally influence member-to-group generalisation. However, I also recognised that testing internal meta-information within the impression formation paradigm was sub-optimal due to the external provision of exemplar information and the necessary artificial injection of meta-information from an external source (i.e., the experimenter). Hence, my preliminary conclusion that external and internal meta-information similarly influence member-to-group generalisation was to be accepted with some caution. To provide a direct test of external vs. internal meta-information source, it is necessary to manipulate these different sources of meta-information within a single paradigm.

In this study, I will investigate external vs. internal meta-information by manipulating the fluency of presented vs. self-generated exemplar information. As in Study 2, exemplar information will be presented (i.e., external information) or self-generated (i.e., internal information). However, this time instead of manipulating the number of exemplars accessible, I will use a novel manipulation of fluency that makes accessing the exemplar information
easier for some participants (i.e., high fluency condition) and more difficult for other participants (i.e., low fluency condition). Specifically, participants in the high fluency condition will complete the reading/retrieval task free from distractions. Participants in the low fluency condition will complete the reading/retrieval task with continual distractions and will require a mouse-click to continue with the task.

It is possible that accessing internal (counterstereotypical) information through self-generation is more likely to affect the subsequent judgment than accessing external information through presented information. This pattern of findings would disconfirm the null finding for sources of meta-information in Chapter 3. However, it would be consistent with a broad literature demonstrating that self-generation is more effective in promoting attitude change than exposure to already-made arguments (e.g., Janis & King, 1954; King & Janis, 1956; for an overview, see Eagly & Chaiken, 1993, pp. 500-505). For example, Wänke et al. (1996, Study 2) asked participants to retrieve or read arguments in support of, or against, the use of public transport (for further details, see Chapter 6). Fluency was manipulated by asking participants to consider either three or seven arguments. Participants who self-generated arguments demonstrated greater attitude change than participants who were presented with the information in the three argument (high fluency) condition. However, interestingly, the advantage of self-generation disappeared in the seven argument (low fluency) condition. According to Wänke et al., it is not only the available content that is used in attitude construction but also the way that content comes to mind. That is, the subjective experience of information content is a crucial variable in building judgments.

Support for Wänke et al. (1996) will emerge in this study if participants self-generating exemplars demonstrate more stereotype change than participants presented with exemplar information, but only under conditions of high fluency (i.e., task distracter: absent). When fluency is low (i.e., task distracter: present) the advantage of retrieval versus presented information should diminish (see also, Wänke et al., 1996).

Manipulating fluency associated with the presented information extends this current investigation to processing fluency in the social domain. In Chapter 3, I defined processing fluency as the ease with which mental operations concerned with processing external stimulus are conducted (see also Rubin et al., 2010). Typically, processing fluency research reveals that high fluency results in more positive judgments of the stimuli than low fluency (Alter & Oppenheimer, 2008). However, much of this research is primarily focused on non-social judgments (e.g. Bornstein et al., 1990; Jacoby & Dallas, 1981; Posner & Keele, 1968; for an exception, see Rubin et al., 2010). For example, Winkielman and Cacioppo (2001, Study 1) investigated the impact of processing fluency on judgment positivity of different line drawings. Winkielman and Cacioppo asked participants to judge simple line drawings of neutral objects such as a house, bird or aeroplane. Prior to viewing the line drawings,
participants were presented a visual prime that was matched (saw a contour that matched the drawing) or mismatched (saw an un-matched contour). Participants judged the images more positively in the matched (high fluency) condition than participants in the mismatched (low fluency) condition. Study 5 will provide will provide a direct test on the role of processing fluency in member-to-group generalisation (for an indirect extension, see Garcia-Marques & Mackie, 1999; see also Chapter 3).

The novel manipulation of fluency will also provide an important extension to retrieval fluency research, and addresses one of the criticisms associated with Studies 1 – 4. The traditional few-versus-many paradigm requires participants to consider different amounts of information (for a review, see Schwarz, 1998). This may mean that participants in the ‘many’ conditions are drawing on other information to formulate their judgments, other than relying strictly on retrieval fluency (e.g., Tormala et al., 2007; Tormala et al., 2002). For example, participants in the many condition could complete their subsequent judgment less positively than participants in the few condition because they performed more cognitive steps, came closer to running out of relevant recall content, or took more time than expected (von Helversen et al., 2008). Although the measure of exemplar vividness ensured there was no change in the vividness of the retrieved exemplars in the previous studies, it is important to control for any (possible) influence of different numbers of exemplars. Thus, the novel fluency manipulation of task distraction will ensure that all participants are exposed to the same number of exemplars.

**The Role of Exemplar Typicality**

The mediating role of exemplar typicality was not clarified through the investigation of the easy-thus-typical heuristic in Studies 1 – 3. Despite the null findings, there are significant reasons to believe that exemplar typicality influences generalisation. Rothbart and John (1985) suggested that exemplar typicality is crucial at both the information processing and retrieval stages of the generalisation process (see also Rothbart, 1996). This means that exemplar typicality should impact on the extent of generalisation amongst participants in both the presented and self-generated conditions. To test the role of exemplar typicality across both presented information and self-generated conditions, and provide a powerful test of typicality, I will use a moderation-of-process design for Study 5 (Spencer et al., 2005). Within this kind of design, the process variable (i.e., exemplar typicality) is included as an independent variable. All participants will be asked to consider elderly exemplars that are either stereotypical, moderately atypical, or extremely atypical. In the presented information conditions, participants will be told that the exemplar information was generated by previous participants who had been asked to identify elderly people that were either inactive (stereotypical), active (moderately atypical) or extremely active (extremely atypical). In the
self-generation conditions, participants will be asked to retrieve inactive, active or extremely active exemplars from their memory.

**Typicality predictions for presented information.** Within impression formation research, exemplar typicality has been identified as a critical variable in mediating change in concentrated-dispersed studies (for a review, see Hewstone, 1994). However, the exact role of typicality is not straight-forward. On one hand, Rothbart (1985) argued that extreme atypicality should maximise generalisation, while on the other hand Hewstone (1994) argued that moderate atypicality should maximise generalisation. The meta-analysis presented in Chapter 2 revealed that both moderately atypical and extremely atypical exemplars result in generalisation, with a non-significant advantage of extremely atypical exemplar information. With the manipulation of exemplar typicality in this study, I aim to reconcile the mixed evidence on the role of exemplar typically in impression formation research.

Stereotypical exemplars do not challenge the group stereotype and invoke dominant responses, thus the impact of the task distraction should be minimal (Rothbart et al., 1996). Therefore, irrespective of the task distracter, participants presented with stereotypical information should display an increased reliance on the out-group stereotype (or less stereotype change) compared to participants presented with counterstereotypical information.

Participants presented with moderately atypical exemplars should still perceive these exemplars as being close in goodness-of-fit to the stereotype. As a consequence, disconfirming information should also be generalised to the group across both levels of task distracter, however there should be an advantage for participants in the non-distracted (high fluency) condition. Garcia-Marques and Mackie (1999, Study 3; for a detailed review, see Chapter 3) presented participants with stereotype disconfirming exemplars where the bulk of the information was moderately disconfirming. Some participants were asked to rehearse a 9-digit number while reading the sample information (low processing fluency), others while rating the group (low retrieval fluency), and some did not receive any rehearsal task (high fluency). Participants high in fluency displayed significant stereotype change, while participants low in fluency, during encoding or retrieval, did not display stereotype change. For my research, this suggests that participants presented with moderately disconfirming information should show more stereotype change in the non-distracted (high fluency) condition than in the distracted (low fluency) condition.

The presentation of extremely atypical exemplars will provide a test of the subtyping (Hewstone, 1994) and conversion (Rothbart, 1981) models of stereotype change. The conversion model predicts greater stereotype change amongst participants presented with the extremely atypical exemplars (irrespective of task distraction). The subtyping model predicts that the presentation of extremely atypical exemplars should result in the cognitive exclusion (or subtyping) of this exemplar information (for a recent definition of subtyping, see Stewart...
et al., 2006). Hence, participants presented with extremely atypical exemplars should not generalise to the out-group. However, Hewstone (1994) argued that subtyping is a cognitively demanding task and can, therefore, only occur under conditions of sufficient cognitive resources (or high fluency). Therefore, these two models differ in predicting the impact of extremely atypical exemplars. Under conditions of extreme atypicality and high fluency, the subtype model predicts that there will be no stereotype change (i.e., evidence of subtyping) and the conversion model predicts that there will be stereotype change. Under conditions of low fluency, both the subtype and conversion models predict that there will be stereotype change (i.e., evidence of generalisation).

Taken together, the results amongst participants in the presented information condition should reveal a two-way interaction between exemplar typicality and task distracter. The pattern of means amongst distracted participants in the extremely atypical condition will reveal support for either the subtype model or conversion model of stereotype change.

*Typicality predictions for self-generated information.* Across Studies 1 – 3, I tested exemplar typicality as a mediating variable of the generalisation process. The basic premise of the easy-thus-typical heuristic is that under conditions of high fluency participants will perceive the retrieved exemplars as typical, resulting in greater generalisation to the out-group. That is, the internal meta-information associated with easily retrieving exemplars flags typicality, resulting in increased generalisation. In Studies 1 – 3, I used a measurement-of-mediation design and included typicality as a process variable (Spencer et al., 2005). Support for the easy-thus-typical heuristic was not evident. By using a moderation-of-process design, this study has greater potential to reveal support for the easy-thus-typical heuristic.

If my premises are correct, under conditions of high fluency (i.e., non-distracted), participants should associate the ease of retrieval with exemplar typicality and consequently perceive all exemplars to be more typical of the group irrespective of the exact typicality instructions for retrieval. This easy-thus-typical heuristic association should result in generalisation of the exemplars to the out-group judgment, irrespective of exemplar typicality. On the other hand, under conditions of low fluency (i.e., distracted), participants should not associate the ease of retrieval with exemplar typicality, resulting in reduced generalisation as typicality becomes more extreme.

To summarise, among self-generation participants, I also expect a two-way interaction between exemplar typicality and task distraction. Non-distracted participants should show generalisation of the retrieved exemplars to the out-group. Distracted participants should show stereotype maintenance as typicality becomes more extreme. I expect that the easy-thus-typical predictions will be most evident in the extremely atypical conditions, whereby participants retrieving extremely atypical exemplars in the non-distracted
condition (i.e., high fluency) display more stereotype change than participants in the distracted condition (i.e., low fluency).

**Summary of the Aims and Hypotheses**

In summary, in this study I will investigate the source of meta-information by manipulating external and internal meta-information (i.e., fluency). This will allow for a joint test of retrieval fluency and processing fluency. The manipulation of exemplar typicality will reveal the role played by typicality in both the impression formation and self-generation paradigms.

Together, I expect a three-way interaction between information source, task distracter and exemplar typicality to emerge on the change in group stereotyping index. This is expected to reflect a different pattern of results amongst the presented information and self-generation participants. Non-distracted participants presented with extremely atypical exemplars will reveal support for either the subtype model of stereotype change or the conversion model of stereotype change. Amongst self-generated participants, evidence of increased generalisation for non-distracted participants will reveal support for the easy-thus-typical heuristic.

**Method**

**Participants and Design**

Participants were first year psychology students at the University of Newcastle ($N = 375$; 301 female and 74 male), who completed the research for 2% course credit. All participants were 30 years of age or younger ($M = 19.85$ years, $SD = 2.70$). The majority indicated that their first spoken language was English (97%; $N = 365$) and that their cultural background was Australian (93%; $N = 347$).

The design was a three-factor between subjects design: 2 (information source: self-generation vs. presented) x 2 (task distracter: present vs. absent) x 3 (exemplar typicality: stereotypical vs. moderately atypical vs. extremely atypical) with an appended no-exemplar control group. There were between 12 and 27 participants who completed the research in each experimental cell of the design. Due to a computer programming error there were 133 participants allocated to the appended control condition.

**Procedure**

The experiment closely followed the procedure of Study 2, where participants completed the research as an online study using a secure website accessible exclusively to first year psychology students at the University of Newcastle. Similar to Study 2, participants were randomly allocated to either a self-generation or presented information condition. In this study, all participants read/retrieved four exemplars that were stereotypical, moderately
atypical or extremely atypical. Specifically, participants in the stereotypical condition were asked to retrieve/read about elderly people that were defined as “mentally slow on the uptake and/or physically sedentary”. Participants in the moderately atypical condition were asked to retrieve/read about elderly people using the same description from Studies 1 – 4 (i.e., defined as “mentally alert and/or physically on-the-go”). Participants in the extremely atypical condition were asked to retrieve/read about elderly people that were defined as “very mentally alert and/or physically energetic”.

In the presented information condition, I created the descriptions for the moderately and extremely atypical profiles by adapting the presented exemplars from Study 2 by using the descriptions provided by the participants in that study. The four exemplars that were described by Study 2 participants as moderately disconfirming were included as moderately atypical exemplars for this study. Four of the five remaining exemplars were then altered to ensure they were more (i.e., extremely) active. This was achieved by changing some of the adjectives from “very” active to “extremely” active and re-wording some of the activities to induce a sense of extreme atypicality. Then, I created four descriptions for the stereotypical exemplars and asked two independent judges blind to the hypothesis to check these profiles for stereotypicality. Comments provided by the judges were incorporated into the profiles to ensure these exemplars were perceived as stereotypical (for the full list of experimental profiles, see Appendix VIII). All descriptions were between 68 – 74 words in length, and each condition was comprised of two male and two female exemplars. Three were non-famous exemplars, and one was a famous exemplar (Elizabeth Taylor as the stereotypical exemplar; Sean Connery as the moderately atypical exemplar; Queen Elizabeth II as the extremely atypical exemplar). To minimise order effects, the non-famous exemplars were presented randomly, however the single famous exemplar was always presented in position three (out of four).

Unlike in the previous studies, participants were not asked to consider a different number of exemplars; instead, all participants were asked to consider four exemplars. To manipulate fluency, participants were either distracted during the exemplar retrieval/reading task (task distracter: present; i.e., low fluency) or were allowed to complete the retrieval/reading task free from distraction (task distracter: absent; i.e., high fluency). Distracted participants were interrupted during the retrieval/reading task every 28 seconds by an 8 x 6 cm box appearing in the centre of the screen with the message: “Communicating with server”. Participants were unable to continue with their retrieval/reading task until they clicked the ‘OK’ button inside the box. The distraction box was designed to look similar to a Microsoft Word message box that would legitimately appear on screen during computer work.
As in Study 2, participants were asked to provide a description (self-generation participants) or their impression (presented information participants) of each exemplar. This was completed at the time of exemplar generation/presentation. The number of words in these descriptions will serve as a measure of exemplar vividness.

**Dependent Measures.** The dependent measures closely followed Study 2. A stereotypicality index was created by combining the responses to the adjectives ‘active’ (reverse scored; R), ‘fast’ (R), ‘strong’ (R), ‘forgetful’, ‘frail’, ‘inactive’, ‘slow’ and ‘weak’ “described the elderly in general” (1 = Not at all; 6 = Extremely). The stereotypicality index (Cronbach’s $\alpha = .79$) was subtracted from the mean of the control group (dependent measures only; $M = 4.57, SD = .62$) to create a differential score whereby positive scores indicate stereotype reduction and negative scores indicate an increase in stereotype responding. Two participants were more than 2.5 standard deviations away from their group mean on this index, and were replaced with the mean for their condition.

Dispersion was measured by asking participants to indicate “how similar you consider the elderly, as a group, to be on each adjective” (1 = Not at all similar; 6 = Extremely similar). A reliable measure of dispersion (Cronbach’s $\alpha = .84$) was created by averaging the similarity ratings from the traits ‘forgetful’, ‘frail’, ‘inactive’, ‘slow’ and ‘weak’. A differential score for dispersion was created by subtracting the dispersion index from the mean of the control group ($M = 3.98, SD = 1.00$). After this manipulation, positive (negative) scores indicate increased (reduced) perceived out-group dispersion.

Prejudice was measured by asking participants to judge their feelings towards the elderly as a group by indicating how well the traits ‘kind’, ‘likeable’ and ‘pleasant’ characterised the elderly (1 = Not at all; 6 = Extremely). These traits were reverse scored and then combined to create a reliable index of prejudice (Cronbach’s $\alpha = .83$). When subtracted from the mean of the control condition ($M = 2.36, SD = .73$), positive (negative) scores on the prejudice change index indicate a decrease (an increase) in out-group prejudice. Seven participants were over 2.5 standard deviations away from their group mean on this index, and were replaced with the appropriate mean for their condition.

Following the out-group measures, participants were asked to complete measures of exemplar likeability. In this study, the easy-thus-liked heuristic was measured across exemplars by asking participants to indicate how much they ‘liked’ each individual exemplar. Specifically, participants indicated how “likeable” they considered each exemplar (1 = Not at all; 6 = Extremely). The exemplar likeability index was created by averaging the responses.

---

19 The items ‘active’, ‘fast’ and ‘strong’ negatively detracted from the reliability index, and so were not included in this measure of dispersion.

20 There was a data saving error that occurred for 8 (out of 242; 3%) participants in this section of the research. These participants were excluded from the analysis using the exemplar likeability index.
across each exemplar (Cronbach’s $\alpha = .67$). Scores could range from 1 to 6 with higher scores depicting greater perceived likeability. This assessment of the easy-thus-liked heuristic more closely taps the likeability of each retrieved exemplar, rather than relying only on a single global out-group measure, such as the feeling thermometer (e.g., Studies 2 & 3).

The following section of the research included manipulation checks. A self-reported task ease measure was used to check the success of the task distraction manipulation: Participants were asked to indicate how ‘easy’ they found the reading/retrieval task ($1 = \text{Not at all}; 6 = \text{Extremely}$). The manipulation check of exemplar typicality asked participants to “indicate what type of elderly people” they read about or retrieved from their memory. They could select either “inactive”, “moderately active” or “extremely active”.

Following the manipulation checks additional process variables were then measured. This included an unrequested cognition measure, which asked participants how many inactive (moderately or extremely atypical conditions) or active (stereotypical condition) elderly people came to mind during the experiment\textsuperscript{21}. Prior contact was measured by asking participants to indicate how many elderly people they knew across the following three categories: (1) family, (2) friends, and (3) elderly people known through media or stories. For each category participants were then asked to indicate how many were considered active. Following the same procedure as Study 2, a mean ratio index was created that ranged from 0 to 1; where a score of 1 indicated that all of the known elderly people were considered to be active\textsuperscript{22}. Twelve responses were higher than one and these were set to a score of 1.

**Results**

**Manipulation Checks**

**Task distracter.** The responses to the self-reported task ease measure was used to check for any differences between participants in the distracted vs. non-distracted conditions. A 2 (information source: self-generation vs. presented) x 2 (task distracter: present vs. absent) x 3 (exemplar typicality: stereotypical vs. moderately atypical vs. extremely atypical) between subjects ANOVA on self-reported task ease question revealed a marginal main effect of task distracter, $F (1, 229) = 2.98, p = .08$. As expected, participants in the task distraction present condition found it more difficult to complete the retrieval/reading task ($M = 3.91, SD = 1.32$) than participants in the task distraction absent condition ($M = 3.66, SD = 1.40$). There were no other effects, all $ps \geq .19$. This suggests that manipulation of task distracter was appreciably

---

\textsuperscript{21} There was a data saving error that occurred for 12 (out of 375; 3%) participants in the unrequested cognitions section. These participants were excluded from the analysis of unrequested cognitions.

\textsuperscript{22} There was a data saving error that occurred for 11 (out of 375; 3%) participants in the prior contact section. These participants were excluded from the analysis using the prior contact index.
strong (on a 1–6 scale), although future research might look at strengthening this result by increasing distraction perhaps by making the distraction more frequent.

**Perceived exemplar typicality.** The manipulation check for exemplar typicality required participants to select whether they had been asked about inactive, moderately active or extremely active exemplars. Of the 242 experimental participants, 215 (or 89%) selected the correct exemplar typicality category. A chi-square test revealed that participants were not randomly selecting an exemplar typicality category, \( \chi^2(2) = 9.78, p < .05 \).

**Testing Basic Effects: Group Judgments**

**Change in group stereotyping.** Preliminary analyses revealed no effects of gender, and so this variable will no longer be considered. Descriptive values for each of the outcome measures are presented in Table 8.1 as a function of information source, exemplar typicality and task distracter.

Change in group stereotyping was subjected to a 2 (information source: self-generation vs. presented) x 2 (task distracter: present vs. absent) x 3 (exemplar typicality: stereotypical vs. moderately atypical vs. extremely atypical) between subjects ANOVA. The predicted three-way interaction was significant, \( F(2, 230) = 3.40, p = .04 \). I followed up on the three-way interaction by investigating the results for exemplar typicality and task distracter at each level of information source.

Amongst participants in the presented information condition, there was a marginal main effect of exemplar typicality, \( F(2, 119) = 2.64, p = .08 \). Participants in the stereotypical exemplar conditions demonstrated reduced stereotype change (\( M = -.08, SD = .61 \)), compared to participants in the moderately atypical and extremely atypical conditions (\( M = .20, SD = .64 \), and \( M = .18, SD = .83 \), respectively). The main effect of task distracter was non-significant, \( F(1, 119) = 2.80, p = .10 \). A significant two-way interaction between task distracter and exemplar typicality was also found, \( F(2, 119) = 3.35, p = .04 \). Figure 8.1 depicts this interaction. Below, I present the results for following up the two-way interaction by splitting along exemplar typicality.

---

23 The main effects were all non-significant, \( F_s \leq .09 \). There was a significant two-way interaction between information source and task distracter, \( F(1, 230) = 4.74, p = .03 \). In line with the bookkeeping effect, presented information participants demonstrated greater change in group stereotyping when presented with a task distracter (\( M = .21, SD = .71 \)) than without a task distracter (\( M = -.02, SD = .70 \)). In line with a retrieval fluency effect, self-generated participants demonstrated greater change in group stereotyping when retrieving exemplars without a task distracter (\( M = .24, SD = .68 \)) than with a task distraction (\( M = .07, SD = .45 \)). The other two-way interactions were non-significant, both \( F_s < 1 \).
Table 8.1
*Outcome Measures as a Function of Information Source, Exemplar Typicality and Task Distraction (SDs shown in parenthesis)*

<table>
<thead>
<tr>
<th>Information Source</th>
<th>Presented</th>
<th>Self-generation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exemplar Typicality</td>
<td>Exemplar Typicality</td>
</tr>
<tr>
<td><strong>Stereotypicality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>-.14 (.53)</td>
<td>.28 (.82)</td>
</tr>
<tr>
<td>Present</td>
<td>-.01 (.68)</td>
<td>.15 (.53)</td>
</tr>
<tr>
<td><strong>Dispersion</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>-.06 (.90)</td>
<td>-.15 (1.39)</td>
</tr>
<tr>
<td>Present</td>
<td>-.09 (1.08)</td>
<td>.05 (.84)</td>
</tr>
<tr>
<td><strong>Prejudice</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>-.04 (.81)</td>
<td>.05 (.62)</td>
</tr>
<tr>
<td>Present</td>
<td>-.16 (.71)</td>
<td>.08 (.70)</td>
</tr>
</tbody>
</table>

**Notes.** Positive scores depicting greater change in group stereotyping from the baseline.

** Significantly different from zero at \( p < .05 \), * \( p = .07 \).
Participants presented with stereotypical exemplars were expected to demonstrate similar levels of generalisation irrespective of task distraction. This prediction was supported with no significant difference between participants presented with stereotypical exemplars and a task distracter \( (M = -.01, SD = .68) \) and those presented with stereotypical exemplars without a task distracter, \( (M = -.14, SD = .53; F < 1) \). It is worthwhile to note that the negative sign indicates generalisation in the direction of increased stereotype reliance. That is, these participants became more stereotypical in their judgments of the out-group than participants in the control condition, demonstrating generalisation of the exemplar material. These changes, however, were non-significant, both \( ts < 1 \).

Participants presented with moderately atypical exemplars in the non-distracted condition were expected to demonstrate greater generalisation of the presented material compared to the distracted participants. This prediction was not supported. There was no difference amongst participants presented with moderately atypical exemplars without a task distracter \( (M = .28, SD = .82) \) and participants presented with moderately atypical exemplars with a task distracter \( (M = .15, SD = .53), F < 1 \). The change in group stereotyping was also non-significant compared to the baseline, all \( ts \leq 1.29 \).

The manipulation of task distracter amongst extremely atypical exemplars was expected to provide a strong test for either the conversion model or subtype model of stereotype change. The results revealed support for the subtype model. Participants in the
distracted condition demonstrated greater stereotype change ($M = .55$, $SD = .83$) than participants in the non-distracted condition ($M = -.09$, $SD = .71$). This difference was significant, $F (1, 43) = 7.55, p < .01$. In line with the subtype model, participants in the distracted condition were unable to exclude the disconfirming exemplar information from their group stereotype, as they did not have sufficient cognitive resources available to engage in subtyping. Stereotype change was in fact found to be significant only in the distracted condition ($t (18) = 2.87, p = .01$; non-distracted condition, $t < 1$).

When following up on the three-way interaction for participants in the self-generation condition, there were no significant main effects (all $F$s $\leq 2.01$, $p$s $\geq .16$). Although the two-way interaction between task distracter and exemplar typicality was found to be non-significant ($F < 1$), I pursued simple effect analysis. Figure 8.2 displays the results for the self-generation participants.

![Figure 8.2: Stereotype change by exemplar typicality and task distracter for self-generation participants (error bars: 95% CI).](image)

* Significantly different from zero at $p < .07$.

Simple effect analysis revealed that participants self-generating extremely atypical exemplars displayed evidence of the predicted retrieval fluency effect. Participants in the task distraction absent condition (i.e., high fluency) reported greater stereotype change ($M = .42$, $SD = .66$) than participants in the task distraction present condition (i.e., low fluency; $M = .09$, $SD = .34$). This difference, however, did not reach standard levels of significance, $F (1, 35) = 2.44, p = .12$. Moreover, in support of the easy-thus-typical heuristic, participants who self-
generated extremely atypical exemplars in the absence of a task distracter displayed a significant change from the baseline, $t (24) = 3.14, p = .004$. All other conditions were not significantly different from the baseline, all $ts \leq 1.20$.

To test Wänke et al.’s (1996) prediction that self-generation is a more effective approach to attitude change under conditions of high fluency, I investigated the joint role of information source and typicality separately for each level of task distraction. Support for Wänke et al.’s predictions was expected to reveal greater stereotype change in the self-generation conditions compared to the participants in the presented information conditions.

A 2 (information source: self-generation vs. presented) x 3 (exemplar typicality: stereotypical vs. moderately atypical vs. extremely atypical) between subjects ANOVA amongst participants in the non-distracted condition revealed a marginal main effect of information source, $F (2, 124) = 3.07, p = .08$. In line with Wänke et al. (1996), participants in the self-generation condition demonstrated more stereotype change ($M = .24, SD = .68$) than participants in the presented condition ($M = -.02, SD = .70$). The ANOVA results also revealed a trend for an interaction between exemplar typicality and information source, $F (2, 124) = 2.55, p = .08$. Simple effect analysis revealed that the exemplar typicality by information source interaction was driven by participants in the extremely atypical conditions. Participants in the self-generation extremely atypical condition changed their stereotype more ($M = .42, SD = .66$) than participants in the presented extremely atypical condition ($M = -.09, SD = .71$), $F (1, 49) = 6.78, p = .01$. In the moderately atypical and stereotypical conditions, the effects were non-significant, all $ps \geq .18$.

The 2 (information source: self-generation vs. presented) x 3 (exemplar typicality: stereotypical vs. moderately atypical vs. extremely atypical) between subjects ANOVA carried out amongst participants in the distracted condition revealed no effects (all $Fs \leq 2.16, ps \geq .12$). In line with Wänke et al. (1996), these results suggest that stereotype change is maximised in self-generated conditions but only when the retrieval is experienced as easy.

**Change in group dispersion and group prejudice.** A 2 (information source: self-generation vs. presented) x 2 (task distracter: present vs. absent) x 3 (exemplar typicality: stereotypical vs. moderately atypical vs. extremely atypical) between subjects ANOVA was used to test change in group dispersion and change in group prejudice. There were no significant interactions (all $Fs \leq 1$) or main effects (all $Fs \leq 2.34, ps \geq .10$) on these two outcome measures, replicating the results in my previous studies.

**Testing Mediating Processes**

**Unrequested cognitions.** On average, participants indicated they thought of approximately 4 unrequested elderly individuals ($M = 3.75; SD = 4.93$), with 7 participants indicating they thought of more than 10 unrequested elderly people (3%).
To test the potential mediating role of unrequested cognitions on the obtained effects, I used the three preliminary steps (ANOVA, ANCOVA and correlation analysis) to test for a mediating variable that I used in the previous studies. The first step used unrequested cognition as a dependent variable in the three-way ANOVA (information source, task distraction and exemplar typicality). The three-way interaction, two-way interactions and main effects were all non-significant, all $F_s \leq 2.15$, $p_s \geq .12$. Thus, the results using unrequested cognitions did not mimic the findings of the significant three-way ANOVA using change in group stereotyping, $F(2, 230) = 3.40, p = .04$. The second step used number of unrequested cognitions as a covariate in the three-way ANCOVA with change in group stereotyping. This analysis revealed a slight reduction in significance for the three-way interaction from $F(2, 230) = 3.40, p = .04$ to $F(1, 217) = 2.38, p = .09$. The third step revealed no correlation between unrequested cognitions and change in group stereotyping, $r = .07, p = ns$. As only one of the three preliminary tests showed support for mediation, it is apparent that unrequested cognitions are not responsible for my key results. This finding is in line with Study 2.

**Perceived prior contact.** I followed the same three-step analysis (ANOVA, ANCOVA and correlation) to test the mediating role of perceived prior contact on the three-way interaction. When the prior contact index was entered as the dependent variable on the three-way ANOVA all interactions and main effects were non-significant, $F_s \leq 1.12$, $p_s \geq .29$. When prior contact was entered as a covariate in a three-way ANCOVA with change in group stereotyping there was slight change for the three-way interaction from $F(2, 230) = 3.40, p = .04$ to $F(1, 218) = 2.59, p = .07$. There was a correlation between prior contact and change in group stereotyping, $r = .17, p = .01$. This suggests that as stereotype change increases there is an increase in perceived prior contact. The lack of significance in the ANOVA indicates that perceived prior contact did not act as a mediating variable in the obtained effects.

**Easy-thus-liked heuristic.** The easy-thus-liked heuristic was also tested using the three preliminary steps to identify a mediating variable. First, the exemplar likeability index was entered as the dependent variable in the ANOVA (information source, task distraction and exemplar typicality). The results of the ANOVA on the exemplar likeability index (non-significant three-way interaction, $F < 1$) did not mimic the results obtained on change in group stereotyping. Interestingly, there was a significant two-way interaction between exemplar typicality and information source, $F(2, 222) = 11.86, p < .001$ on the exemplar likeability index. This was qualified by a main effect of information source, $F(1, 222) = 23.01, p < .001$ and a main effect of exemplar typicality $F(2, 222) = 28.36, p < .001$. The means reflected greater exemplar liking in the self-generation conditions ($M = 4.79, SD = .85$) than in the presented information condition ($M = 4.22, SD = .92$). The means also revealed less liking for the stereotypical exemplars in both the presented information ($M = 3.40, SD =$.
.72) and self-generated condition (\(M = 4.60, SD = .99\)). Followed by extremely atypical exemplars (presented information: \(M = 4.40, SD = .64\); self-generated: \(M = 4.80, SD = .72\)). Greatest liking emerged for moderately atypical exemplars in both the presented information (\(M = 4.90, SD = .66\)) and self-generated conditions (\(M = 4.92, SD = .83\)).

The second step in the preliminary investigation of a mediating variable includes exemplar likeability as a covariate on the three-way ANCOVA using change in group stereotyping. There was no change in the results with exemplar likeability included as the covariate for the three-way interaction, from \(F(2, 230) = 3.40, p = .04\) to \(F(2, 229) = 3.39, p = .04\). Finally, there was no correlation between exemplar likeability and change in group stereotyping (\(r = .04, p = ns\)). From these analyses, it appears that exemplar likeability is not a mediating variable.

**Exemplar vividness.** Exemplar vividness was tested by measuring the number of words provided by participants in the exemplar description task (word range from 1 – 137 words; \(M = 25.02; SD = 16.22\)). Separate means were created for the length of descriptions for the first two exemplars and the last two exemplars. These means were then compared across conditions. There was no significant change in the length of description between the first two exemplars (\(M = 25.25, SD = 16.73\)) and the last two exemplars (\(M = 24.77, SD = 17.37\)), \(t < 1\). In line with my previous studies, this indicates that participants were not relying on exemplars of a poorer quality in the latter stages of the retrieval/reading task.

**Summary of Main Results**

The predicted three-way interaction between information source, task distraction and exemplar typicality was significant. Further investigation of this effect revealed a general stereotype change advantage amongst participants who were self-generating exemplars over participants who were presented with exemplar information under conditions of high fluency (i.e., distraction absent conditions). In the presented information condition, the subtype model of stereotype change received support with no evidence of subtyping amongst the distracted extremely atypical participants. In the self-generation condition, support for the easy-thustypical heuristic emerged with a significant change in group stereotyping amongst participants easily retrieving extremely atypical exemplars. The alternative explanations of unrequested cognitions, perceived prior contact, exemplar likeability and exemplar vividness were all ruled out.

**Discussion**

The goal of Study 5 was to investigate the impact of external (processing fluency) and internal (retrieval fluency) meta-information on member-to-group generalisation and assess the role of typicality in both the self-generated and impression formation paradigms. This was
achieved through the use of a three-factor design that avoided confounding fluency with sample size. There was a significant three-way interaction between information source, task distracter and exemplar typicality that was not explained by the influence of unrequested cognitions, perceived prior contact or exemplar vividness. I will now discuss the results from the significant three-way interaction.

**External vs. Internal Meta-Information**

Manipulating information source and task distraction allowed external and internal meta-information to be tested within this final experimental study. Under conditions of high fluency (i.e., non-distracted) there was an advantage of internal meta-information over external meta-information, as revealed by the marginal main effect of information source on the stereotype change index. This main effect demonstrated that participants self-generating exemplars displayed greater stereotype change than participants who were presented with out-group exemplars. This finding supports the research by Wänke et al. (1996) who found that participants self-generating information displayed greater attitude change than participants presented with information (see also Fazio et al., 1995). Under conditions of low fluency, however, there was no difference between internal and external meta-information. This was evidenced by non-significant effects of information source in the low fluency (i.e., distracted) condition. To further elaborate on the source of the meta-information, I will now discuss the results separately for external and internal meta-information.

**External Meta-Information and Exemplar Typicality**

The presentation of out-group information under different conditions of task distraction manipulated external meta-information, effectively extending a processing fluency manipulation to the domain of social judgements (for mediating evidence of processing fluency and migrant bias, see Rubin et al., 2010). As expected, I found an interaction between task distracter and exemplar typicality amongst participants presented with out-group exemplar information. I will now discuss this interaction across each level of exemplar typicality.

Participants presented with extremely atypical exemplars generalised to the group only under conditions of low processing fluency (task distracter: present). When processing fluency was high, no generalisation was found. This finding is consistent with the results of Yzerbyt et al. (1999). Yzerbyt et al. used a cognitive load manipulation in a standard impression formation study and found that extremely atypical information was included in out-group judgments only when the participants were busy attempting to recall an 8-digit number. Yzerbyt et al. interpreted these findings as reflecting the fact that there were insufficient cognitive resources available to allocate these extreme members to a subtype.
Similarly, in the present study participants failed to subtype the extremely atypicality exemplars when they were presented with a task distraction. Together with Yzerbyt et al.’s results, this study provides direct support for Hewstone’s (1994) subtype model of stereotype change.

By definition, moderately atypical exemplars are closer in goodness-of-fit to the group stereotype than extremely atypical exemplars, therefore the presentation of moderately atypical exemplars was expected to generalise to the out-group across both levels of task distracter. It was also expected that non-distracted (high fluency) participants would display greater generalisation of the exemplar information to the out-group than distracted (low fluency) participants, due to the greater availability of cognitive resources to engage in stereotype change (Garcia-Marques & Mackie, 1999). The results did not support these predictions. It is possible that the moderately atypical exemplars were not considered to be sufficiently atypical to produce stereotype change in these participants. It is also possible that the distraction task was not strong enough to evoke a sense of low fluency. Further research might focus on the use of a more effective distraction task to help reconcile differences between this study and the previous research by Garcia-Marques and Mackie (1999).

Participants presented with stereotypical exemplars displayed the predicted increased reliance on the group stereotype. The means were in the direction of a processing fluency effect, with more stereotyping amongst participants high in fluency (i.e., non-distracted) than in the low fluency (distracted) condition. However, this difference was non-significant.

In general, it seems that when the typicality of the exemplar does not challenge the group stereotype (i.e., stereotypical and moderately atypical exemplars) the external meta-information associated with high vs. low processing fluency does not significantly influence generalisation. However, when the exemplar typicality challenges the group stereotype (i.e., extremely atypical exemplars), the external meta-information associated with low processing fluency inhibits subtyping. It seems likely that participants high in fluency recognise the extreme atypicality of the exemplars and engage in additional cognitive processes to exclude them from their group judgment.

**Internal Meta-Information and Exemplar Typicality**

By asking participants to self-generate exemplars that varied in typicality, my aim was to clarify the status of exemplar typicality for the retrieval fluency effect. In this study, I found that self-generating extremely atypical exemplars under conditions of high fluency resulted in greater stereotype change than under conditions of low fluency. Although this finding did not reach standard levels of significance, it is consistent with the basic retrieval fluency effect that emerged in Study 1, and was also present in Studies 2 and 3. In addition, participants self-generating extremely atypical exemplars under conditions of high fluency
were the only group to demonstrate a significant change from the baseline, supporting the conclusion that member-to-group generalisation occurred amongst these participants.

It seems that internal meta-information associated with easy retrieval is used to facilitate generalisation to out-group judgments, and this effect occurs only when the retrieved exemplars are extremely atypical. This finding clarifies the validity of the easy-thus-typical heuristic (for details, see Chapter 4). Recall, the easy-thus-typical heuristic was built from extensive social psychological evidence that indicates people readily use member typicality as a basis to include counterstereotypical group members in the group judgment. It seems that when exemplar retrieval is perceived as easy, the sense of ease is associated with the exemplar resulting in increased generalisation, and this effect is maximized as the exemplar becomes increasingly atypical.

Participants retrieving moderately atypical exemplars were expected to display a retrieval fluency effect, with greater stereotype change in conditions of high fluency than in conditions of low fluency. Although the means were in the predicted direction, with participants high in retrieval fluency displaying more stereotype change than participants low in retrieval fluency, this difference was non-significant. This suggests that internal meta-information may not impact on generalisation after retrieving moderately disconfirming exemplars. However, this explanation is unlikely as participants in this moderately atypical retrieval condition were provided with the same retrieval request that was provided to participants in Studies 1 – 4. Instead, it may be that the task distraction manipulation was insufficient to generate different conditions of task ease. However, this also seems unlikely as the manipulation check of task distracter revealed that the manipulation was successful. It is possible that this null effect may have just been an anomaly within the results, and it would be desirable for future research to further investigate this effect.

It is interesting that retrieving stereotypical exemplars did not result in stronger group stereotyping. It is possible that participants retrieving stereotypical exemplars may have recognized they only had limited stereotypical exemplars available, and this difficulty resulted in stereotype change. Oftentimes, there is limited personal experience with stereotyped out-groups resulting in many stereotypes being formed through acculturation (e.g., Hebl & Heatherton, 1998), literature (e.g., Mullen, 2004) and the media (e.g., Gilliam & Iyengar, 2000). In support of this ‘limited contact’ argument, Price and Hsu (1992) found that much of people’s information about stigmatized out-groups (in their research, people with HIV/AIDS) comes from the media. It is possible that retrieving stereotypical exemplars may highlight limited direct experience with the out-group and as a consequence flag the unreliable nature of the retrieved information, resulting in reduced out-group stereotyping.
Implications and Limitations

There are some important implications associated with the finding that external and internal sources of meta-information influence member-to-group generalisation differently according to exemplar typicality. For instance, the finding that exemplar typicality influenced retrieval fluency adds weight to the call by Tormala et al. (2007) that the meta-cognitive process of fluency may be mediated by a number of different variables. My research on fluency effects in member-to-group generalisation suggests that the broader concept of fluency is most likely mediated by factors that are contingent on both the judgment item (e.g., self-judgments vs. other-judgments) and the nature of the retrieval request (e.g., memory retrieval vs. novel thought generation). Future research in retrieval fluency would be well served to investigate the flexibility of mediating variables influencing judgments across a variety of domains.

Building on the point above, the lack of effect of the easy-thus-liked heuristic does not indicate that this is an invalid heuristic for fluency research. Rather, it suggests that the easy-thus-liked heuristic is simply not relevant to the current judgment domain. In particular, I believe that the lack of effects for this affective heuristic may be best explained through the matching hypothesis (see Chapter 2; Paolini et al., 2007). Recall, with the matching hypothesis I argued that generalisation is maximised when the out-group experience is matched to the outcome measure. In this study, the retrieval of exemplar information is a cognitive experience and the easy-thus-liked heuristic is an affective measure. Thus, the lack of effect of the easy-thus-liked heuristic provides indirect support for the role of the matching hypothesis.

Mayer and Tormala (2010) provide more direct support for the role of fluency in the matching hypothesis. Mayer and Tormala (2010, Study 2) found that fluency mediated the impact of a positive message about blood donation. In the first stage of their study, participants were primed with a negative message about blood donation. This message was either framed in affective terms (e.g., “the nurse gripped my arm with one hand and slid the large spike of the needle into my skin”, p. 447), or in cognitive terms (e.g., “because of the physical effects, most first-time donors (about 75%) say they will not donate blood again”, p. 447). Participants then received a second message that contained several arguments in favour of blood donation. The arguments presented in the second message were framed in either cognitive or affective terms. For example, participants read “I think (I feel) that blood donation is the most fantastic thing I can do with 30 minutes of my free time (p. 446)”. Participants then reported their attitudes towards blood donation and indicated how easy they processed the second message (i.e., a measure of processing fluency). Support for the matching hypothesis emerged, with participants in the affective (cognitive) prime condition responding more positively to feel (think) message framing. In addition, it was found that this
matching effect was mediated by processing fluency. Under matched conditions, the second message was easier to process and this resulted in more persuasion than unmatched conditions. Future fluency research could further explore the role of matching fluency experiences with the affective vs. cognitive nature of the subsequent judgment.

Finally, this study introduced a novel manipulation of fluency, through the use of a task distraction. This manipulation avoided confounding fluency with sample size and revealed some evidence for a retrieval fluency effect amongst participants self-generating extremely atypical exemplars. However, it should be noted that the manipulation check revealed only marginal success for the use of a task distracter to induce a sense of task difficulty. I believe that future research should aim to strengthen this manipulation, and therefore strengthen the conclusions drawn above by investigating fluency without manipulating the number of exemplars.

**Conclusions**

The findings from this study demonstrate that external and internal sources of meta-information each play a unique role in stereotype change. Extending the social meta-cognitive research on member-to-group generalisation (see Chapter 2), exemplar typicality was found to influence generalisation across differing sources of meta-information. Ultimately, this study highlights the influential and dynamic role of meta-cognition in member-to-group generalisation.
CHAPTER 9

GENERAL DISCUSSION:
SUMMARY OF FINDINGS, APPLICATIONS, AND FUTURE RESEARCH

Stereotyping, prejudice and discrimination constitute major social problems and social psychologists have been active in searching for remedies (Yzerbyt et al., 1999). Negative stereotypes about social groups (e.g., defined by race, gender, religion, occupation) frequently lead to prejudiced feelings and discrimination against members of those groups (Queller & Smith, 2002). Traditionally, social cognitive interventions to reduce stereotyping have used information-based programs, such as interventions of mass communication aimed to alter biased group perceptions by conveying stereotype-incongruent messages. A new approach to promoting stereotype change may be achieved through programs based on raising the awareness of the ways in which people monitor and carry out member-to-group generalisation (e.g., Paolini et al., 2009; Yzerbyt, Lories & Dardenne, 1998).

This research investigated the impact of fluency on member-to-group generalisation and has added to the scant literature that explicitly demonstrates the applicability of meta-cognitive processes to the process of member-to-group generalisation (e.g., Paolini et al., 2009; Yzerbyt et al., 1999). Through the use of a self-generation paradigm, I have shown that internal and external meta-cognitions act as important moderating factors in the process of member-to-group generalisation. Furthermore, I have provided evidence for a dual process account. I have demonstrated that retrieval fluency operates as an automatic process, whereby both the retrieved content and the meta-information are used to build an out-group judgment. However, alerting people to the influence of fluency on their judgment is sufficient to evoke a controlled process that wipes off the automatic retrieval fluency effect.

In this chapter, I will firstly summarise the key themes and findings identified in each chapter. I will then discuss the contribution of my research in providing a new direction for understanding and researching member-to-group generalisation. I will then look at the methodological and social implications of my research. Finally, I will identify some limitations associated with this research and offer suggestions for future work.

Summary of Key Findings

In Chapter 1, I pointed out that researchers have approached member-to-group generalisation from a social-cognitive or contact-hypothesis tradition. Social-cognitive
researchers adopted the impression formation paradigm to focus on understanding moderating factors of generalisation. Contact-based researchers focused on applying these moderating factors to facilitate generalisation in face-to-face contact situations. In Chapter 2, I meta-analytically reviewed the impression formation literature and concluded that generalisation operates as a default process (see also Bless & Schwarz, 1998; Paolini, 2001). I reviewed evidence supporting a matching hypothesis (Paolini et al., 2007), whereby generalisation was maximised for cognitive out-groups using cognitive measures, and this was attributed to the cognitive nature of the impression formation task. The general trend that emerged from the investigation of moderating variables in Meta-Analysis Study 1 was argued to support Schwarz and Bless’ (1992) inclusion-exclusion model.

In Chapter 3, I elaborated on Paolini et al.’s (2009) distinction between the impact and source of meta-information to argue that meta-information can either facilitate the inclusion or exclusion of the exemplar material from group judgments. I identified fluency research as a particularly interesting and promising meta-cognitive process. I then reviewed research from the retrieval fluency tradition that suggested that meta-information associated with exemplar retrieval was likely to influence generalisation. I proposed that the investigation of fluency meta-information would enhance our understanding of both member-to-group generalisation and fluency processes. I conducted a preliminary test on the role of meta-information on member-to-group generalisation by meta-analytically reviewing impression formation literature that investigated (often implicitly) the direction and source of the meta-information. From this review, I concluded that meta-information was able to facilitate inclusion or exclusion of the exemplar material in out-group judgments, and that more explicit research would enhance our understanding of this important process.

In Chapter 4, I presented the theoretical basis for my dual process account, by suggesting that fluency meta-information is an automatic meta-cognitive process that could be interrupted through engaging bias correction. I then introduced a self-generation paradigm that would be used to test the role of internal meta-information in member-to-group generalisation. The self-generation paradigm required participants to retrieve counterstereotypical exemplars from their memory and then complete out-group judgments.

The self-generation paradigm was then used to investigate meta-cognitions within member-to-group generalisation across five empirical studies. In Chapter 5, I presented a pilot study that prepared the materials for the self-generation paradigm and provided a first empirical test of this hybrid approach to studying member-to-group generalisation. Specifically, in Study 1 participants were asked to retrieve either three or nine active elderly exemplars. Explicit (trait rating tasks) and implicit (IAT) stereotype change measures revealed the predicted retrieval fluency effect, whereby participants in the three-exemplar condition changed their group stereotype significantly more than participants in the nine-
exemplar condition. The congruent results on the implicit and explicit measures of stereotype change were used as initial evidence for a dual process account of meta-information in member-to-group generalisation. Furthermore, I argued that an automatic process must be applicable to the subsequent judgment (Anderson et al., 2007). The lack of stereotype change on the retrieval-irrelevant judgments items was used as additional support for the dual process account. Altogether, I argued that participants automatically use the meta-information associated with the retrieval process to help build up their out-group judgments.

In Chapter 6, information source was manipulated along with the number of exemplars and an explicit (congruent) meta-information cue about the retrieval process was introduced. In Study 2, two key effects emerged. Firstly, in the absence of any meta-information cue the self-generation participants replicated the retrieval fluency effect identified in Chapter 5; while the reverse effect emerged for participants in the presented information condition. Participants presented with nine-exemplars changed their stereotype more than participants presented with three-exemplars. Secondly, the explicit meta-information cue was found to wipe off the influence of the spontaneously occurring meta-information effect. This was reported as further evidence for a dual process account.

In Chapter 7, alternative explanations for the dual process account of fluency were tested. Specifically in Study 3, I revealed that the automatic effects of retrieval fluency were not accounted for by the influence of cognitive fatigue or differences in depth of processing. Thus, the results of this study supported the explanation of an automatic retrieval fluency effect in member-to-group generalisation. In Study 4, I revealed that both congruent and incongruent meta-information cues wiped-off retrieval fluency effects. Interestingly, in this study the inclusion of a congruent meta-information cue reversed the retrieval fluency effect. Participants in the nine-exemplar condition were found to change their explicit group stereotype more than participants in the three-exemplar condition. Together, the evidence in this study supported the bias correction account in the dual process model.

Finally, in Chapter 8, I presented a test of internal and external meta-information by disrupting the retrieval (internal) and the processing (external) of exemplar information. In Study 5, exemplar typicality was also systematically varied. This study revealed that both internal and external sources of meta-information play a unique role in stereotype change that is differently influenced by exemplar typicality. In particular, internal meta-information associated with easy retrieval facilitates inclusion of extremely atypical exemplars while external meta-information associated with easy processing facilitates exclusion of extremely atypical exemplars. These results were used as support for the important and unique role of meta-cognitive processes in member-to-group generalisation.
A Broader Theoretical and Empirical Approach to Member-to-Group Generalisation

Across five empirical studies, I have shown that fluency influences the process of member-to-group generalisation. Hence, this research seriously extends member-to-group generalisation research into the area of social meta-cognition and builds on early research in this direction (see also Paolini et al., 2009; for a review, see Jost et al., 1998). I detail these extensions below.

Information and Meta-Information

Many researchers have argued that meta-information plays a large role in all judgments, with Jacoby, Kelley and Dywan (1989) arguing that research focusing only on information content is negligent (also see Clore, 1992; Jost et al., 1998; Schwarz, 1998). It is evident from this research that the retrieved content is not the only source of information that is used during judgment construction. Rather, information and meta-information play a unique and important role in out-group judgements (for an early overview, see Clore, 1992).

Presented vs. self-generated information. The evidence from the empirical studies in this thesis demonstrates that the different sources of exemplar information impact differently on generalisation. To provide useful guidelines for promoting change in negative out-group stereotypes, social policy makers need to consider the exact source of the information. For presented information, it would seem that maximized generalisation is a case of the ‘more the merrier’. On the other hand, for self-generated information, it would seem that generalisation is a case of the ‘less the merrier’.

Participants in Study 2 who were presented with disconfirming information in the absence of a meta-information cue generalised more to the out-group as the sample size increased, reflecting a case of the ‘more the merrier’. Similarly, the meta-analysis presented in Chapter 2 also supported a disconfirming sample size effect. Adopting Jost et al.’s (1998) expanded approach to understanding social cognition, people seem to rely on the law of large numbers and ‘add’ more weight to the out-group exemplar information as it increases in size. Hence, increases in the sample size contribute to increased generalisation.

Participants in Studies 1 – 3 who self-generated disconfirming information in the absence of externally injected meta-information cues generalised more to the out-group as the sample size decreased, reflecting a case of the ‘less the merrier’. That is, participants high in fluency (three-exemplar) displayed more stereotype change than participants low in fluency (nine-exemplar). This effect occurred even though (paradoxically) high-fluency participants had retrieved less counterstereotypical exemplars than low-fluency participants. Hence, decreases in the sample size contribute to increased generalisation.

Internal vs. external meta-Information. The contrasting effects of presented and self-generated information were attributed to the role of qualitatively different sources of
meta-information. This thesis identified and compared two different sources of meta-information: (1) internal meta-information (i.e., participant controlled) which is associated with retrieval fluency, and (2) external meta-information (i.e., experimenter controlled) which is associated with processing fluency. In the meta-analysis presented in Chapter 3, I provided preliminary evidence from impression formation studies that both internal and external sources of meta-information influence member-to-group generalisation. However, the strongest test for internal meta-information on member-to-group generalisation came from Studies 1–5. The five empirical studies all investigated the role of internal meta-information through retrieval fluency, and enabled me to demonstrate that retrieval fluency impacts on member-to-group generalisation. The use of a no-information control condition demonstrated that meta-information associated with easy retrieval (or high fluency) results in assimilation effects while the meta-information associated with difficult retrieval (or low fluency) does not result in contrast effects, rather in reduced assimilation (but see, Sanna et al., 2009). It is worthwhile to note that the participants in Studies 1 and 3 who retrieved nine exemplars did display stereotype change from the baseline. The change amongst the nine-exemplar participants was just not as extreme as the participants who retrieved three exemplars. Altogether this evidence demonstrates that the internal meta-information of retrieval fluency is a particularly influential moderator of generalisation.

The final empirical study, with presented and self-generated conditions, offered a direct comparison of internal vs. external meta-information. In this study, internal meta-information was manipulated by making the retrieval of four exemplars either easy or difficult, while external meta-information was manipulated by making the processing of four presented exemplars either easy or difficult. In the easy (non-distracted) condition, stereotype change was maximized when participants retrieved their own exemplars from memory rather than when they were presented with the same information (see also, Wänke et al., 1996). That is, when meta-information is associated with ease (i.e., the non-distracted participants), internal meta-information (i.e., self-generation) is more efficient at promoting change than external meta-information (i.e., presented information). On the other hand, and in line with the conclusions from the Meta-Analysis Study 2, when meta-information is associated with difficulty (i.e., the distracted participants), there was no benefit of internal meta-information over external meta-information.

It is worth noting that the provision of a meta-information cue in Studies 2–4 can be treated as a case of external meta-information information. In the meta-information cue condition participants were presented with a congruent cue (Studies 2–4) or an incongruent cue (Study 4). In Studies 2 and 3, participants presented with a congruent cue displayed no significant difference between the retrieval of a few exemplars and the retrieval of many exemplars. Participants in Study 4 presented with a congruent cue, however, demonstrated a
significant ‘reversal’ of the retrieval fluency effect (see also Schwarz et al., 1991, Study 3). The different results across the studies using a congruent meta-information cue are discussed more extensively below in the section on future work. For now, it is important to note that the direction of generalisation on the stereotype change index always indicated stereotype change, even in the case of the significant reversal of retrieval fluency identified in Study 4 after the presentation of a congruent meta-information cue. Hence, external meta-information in the form of explicit meta-information cues evokes a bias correction, but does not produce an over-correction in the form of a contrast effect.

**Summary.** In combination, these findings converge on the conclusion that accessing exemplar information from memory is not a simple by-product of cognitive processing; rather the actual process also carries other critical information. In particular, both information and meta-information impact on, and shape, the process of member-to-group generalisation. These findings should highlight to social-cognitive researchers the heavy role played by subtle cues that populate most research paradigms and may be (incidentally) presented during generalisation research (e.g., cover stories). Apparently, it is not only what comes to mind that determines out-group judgments but also how it comes to mind.

**A Dual Process Account**

I proposed a dual process framework to account for the effect of meta-information on member-to-group generalisation. With this model, I argued that meta-information automatically influences out-group judgments. However, the automatic application of meta-information can be controlled through the provision of an explicit meta-information cue.

In my work, I have explicitly extended the conceptualisation of fluency by arguing that the process of using information and meta-information to formulate a judgment is automatic. This point was supported by the evidence for a marginal retrieval fluency effect from the implicit IAT measure in Study 1. I should note that the results of the IAT presented in Study 1 are not sufficient to build a strong argument for automaticity. Greenwald and Banaji (1995) pointed out that a critical test to identify an automatic process is through the provision of a cue. Specifically, Greenwald and Banaji argued that an automatic effect “may be reduced, eliminated, or reversed when subjects are made aware of the nature of the manipulation” (p. 9). To address this issue, two additional tests of the dual process account were included. Firstly, in Chapter 5, I included a test of applicability. With this test, I found support for Anderson et al.’s (2007) requirement that an automatic process should be applicable to the subsequent judgment. Secondly, a meta-information cue was included in Studies 2 – 4. With the inclusion of a cue, participants were alerted to the meta-cognitive nature of this task. This encouraged a more thoughtful evaluation of the exemplar.
information, and triggered a correction process that wiped off the automatic influence of the meta-information on the group judgement.

It is worth stressing that my dual-process interpretation is slightly different to that of previous retrieval fluency accounts. Most retrieval fluency researchers have followed Schwarz et al.’s (1991) interpretation of the effect as reflecting misattribution (for a review, see Schwarz, 1998; see also, Hansen & Wänke, 2008). According to the mis-attribution account, ease of retrieval is initially mis-attributed to the retrieved content (e.g., “if it so easy to recall these examples there must be many more in my memory”). Accordingly, the high retrieval fluency is mis-attributed to the retrieved content resulting in more positive (or extreme) judgments. However, if the ease of retrieval is attributed to an irrelevant source, such as background music (i.e., an explicit meta-information cue), the fluency is then excluded from the subsequent judgment (e.g., Sanna & Schwarz, 2003; Winkielman & Schwarz, 2001).

The use of a control condition in my research on out-group judgments has shown that participants in the few and the many conditions both change their judgment in the direction of the retrieved content. Since, participants in the many condition are appropriately demonstrating generalisation, it is apparent that in the context of member-to-group generalisation, it is not a ‘mis-attribution’ of the impact of the meta-information, rather my evidence suggests that fluency affects the ‘extent’ of information inclusion (see also Rothman & Hardin, 1997).

Following the presentation of a meta-information task cue, both Schwarz et al.’s (1991) mis-attribution account and my own dual process model argue for bias correction. According to Schwarz et al. the presentation of a congruent task cue renders the subjective retrieval experience as un-diagnostic which makes the meta-information irrelevant to, and therefore not included in, the judgment at hand. Similarly, I have argued that when participants are made aware of the meta-information they attempt to control for this meta-cognitive influence by engaging in a process of bias correction.

My dual process account of meta-cognition in member-to-group generalisation extends the two previous dual process accounts of member-to-group generalisation (for a review on dual process models in social inference, see Kruglanski & Orehek, 2007). Firstly, the inclusion-exclusion model by Schwarz and Bless (1992; Bless & Schwarz, 1998) articulates that (the same) out-group exemplar information can be included or excluded from the group judgment, depending on the nature of contextual cues. Schwarz and Bless argue that the default approach to information processing is one of inclusion, resulting in assimilation of the exemplar information in the out-group judgment (also see, Chapter 2). However, the presentation of contextual cues can encourage exclusion of the exemplar information, resulting in a contrast effect (Bless & Schwarz, 1998). My research provides a direct
extension to Schwarz and Bless by demonstrating that the ‘inclusion’ aspect of their model should also reflect the inclusion of meta-information. In addition, my research also extends Schwarz and Bless’ conceptualisation of ‘exclusion’ cues. The presentation of a meta-information task cue (Studies 2 – 4) could be considered a meta-information exclusion cue. However, the meta-information cue did not result in a contrast effect, rather it merely wiped-off the influence of the internal meta-information associated with the process of information retrieval. Thus, meta-information contextual cues result in a lessened extent of generalisation, rather than a contrast effect.

The second dual process account of member-to-group generalisation was presented by Paolini and colleagues (2009). In this model, and in line with Schwarz and Bless (1992), Paolini et al. argued that meta-cognitive processes would either amplify or attenuate member-to-group generalisation. However, the testing ground of this model is limited to the specific meta-cognitive process of accountability. By applying fluency considerations to member-to-group generalisation, I have demonstrated that this meta-cognitive process plays an influential role in out-group judgments. Together, Paolini et al.’s work in accountability and my work in fluency should encourage further work in the application of meta-cognition to member-to-group generalisation (see also, Jost et al., 1998).

**Mediating Processes of the Fluency Effects**

Almost two decades of research on retrieval fluency effects has uncovered surprisingly little about the underlying mechanisms responsible for the retrieval fluency effects (for a similar point, see Caruso, 2008; Ruder & Bless, 2003). However, recently significant progress has been made (e.g., Tormala et al., 2007). Adding to the investigation of mediating processes, I investigated the easy-thus-liked and the easy-thus-typical heuristics (Studies 1 – 5), unrequested cognitions (Studies 2 & 5), perceived prior contact (Studies 2 & 5), exemplar vividness (Studies 1 – 5), cognitive fatigue (Study 3), and depth of cognitive processing (Study 3) as potential mediating variables.

None of the variables demonstrated mediating effects on retrieval fluency. I argued that the positioning of the mediator variable may have contributed to the weak effects. As the mediating measures always followed the measures of stereotypicality, it was not possible to know whether the mediating ratings reflected a process active at the time of forming the group judgments or whether they reflected a process initiated only after expressing the group judgments. To account for this problem, in the final study, typicality was included as a moderating variable. Typicality was further explored in Study 5 due to the strong role of exemplar typicality identified in previous member-to-group generalisation research (e.g., Bodenhausen, et al., 1995; Johnston & Hewstone, 1992). Typicality was manipulated by asking participants to retrieve or read about exemplars that were stereotypical, moderately
atypical or extremely atypical. In Study 5, it was found that the retrieval fluency effect was detected under retrieval of extremely atypical exemplars. That is, when retrieval was experienced as easy participants readily generalised extremely atypical exemplars to their outgroup judgments, where they did not when retrieval was difficult.

Importantly, throughout this thesis I proposed that typicality does not supersede other identified mediating processes. Rather, it adds to the list of previously identified variables that have explained retrieval fluency (e.g., Caruso, 2008; Tormala et al., 2002). Similar to Tormala et al. (2007), I do not suggest that the easy-thus-typical heuristic is an all-or-none affair. Instead, I suggest that the exact mediating processes that account for retrieval fluency effects are specific to the judgment domain at hand and to the retrieval task itself. As each of these varies, the influencing factors will vary accordingly. The support for the easy-thus-typical heuristic within a moderation model should operate as a springboard for future research that endeavours to more fully appreciate the diverse nature of retrieval fluency.

**Extending the Matching Hypothesis**

I found generalisation using the cognitive outcome measure of stereotypicality but not using an affective outcome measure of prejudice. In addition, there was no support for the easy-thus-liked heuristic across Studies 1 – 5. It is unlikely that these null findings can be attributed to poor measuring of the easy-thus-liked heuristic, as several measurement approaches were attempted. In Study 1 I used an out-group prejudice measure, in Studies 2 – 4 I used an out-group feeling thermometer, and in Study 5 I used an exemplar likeability index. It is most likely that the lack of generalisation on prejudice measures and the lack of mediation on the easy-thus-liked heuristic indicate that affective variables are unable to mediate a cognitively loaded activity.

The lack of change on the prejudice measure and the easy-thus-liked heuristic are consistent with the matching hypothesis. From this perspective, affective measures are less suitable to tap cognitive-based experiences (i.e., information retrieval; see also Mayer & Tormala, 2010). In Chapter 2, I introduced the matching hypothesis (Paolini et al., 2007) to account for the general advantage of using cognitive outcome measures to assess for member-to-group generalisation in the social cognitive literature (see also Haddock et al., 2008; Huskinson & Haddock, 2004; Pettigrew & Tropp, 2006). Paolini initially proposed the matching hypothesis to account for changes in contact research. However, my data suggests that the matching hypothesis may be extended to the self-generation paradigm. In general, the cognitive nature of retrieval tasks is best assessed through cognitive outcome measures.

It is worth noting that the matching hypothesis is unable to account for the lack of change on measures of group dispersion. Interestingly, I revealed in Chapter 2, that dispersion measures in impression formation studies have been used almost exclusively to investigate
generalisation in cognitively based out-groups (such as employment groups or student majors). In that chapter, I also noted that dispersion measures were not adopted in contact-based research, where the focus is on affective out-groups (see, Pettigrew & Tropp, 2000). Hence, it may be that dispersion is not a suitably matched measure to gauge changes in the representation of affective groups. To support this conclusion, further work should include measures of dispersion that assess generalisation with affective out-groups across contact research, impression formation research and self-generation research.

**Methodological and Social Implications**

*The Self-generation Paradigm*

A merit of this research is the introduction of a new paradigm. The self-generation paradigm draws on the strengths of both the social cognitive literature and the contact literature. Accessing an individual’s own experiences with the out-group, rather than presenting a list of unfamiliar exemplars, enhances the ecological validity of this approach to member-to-group generalisation. Requesting a specific number and type of exemplars is nonetheless more controlled and less susceptible to a variety of extraneous variables than surveying undefined prior history of face-to-face contact.

Methodologically, this research has extended our understanding of retrieval fluency and improved the few-versus-many paradigm in several ways. Through a manipulation of depth of processing and cognitive fatigue (Study 3), I ruled out two novel alternative explanations that may have accounted for retrieval fluency effects. Similarly, checks of exemplar vividness, unrequested cognitions and prior contact with the out-group did not influence the results. Most importantly, in the context of member-to-group generalisation the inclusion of the no-information control group has demonstrated that fluency effects involve inclusion of the retrieved material in the few-exemplars condition and reduced inclusion in the many-exemplars condition.

*Social Implications: Improving Social Policy*

The data presented in this thesis shows that self-generation is particularly effective at producing generalisation under conditions of high fluency (Wänke et al., 1996; for a similar point in attitude research, see Janis and King, 1954; King & Janis, 1956). However the use of retrieval as a means for social change must be carefully constructed to manage two main issues. Firstly, it is important that the retrieval of stereotype disconfirming information is experienced as easy. For social policy makers this means that prior to implementing self-generation based approaches to change, increasing the exposure of counterstereotypical exemplars to the population is paramount. Secondly, when retrieval is more difficult, self-
generation facilitates stereotype change only when people are made aware of the biasing influence of the meta-information. For social policy makers this means that educating people about the paradoxical effects of retrieval difficulty and the role of biasing influences on their judgments may have sizeable and long-reaching effects on the well-being of discriminated groups in society. Although educating people on the influence of meta-cognitive processes has received mixed results (for supportive evidence, see Allan & Johnson, 2009; Wegener & Petty, 1995; for unsupportive evidence, see Macrae et al., 1994). Successful education programs should aim to increase the cognitive processing of individuals resulting in less biased and more accurate social judgments.

Traditionally, social policy has been influenced by the findings from contact research and impression formation research. My findings add important considerations for both of these research traditions. In the area of contact research, the influence of retrieval fluency in member-to-group generalisation may explain the results of a number of correlational contact-based studies. For example, requesting participants to indicate the number of their out-group friends (e.g., Paolini et al., 2004a) prior to judging the out-group may also be subjected to the retrieval fluency effects detected in this work. Hence, inconsistencies in the contact research may perhaps be reconciled through a consideration of internal meta-information. Consistent with this argument, Pettigrew and Tropp (2006) revealed that intimate forms of contact (i.e., contact with out-group friends) over more generic forms of contact (i.e., contact with acquaintances) result in greater generalisation. This might simply reflect greater fluency of more personal, rather than less personal, contact experiences.

In addition, my research also adds a new variable of consideration for research on imagined intergroup contact (Stathi & Crisp, 2007; Turner, Crisp, & Lambert, 2007). These new ‘contact’ interventions are focusing on the role of imagined intergroup contact especially for situations when actual face-to-face contact is not always possible due to intergroup segregation. Research on imagined contact shows that asking people to imagine contact situations results in similar generalisation effects to face-to-face contact (e.g., Stathi & Crisp, 2007). My data would suggest that the ease or difficulty associated with completing the ‘imagination’ task would also influence the success of the generalisation. In light of my results, researchers engaging in imagination tasks would be well served to ensure that the task is high in fluency for maximised results.

In the area of impression formation research, it is worth considering that in most contexts where people make group relevant judgments, it is not immediately after being presented with exemplar information (as is the case with the impression formation paradigm). Rather, most people make group relevant judgments at a latter point, oftentimes many days after the exposure to the out-group exemplar/s (as is the case with the self-generation paradigm). Hence, what I found in the self-generation paradigm should inform social
interventions more practically than research relying strictly on impression formation paradigms. Since the general theme is a case of ‘less is more’, ensuring that positive disconfirming exemplars are highly accessible should encourage easy retrieval, with the end result being more positive out-group judgments.

**Limitations of the Present Work and Considerations for Future Work**

This research is limited by a few factors. One limitation is the focus on a single out-group. Although, I do not expect that fluency effects identified in this research will be limited to the stereotype of the inactive elderly. Rothman and Hardin (1997) have shown that retrieval fluency impacts on gender out-group judgments. Further research should address the role of fluency in stereotyping towards different out-groups.

Another limitation of this research is the reliance on the retrieval of multiple disconfirming exemplars. In Studies 1 – 3 experimental participants were asked to retrieve at least three exemplars and in Studies 4 and 5 experimental participants were asked to retrieve at least four exemplars. It would be pragmatically interesting to check if retrieval fluency effects hold in the context of retrieving a single counterstereotypical member. This seems especially relevant to contexts in which discriminated groups constitute a small numerical minority. For example, Indigenous Australians represent only 2.5% of the Australian population (Australian Bureau of statistics, 2009), yet face many negative out-group stereotypes (e.g., Alcohol abuses). Despite the small representation in the population most Australians should be able to draw on one counterstereotypical example of a non-drinking Indigenous Australian (e.g., Cathy Freeman, Evonne Goolagong-Cawley, Anthony Mundine or Charlie Perkins). Wänke et al. (1996, Study 6) provided evidence of retrieval fluency effects in one-exemplar research. In the research by Wänke et al. participants were asked to generate one reason (versus 10 reasons) for purchasing a BMW car instead of a Mercedes. Participants were then asked to indicate their preference for either a BMW or a Mercedes. It was found that participants in the one-reason condition were more likely to prefer a BMW than participants in the 10-reason condition. It would be interesting to see if one-exemplar retrieval effects hold for social out-groups, such as Indigenous Australians.

The majority of the present research focused on retrieval fluency in low-elaboration-judgment conditions (Study 1 and the cue-absent conditions in Studies 2 & 3). The manipulation of depth of processing in Study 3 introduced a high-elaboration-judgment condition. Although this manipulation had no significant influence on the results of this study, the results from the cue-present conditions in Studies 2 and 3 suggest that inducing high elaboration should remove the influence of retrieval fluency. Further support for the role of high-elaboration reducing fluency effects can be found in Rothman and Schwarz (1998). Rothman and Schwarz asked participants to generate three or eight risk factors for heart
disease. When heart disease was considered self-relevant (i.e., high-elaboration) participants displayed a content-based effect rather than a retrieval fluency effect. On the other hand, Kühnen (2010) still found a retrieval fluency effect under conditions of high-elaboration and a content-based effect under conditions of low-elaboration. Kühnen’s research, however, included interpolated manipulations between standard retrieval fluency manipulations and key judgments (e.g., an accuracy motivation in Study 1, a power manipulation in Study 4) that may have unwittingly induced high-elaboration and drive the reversal of retrieval fluency effects. Future research could investigate the role of high-elaboration to further understand the controlled aspect of my dual process model.

This research did not address the role of individual differences in the application and use of meta-information associated with fluency. Research has demonstrated that greater personal involvement in the topic of judgment promotes systematic processing of the content (e.g., Chaiken et al., 1989). In addition, individuals high in the need for cognition have been shown to engage in more thorough cognitive processing of material, and are less likely to be influenced by heuristic processing strategies (Cacioppo & Petty, 1982), and respond with greater receptivity to cognitive information (Haddock et al., 2008). Since the simple inclusion of a meta-cognitive cue was sufficient to engage controlled processing, it seems likely that individuals with a greater involvement in the topic or with a natural tendency to engage in controlled processing would be less influenced by retrieval fluency. Future research would be well placed to focus on the moderating role of these individual differences.

Other interesting individual difference variables that may impact on the extent of bias correction include the motivation to suppress prejudice (e.g., Dunton & Fazio, 1997) and the motivation to suppress stereotyping (Maass, Castelli & Arcuri, 2000). Maass et al. argued that people differ in their motivation to suppress socially undesirable thoughts, such as racist attitudes or stereotyped judgments. In their research, participants were asked to interact with two different confederates. One confederate was introduced as a healthy person while the other was introduced as a person with AIDS. After the interaction participants were asked to indicate their level of willingness to interact with the out-group member in the future. Maass et al. coded the body posture of the participants during the interactions and found that high stereotype suppressors showed a more tense body posture towards the ill confederate than towards the healthy confederate, reflecting implicit stereotyping at the non-verbal level. However, on the verbal measure assessing willingness for future interaction, high suppressors expressed more willingness to meet a person with AIDS than a healthy person. This paradoxical dissociation between the non-verbal and verbal indicators is likely to reflect an overcompensation effect along the verbal measure. High suppressors may try so much not to look prejudiced that their judgments are over-corrected (for over-correction towards gay men, see Aberson & Dora, 2003). Motivations to suppress stereotyping might explain a substantial
amount of variance in the degree of bias correction that my participants engaged in, with highly motivated participants more likely to produce a significant reversal (or over-correction).

The individual difference in suppression motivation might account for the difference in congruent cue conditions across Studies 2 – 4. In Studies 2 and 3, participants presented with a congruent cue displayed a null effect of retrieval fluency. In Study 4, participants presented with a congruent cue displayed a significant reversal of the retrieval fluency effect. Interestingly, participants in Study 4 were completing the research in a one-to-one laboratory testing environment, while participants in Studies 2 (anonymous web-collection) and 3 (mass class-testing) were in a less personal testing environment. This difference in testing environments may have stimulated a greater motivation to suppress the influence of fluency effects amongst participants in Study 4. Future research would be well positioned to measure different motivational influences and assess the impact of different testing environments on the manipulation of fluency and the meta-information task cue.

Extending member-to-group generalisation research to the area of fluency has demonstrated the broader value in understanding the impact of meta-cognition. However, this research is limited by the single focus on the meta-cognitive process of fluency. There are other meta-cognitive processes that may be explicitly investigated in member-to-group generalisation. For example, Leyens, Yzerbyt and Schadron (1992) studied the metacognitive process of judgeability. Their Social Judgeability Model suggests that people will only judge individuals when they feel that the information they have available is both sufficient and relevant to the impending decision, in other words the target is perceived as being judgeable. It seems very relevant to extend the Social Judgeability Model from interpersonal judgments to intergroup judgments. Another area of meta-cognitive research might investigate the influence of different feelings as a form of meta-information. For example, feelings of familiarity have been found to influence subsequent judgments. Consistently exposing people to the same propositional stimulus increases the sense of familiarity people have with the stimuli and in turn increases the likelihood that people accept the stimulus items as valid (Jacoby, Kelley, Brown & Jasechko, 1989; Kruglanski, Freuend & Bar-Tal, 1996; Zajonc, 1980), and positive (e.g., Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2010). It seems reasonable that consistently exposing people to disconfirming exemplars will increase the perceived validity in the exemplar information resulting in acceptance of disconfirming exemplars and a change in negative out-group stereotypes (cf., Banaji & Greenwald, 1995). Continuing the investigation of meta-cognitive influences on the process of member-to-group generalisation will further inform and enhance our understanding of this type of social inductive reasoning.
Concluding Remarks

I opened this thesis reporting the comments made by Senator Peter Breen to NSW state parliament in 2005 about Asian drivers being the worst in the world. I wondered how many good Asian drivers Mr Breen would need to come across before he changed his stereotype. The answer to this question is neither easy, nor is it straightforward. But my work suggests that it would certainly have something to do with the ease with which Mr Breen retrieved the exemplars at the time of constructing his parliamentary speech. Perhaps Mr Breen drove to work that morning and was cut-off by an Asian driver or was delayed by a slow Asian driver struggling to complete a reverse park. These examples would have come to mind with high fluency, and overly influenced his statement to parliament. What if that morning he drove to work and, instead, saw a number of good Asian drivers carefully and precisely navigating their way through the city traffic? It would seem that these easily retrieved examples would have generalised to Mr Breen’s stereotype, and resulted in a less stereotypical address (and maybe even in a day that didn’t end with a massive public backlash from his politically incorrect statement!). The answer to my initial question would also have something to do with the typicality of the exemplar. It seems that extremely atypical exemplars are more likely to trigger fluency effects.

Stereotypes are not a thing of the past. They are alive and well in modern society. In an effort to improve negative intergroup relations, this thesis has identified the self-generation paradigm as a new method of conducting research on member-to-group generalisation. Through the use of this paradigm, I have found that retrieval fluency effects support a case of ‘less is more’ in the area of member-to-group generalisation. Used alone, or in concert with social-cognitive or contact strategies, this meta-cognitive approach may enable us to improve negative out-group judgments and troubled intergroup relations. Hopefully, from this research and follow-up investigations, researchers can work towards developing new ways to reduce the impact of negative out-group stereotypes.
REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.


References


the contact hypothesis among majority and minority groups in three European


References


References


Hewstone, M., & Brown, R. (1986). Contact is not enough: An intergroup perspective on the ‘contact hypothesis’. In M. Hewstone, & R. Brown (Eds.), *Contact and conflict in intergroup encounters* (pp. 1-44). Oxford: Blackwell.


References


References


## Appendix I: Concentrated-Dispersed Meta-Analysis

### Coding of Tests Meeting Inclusion Criteria for Concentrated-Dispersed Meta-Analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study</th>
<th>g</th>
<th>r</th>
<th>N</th>
<th>Size</th>
<th>Disc Size</th>
<th>DV</th>
<th>Typical</th>
<th>Eval</th>
<th>Gval</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hantzi (1995)</td>
<td>1</td>
<td>.65</td>
<td>.31</td>
<td>80</td>
<td>6</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone et al. (2000)</td>
<td>1</td>
<td>.55</td>
<td>.26</td>
<td>76</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone et al. (1992b)</td>
<td>1</td>
<td>-.66</td>
<td>-.31</td>
<td>40</td>
<td>6</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone et al. (1992b)</td>
<td>1</td>
<td>.39</td>
<td>.19</td>
<td>40</td>
<td>6</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone &amp; Hamberger (2000)</td>
<td>1</td>
<td>.52</td>
<td>.25</td>
<td>68</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone &amp; Hamberger (2000)</td>
<td>2</td>
<td>.97</td>
<td>.43</td>
<td>64</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone et al. (1994)</td>
<td>1</td>
<td>2.49</td>
<td>.78</td>
<td>24</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone et al. (1994)</td>
<td>2</td>
<td>.92</td>
<td>.42</td>
<td>24</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Hewstone et al. (1994)</td>
<td>3</td>
<td>.90</td>
<td>.41</td>
<td>24</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Johnston &amp; Hewstone (1992)</td>
<td>1</td>
<td>.83</td>
<td>.38</td>
<td>78</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Johnston &amp; Hewstone (1992)</td>
<td>2</td>
<td>.87</td>
<td>.40</td>
<td>42</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Johnston et al. (1994)</td>
<td>1</td>
<td>.51</td>
<td>.25</td>
<td>118</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Johnston et al. (1994)</td>
<td>2</td>
<td>.97</td>
<td>.43</td>
<td>126</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Johnston et al. (1994)</td>
<td>3</td>
<td>1.16</td>
<td>.50</td>
<td>60</td>
<td>8</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Moreno &amp; Bodenhausen (1999)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>48</td>
<td>6</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Queller &amp; Smith (2002)</td>
<td>3</td>
<td>.46</td>
<td>.23</td>
<td>50</td>
<td>18</td>
<td>18</td>
<td>S</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Queller &amp; Smith (2002)</td>
<td>3</td>
<td>-.49</td>
<td>-.24</td>
<td>50</td>
<td>30</td>
<td>18</td>
<td>S</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Weber &amp; Crocker (1983)</td>
<td>1</td>
<td>1.07</td>
<td>.47</td>
<td>40</td>
<td>6</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Weber &amp; Crocker (1983)</td>
<td>1</td>
<td>2.24</td>
<td>.75</td>
<td>40</td>
<td>30</td>
<td>30</td>
<td>S</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>Weber &amp; Crocker (1983)</td>
<td>4</td>
<td>.87</td>
<td>.40</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>S</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Weber &amp; Crocker (1983)</td>
<td>4</td>
<td>-.60</td>
<td>-.29</td>
<td>20</td>
<td>18</td>
<td>18</td>
<td>S</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>C</td>
</tr>
</tbody>
</table>

**Notes.** g = Hedges’ g effect size for each sample; N = Number of participants; DV = outcome measure (S = Stereotypicality; D = Dispersion; P = Prejudice); Size = total out-group sample size; Dis Size = Disconfirming sample size; Typical = Typicality of exemplars (1 = typical; 2 = moderately atypical; 3 = extremely atypical; 4 = mixed typicality; 5 = other/unspecified); Eval = Valence of the exemplar information (1 = positive; 2 = negative; 3 = neutral; 4 = mixed; 5 = unspecified); Gval = Valence of the group stereotype (1 = positive stereotype; 2 = negative stereotype; 3 = mixed); Type = Out-group type (A = affective; C = cognitive).
Appendix II: Pictures Included in the IAT
Appendix III: Experimental Profiles Used in Study 2

Name: Frank Wilkinson
Approximate Age: 78
Brief description of what makes this person active: Frank is a very energetic man. He walks to the store each morning to buy the paper. After reading through the news he then tries to complete the crosswords in the back section of the paper. After lunch he walks to the local primary school where he volunteers to help students with reading difficulties. He makes sure that he gets about 30 minutes of exercise by walking every day.

Name: Beryl McCauley
Approximate Age: 82
Brief description of what makes this person active: Beryl is a very busy woman. As she never learnt to drive a car, she walks into town to get her groceries every two or three days. This is about a two-kilometre walk each way. Beryl is very involved in her local church, organising different fund raising activities and teaching Sunday school. She also loves to play scrabble and enjoys competing in tournaments where she performs often makes it through to the final rounds.

Name: John Howard
Approximate Age: 68
Brief description of what makes this person active: John Howard was the Prime Minister of Australia until 2007. While not everyone agreed with his policies, he would certainly be described as very dynamic. As Prime Minister, he had an extremely intellectual job and was required to constantly deal with new ideas, problems and issues. Even now, he remains one of the longest serving members in the Australian Liberal Party and can be seen walking every day to remain physically fit.

Name: Winn Jenkins
Approximate Age: 96
Brief description of what makes this person active: At 96 Winn lives independently and maintains a busy schedule of appointments and meetings. She attends regular ProBus meetings (short for professional business people). At these meetings she remains informed about different government policies and taxation changes. A few years ago she attended a computer course and learnt to use the Internet, and now she does stock trading online. She also loves to play bowls and plays social games every week.
Name: Lloyd Hudson
Approximate Age: 70
Brief description of what makes this person active: Lloyd is recently retired. When he stopped working he decided to take up triathlons to help him maintain a healthy lifestyle. Instead of just competing in short distance events he competes in the Ironman events (these are the really long races that take all day to complete). He usually wins his age division and last year was selected to participate in the World Championships in his age division. He is really an amazing person.

Name: Noel Frankel
Approximate Age: 74
Brief description of what makes this person active: Noel owns and runs the local news agency. He works very long hours, opening his doors at 6 am and then closing at 6 pm. You can see Noel in the store 6 days a week. He says he takes Sundays off to balance the books and organize payments to staff and suppliers. The business is very successful and he makes sure he learns the names of all his regular customers.

Name: Queen Elizabeth II
Age: 81
Brief description of what makes this person active: Queen Elizabeth is the queen of the United Kingdom and rules as queen over many states and their territories. She leads a very hectic life, full of engagements and attends possibly hundreds of official duties every year. She frequently travels overseas, for example, she officially visited to the United States last year. She is mentally alert in that she has an excellent knowledge of geopolitics and is still very aware of current world issues.

Name: Sean Connery
Age: 77
Brief description of what makes this person active: Sean Connery is a Scottish actor and producer. At 70 years he released one of his best films, Finding Forrester. More recently he stared in “The League of Extraordinary Gentlemen”. His role in films requires a high level of fitness and demands great mental capacity. He is also very involved in charity work and created his own Scottish charity to support deprived children in Edinburgh as well as the Scottish film industry.

Name: Dorothy Hyde
Age: 71
**Brief description of what makes this person active:** Dorothy is heavily involved in local charity work. She volunteers at the Salvation Army and works two or three days a week in one of their second hand shops. She tries to keep herself fit by swimming regularly. Although she says that she doesn’t have much time because of her other commitments, like minding her grandchildren, she still swims over a kilometre at the local pool every second day.
Appendix IV:

Study 2 Mediation Analysis of the Easy-Thus-Liked Heuristic

The Restricted Design

Using the restricted design, I followed the three preliminary tests used in Study One to identify promising mediating variables. The first step compares the results on the feeling thermometer and stereotype change along the number of exemplars (three vs. nine) factor for self-generation/meta-information cue-absent participants. On the stereotype change measure number of exemplars produced a marginally significant effect ($F(1, 54) = 3.16, p = .08$), but a non-significant effect on feeling thermometer, $F < 1$. In the second step, when the feeling thermometer was included as a covariate, the effect of number of exemplars remained similar from, $F(1, 54) = 3.16, p = .08$, to $F(1, 53) = 3.12, p = .08$. Participants in the three-exemplar condition changed their stereotype more ($M = .39, SD = .70$) than participants in the nine-exemplar condition ($M = .07, SD = .65$). There was no correlation, Pearson $r = .08, p = ns$. The lack of evidence across these three preliminary steps suggests that there is no support for the easy-thus-liked heuristic and precludes the need for any further mediation analysis.

Analysis of all Conditions

I adopted the standard three-step strategy from Study One for preliminary testing of mediating variables. The first step compared the results between the feeling thermometer and stereotype change on the three-way ANOVA (information source, number of exemplars, meta-information cue). The results on the stereotype change measure were significant, however, there was no significant three-way interaction using the feeling thermometer, $F < 1$. In the second step, the feeling thermometer was entered as a covariate on the significant three-way interaction (information source, number of exemplars, meta-information cue) with change in group stereotyping as the dependent variable. The results of this analysis remained significant, from $F(1, 223) = 4.72, p = .03$ to $F(1, 222) = 5.31, p = .02$. The final step tested for correlation between the two variables: change in group stereotyping and the feeling thermometer. There was a small significant positive correlation between change in group stereotyping and the feeling thermometer, Pearson $r = .13, p = .05$ (2-tailed). This analysis indicates that increased feelings of warmth towards the elderly are positively related to greater stereotype change. The lack of significant effects on the ANOVA and ANCOVA analyses fail to support the easy-thus-liked hypothesis.
Appendix V:

Study 2 Mediation Analysis of the Easy-Thus-Typical Heuristic

The Restricted Design

On the stereotype change measure number of exemplars produced a marginally significant effect \( F(1, 54) = 3.16, p = .08 \), but a non-significant effect on the typicality index, \( F < 1 \). Next, I included the typicality index as a covariate in the one-way ANOVA along change in group stereotyping. The effect of exemplar number remained substantially unchanged, from \( F(1, 54) = 3.16, p = .08 \) to \( F(1, 53) = 2.93, p = .09 \). The third step assessed the relationship between the critical variables. There was a non-significant correlation between typicality and stereotype change, Pearson \( r = .08, p = ns \).

The above three steps (ANOVA, ANCOVA and correlation analysis) were repeated using the frequency index. There was a non-significant effect of number of exemplars on the frequency index, \( F(1, 54) = 1.37, p = .25 \). Next, I included the frequency index as a covariate in the one-way ANOVA along change in group stereotyping. There was no statistically observable role of the frequency index from \( F(1, 54) = 3.16, p = .08 \) to \( F(1, 53) = 3.15, p = .08 \). There was no correlation between the frequency index and stereotype change, \( p = ns \).

The lack of significant effects on the ANOVAs and the lack of change on the ANCOVAs and the lack of correlation between the predictor variables and the outcome variable suggest there is no need to continue with the full model assessment of mediation for the easy-thus-typical heuristic (Baron & Kenny, 1986).

Analysis of all Conditions

In Study 2, two measures of typicality assessed the easy-thus-typical heuristic: (1) the typicality index, and (2) the frequency index.

A three-way interaction on the typicality index using the 2 (information source: self-generation vs. presented) x 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: absent vs. present) between subjects ANOVA was marginally significant, \( F(1, 223) = 3.46, p = .06 \). This three-way interaction for the typicality index is slightly weaker compared to the ANOVA using the change in group stereotyping index, \( F(1, 223) = 4.72, p = .03 \). Following up on the marginally significant interaction using exemplar typicality revealed no significant two-way interactions (all \( Fs \leq 1.64, ps \geq .20 \)). The only significant main effect was for information source, where the self-generated participants perceived the exemplars to be more typical of the elderly (\( M = 3.10, SD = .80 \)) than the participants in the presented information condition (\( M = 2.64, SD = .75 \)), \( F(1, 223) = 18.69, p < .001 \).

In the second step, I used the typicality index as a covariate on the three-way ANOVA (information source, number of exemplars and meta-information cue) on change in group stereotyping. The three-way interaction remained significant, \( F(1, 222) = 3.87, p = .05 \),
from $F(1, 223) = 4.72, p = .03$. The third step assesses the relationship between the critical variables. There was a weak correlation between typicality and stereotype change, Pearson $r = .10, p = .07$ (one-tailed).

The above three steps (ANOVA, ANCOVA and correlation analysis) were repeated using the frequency index. There was no three-way interaction, $F < 1$, using the frequency index as the dependent variable on the ANOVA with number of exemplars, information source and meta-information cue. Next, the frequency index was entered as a covariate on the three-way ANOVA on change in group stereotyping. There was no statistically observable role from $F(1, 223) = 4.72, p = .03$ to $F(1, 222) = 5.20, p = .02$. There was a significant correlation between the frequency index and stereotype change, $r = .22, p = .001$ (two-tailed).

Exemplar typicality and frequency are positively related to greater stereotype change, suggesting that the more typical the counterstereotypical exemplar the larger the stereotype change. The lack of fully significant effects on the ANOVAs and the lack of change on the ANCOVA indicate that no assessment of the full model of mediation required (Baron & Kenny, 1986).
Appendix VI:

Study 3 Mediation Analysis of the Easy-Thus-Liked Heuristic

Analysis of all Conditions

In the first step to assess the mediating role of the easy-though-liked heuristic, I entered the feeling thermometer as the dependent variable on the 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) between subjects ANOVA for participants in the active elderly condition. The three-way interaction on the feeling thermometer was non-significant, $F < 1$, and did not mirror the results detected on change in group stereotyping. A main effect of meta-information cue emerged, $F (1, 57) = 13.28, p = .001$. Participants without a meta-information cue were higher in warmth ($M = 87.53, SD = 7.81$) than participants provided with a meta-information cue ($M = 76.68, SD = 14.36$). Altogether, this suggests that ‘liking’ is not operating in a way consistent with the results of the ANOVA using change in group stereotyping.

In the second preliminary step, I used the feeling thermometer as a covariate, and ran a 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) ANCOVA on the group stereotyping index. There was no sizeable change in significance for the two-way interaction from, $F (1, 57) = 5.55, p = .02$, to $F (1, 56) = 5.09, p = .03$. Finally, the feeling thermometer did not co-vary with the group stereotyping index, $r = -.12, ns$. This confirms that the pattern of group stereotyping produced by the joint action of number of exemplars and meta-information cue was not explained by variation in likeability.
Appendix VII:

Study 3 Mediation Analysis of the Easy-Thus-Typical Heuristic

Analysis of all Conditions

To first assess the mediating role of typicality, I entered the typicality index as the dependent variable on the 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) between subjects ANOVA. The index for individual typicality ($F(1, 57) = 3.42, p = .07$), closely mirrored the findings for group stereotyping ($F(1, 57) = 5.55, p = .02$). Following up on the marginal interaction on exemplar typicality, participants in the meta-information cue: absent condition supported a retrieval fluency effect. These participants perceived the retrieved exemplars as more typical after retrieving three-exemplars ($M = 3.86, SD = 1.09$) than after retrieving nine-exemplars ($M = 3.33, SD = 1.08$); this difference, however, was statistically non-significant, $F(1, 28) = 1.75, p = .20$. This indicates that as the participants progressed with the task they needed to generate slightly more extreme cases. This finding was reversed for the participants in the meta-information cue: present condition. Here, participants in the nine-exemplar condition perceived greater typicality ($M = 3.98, SD = 1.05$) than participants in the three-exemplar condition ($M = 3.45, SD = 1.21$). This difference was non-significant, $F(1, 29) = 1.68, p = .21$. Overall, these findings indicate that the experimental manipulations affected the exemplar typicality of active elderly exemplars in a way consistent with a retrieval fluency account.

The second preliminary step assessed the role of typicality as a covariate on the 2 (number of exemplars: three vs. nine) x 2 (meta-information cue: present vs. absent) ANOVA using change in group stereotyping. There was a reduction in the size of the effect from, $F(1, 57) = 5.55, p = .02$, to a marginal finding, $F(1, 56) = 3.66, p = .06$. Thus, confirming that the joint action of number of exemplars and meta-information cue was partly explained by variations in exemplar typicality. The main effects were non-significant, all $F$s < 1.

The third preliminary test looked at correlation analyses. Exemplar typicality entertained a meaningful co-variation with the change in group stereotyping index, $r = .22, p = .02$. This suggests that as typicality increases, change in group stereotyping also increases.

Following this three-step preliminary analysis, I carried out a simple path analysis consisting of three regression equations to test a mediational model (Baron & Kenny, 1986, p. 1177): regressing the mediator on the predictor; regressing the criterion on the predictor; regressing the criterion on both the predictor and mediator. Mediation holds if the effect of the predictor on the criterion is reduced whilst controlling for the mediator and all the other beta-weights are significant. Full mediation holds if the predictor variable is no longer significant when the mediator is controlled.

First, a contrast vector was created to represent the special status of the meta-information cue: absent condition using the weights in parentheses: meta-information cue
absent/three-exemplar (2); meta-information cue absent/nine-exemplar (-2); meta-information cue present/three-exemplar (1); meta-information cue present/nine-exemplar (-1). Figure A.1 shows the results of the analyses run with the contrast vector as predictor, the exemplar typicality as mediator, and change in group stereotyping as the criterion.

**Figure A.1.** Mediation model for change in group stereotyping. Cue = meta-information cue. The path coefficients are standardized beta weights; *p < .05.

The contrast vector did not predict a significant change in exemplar typicality (β = .02, p = ns) or stereotype change (β = .14, p = .28). Exemplar typicality predicted a significant increase in change in group stereotyping when controlling for the vector (β = .28, p = .03). When the mediator was entered into the equation, the effect of the contrast vector was slightly reduced and still non-significant (β = .12, p = .33). Sobel’s test (Baron & Kenny, 1986, p. 1177) was used to test the significance of the indirect effect of the predictor on change in group stereotyping via exemplar typicality. This indirect effect was non-significant, Z < 1. Therefore, the typicality of the retrieved exemplars mediated an increase in change in group stereotyping, but this mediation analysis was not fully significant.
Appendix VIII
Experimental Profiles Used in Study 5

Stereotypical Exemplars

Name: Noel Frankel
Approximate Age: 74

Brief description of what makes this person inactive: Noel retired a few years ago and since then he has gradually lost interest in things. He used to play a lot of golf, but now complains that it plays trouble with his hip. He frequently tells you about all his aches and pains, saying that he doesn’t really enjoy walking places and would prefer to drive. He seems happy sitting around during the day and watching the television.

Name: Frank Wilkinson
Approximate Age: 78

Brief description of what makes this person inactive: Frank uses an electric chair to go to the shops, although there is nothing wrong with him, he prefers to use the chair because he says it is easier than walking around the aisles. Each week he buys alcohol and cigarettes, although sometimes he should be spending his money on other items like health care bills. He frequently forgets his grandchildren’s birthdays often sending them cards that arrive weeks late.

Name: Elizabeth Taylor
Age: 77

Brief description of what makes this person inactive: Although Elizabeth Taylor used to lead a very full life in her younger days, nowadays she has become quite sedentary. She is frequently in hospital and is pictured using a wheelchair to get around. A few years ago she had a brain operation and now has other people take care of her day-to-day accounts. She could be considered reclusive and sometimes fails to make scheduled appearances citing illness or personal reasons.

Name: Dorothy Hyde
Age: 69

Brief description of what makes this person inactive: Dorothy spends a lot of her time at the local RSL club where she plays bingo twice a week. She usually spends a lot of time on pension day playing the pokies and then spends her other week days doing knitting. She doesn’t really read and finds the news and other current affairs programs depressing. Sometimes she will flick through the local newspaper to find her horoscope, which she enjoys to read.
Moderately Atypical Exemplars

Name: Frank Wilkinson
Approximate Age: 78

Brief description of what makes this person active: Frank is an active man. He walks to the store each morning to buy the paper. After reading through the news he then tries to complete the crosswords in the back section of the paper. After lunch he walks to the local primary school where he volunteers to help students with reading difficulties. He makes sure that he gets about 30 minutes of exercise by walking every day.

Name: Beryl McCauley
Approximate Age: 82

Brief description of what makes this person active: Beryl is a busy woman. As she never learnt to drive a car, she walks into town to get her groceries every two or three days. This is about a one-kilometre walk each way. Beryl is very involved in her local community, helping to organise different fund raising activities for the Lions Club. She also volunteers at the community soup kitchen once a week to cook and serve food for the homeless.

Name: Sean Connery
Age: 77

Brief description of what makes this person active: Sean Connery is a Scottish actor who retired from acting a few years ago. Since then he has become involved in charity work and created his own Scottish charity to support deprived children in Edinburgh as well as supporting the Scottish film industry through his guest appearances. Although he is not as involved in the media as he once was, he is still frequently photographed attending dinner ceremonies and award evenings.

Name: Dorothy Hyde
Age: 69

Brief description of what makes this person active: Dorothy enjoys regular exercise and tries to keep herself fit by swimming regularly. Although she says that she doesn’t have much time because of her other commitments, like minding her three grandchildren each afternoon, she still swims over a kilometre at the local pool every second day. She also loves to play scrabble and enjoys competing in tournaments where she performs often makes it through to the final rounds.

Extremely Atypical Exemplars

Name: Lloyd Hudson
Approximate Age: 70

Brief description of what makes this person extremely active: Lloyd is recently retired. When he stopped working he decided to take up triathlons to help him maintain a healthy lifestyle. Instead of just competing in short distance events he competes in the Ironman events (these are the really long races that take all day to complete). He usually wins his age division and last year was selected to participate in the World Championships in his age division. He is really an amazing person.

Name: John Munroe

Approximate Age: 68

Brief description of what makes this person extremely active: John Munroe is a heart surgeon, who frequently works over 12 hours a day to ensure that his patients receive the best possible care. He keeps his knowledge on surgical techniques current by attending regular conferences in many different parts of the world. As a heart surgeon, he understands the importance of remaining fit and healthy and, despite his hectic work schedule, jogs every morning and each year completes the Sydney Marathon.

Name: Queen Elizabeth II

Age: 81

Brief description of what makes this person extremely active: Queen Elizabeth is the queen of the United Kingdom and rules as queen over many states and their territories. She leads a very hectic life, full of engagements and attends possibly hundreds of official duties every year. She frequently travels overseas, for example, she officially visited the United States last year. She is mentally alert in that she has an excellent knowledge of geopolitics and is still very aware of current world issues.

Name: Winn Jenkins

Approximate Age: 96

Brief description of what makes this person extremely active: At 96 Winn lives independently and maintains a busy schedule of appointments. She attends regular ProBus meetings (short for professional business people) where she remains informed on government policies and taxation changes. A few years ago she attended a computer course and learnt to use the Internet, and now she does stock trading online. She also loves to play bowls, and still competes in women’s tournaments on a regular basis.