When do Autistic Children Exhibit a Shape Bias? Investigating the Impact of Methodology on Novel Noun Generalisation in Autism and Typical Development

Leigh Keating

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Psychology

Lancaster University

November 2022

1 Contents

Declaration

I declare that this thesis is entirely my own work completed under the supervision of Calum Hartley and Katherine Twomey. None of the work in this thesis has been submitted elsewhere in support of application for another degree at this or any other institution.

The parts of this thesis that are presented in publishable paper format are indicated at the beginning of the relevant chapter along with author contributions for each.

Leigh Keating (2022)

List of Tables

Introduction

Chapter 2

Chapter 3

Chapter 4

Appendix: A

List of Figures

Chapter 1

Chapter 2

Chapter 3

Acknowledgements

There are so many people in my life that have made completing this thesis possible. Firstly, of course, I am deeply grateful to my supervisors, Calum Hartley and Katherine Twomey. Their encouragement and support throughout has not only helped me to develop as a researcher, but has also made the whole process of completing this project such a positive experience. I have been extremely lucky to have such a wonderful team that had my back each step of the way.

I would also like to thank the friends and colleagues I have made during the PhD programme. Thanks to Samantha Russell and Sarah Chadwick for making C3a such a supportive place, even when it went virtual. I could always count on them for a sympathetic ear or some good humour. I would like to thank Nina Harrison, for being my ‘code buddy’, and travel companion on many a training course. I am also grateful to Rachael Cheung and Charlotte Rothwell, who have been on much of this journey alongside me and helped keep me positive in the face of the challenges of doing a PhD during a global pandemic.

I am grateful to all of my family, as everyone one of them has been supportive of my goals. In particular, my parents, Steph and Linda Sims, for always encouraging me, and for doing babysitting duty. Thanks to my nan, Margaret Sims, whose love and support is always a source of inspiration. And to my children: Benjamin, Elliott and Asher, for trying their best to get along while mummy was writing.

I am extremely grateful to the Leverhulme Trust, whose funding made this work possible. And of course, to the schools and children who took part in the research. It was so lovely to work with them all.

Finally, I owe unending thanks to my husband, Mark Keating. He has been my rock throughout my whole journey - from my undergraduate degree as a distance learner, through encouraging me to apply for a Masters programme, and finally to the PhD. In times where I doubted myself, he never did. I am thankful for all the wonderful people in my life who made this thesis what it is, though I suspect without Mark's support, the pages that follow would remain blank.

Thesis abstract

From ~24 months old, typically developing (TD) children often generalise names for solid objects to novel examples that are the same shape. This ‘shape bias’ can facilitate word learning by providing an attentional short-cut, allowing children to accurately generalise novel nouns. Autistic children often experience delays in language acquisition, and difficulty exploiting the shape bias may be a contributing factor. However, extant research findings are highly inconsistent, with autistic children exhibiting a shape bias in certain tasks but not others. Thus, the current research investigated whether variability in autistic children’s shape bias can be explained by methodological differences.

Autistic children (aged 4 to 9 years) and TD children matched on receptive vocabulary (aged 30 months to 4 years) participated in five experiments measuring shape bias in both ‘forced-choice’ and ‘yes or no’ variants of a novel noun generalisation task (NNG). Each task included an ‘online’ condition, where children were asked to generalise a label from a visible standard item to a target, and an ‘offline’ condition where the standard was absent at test, and generalisation had to be completed from memory.

Experiment 1 investigated whether the visibility of the standard affected autistic children’s shape bias in a forced-choice task with high contrast stimuli. We found that both autistic and TD groups generalised by shape regardless of standard visibility. Experiment 2 investigated whether shape was still preferred in a yes/no task, where children had the freedom to include any of the stimuli in the category rather than just the best example. In this task, autistic children were more likely to accept the differently shaped distractors than the TD group. Experiments 3 and 4 used the same tasks with low contrast stimuli to investigate whether the requirement to remember the standard had an impact when object shapes were more similar.

Both groups again exhibited a shape bias in the forced-choice task, however in the yes/no version only the TD group generalised by shape. Finally, Experiment 5 investigated whether an attentional preference for small details could account for shape bias differences in autism in a yes/no task. We found that both autistic and TD children generalised based on global shape rather than a salient local feature.

Overall, our results suggest that methodological variations can explain discrepancies in previous findings regarding shape bias in autism. Autistic children exhibited a strong shape bias in forced-choice tasks, whereas the bias appeared reduced or absent in yes/no tasks requiring children to categorise items individually. This suggests that differences may lie in autistic children's use of the shape bias as a tool for category exclusion decisions, rather than inclusion decisions, raising questions about the role of the shape bias in word and category learning for all children. There may be multiple routes through which attention to shape can contribute to learning and, by identifying which routes are most accessible for autistic children, we can inform teaching methods that work in harmony with their strengths.

Thesis introduction

Background and context of research

Organising the world into categories appears to come naturally in typical development. From as early as three months old, infants can use perceptual regularities between examples to form their first categories (Quinn et al., 1993). During the second year of life, powerful strategies start to develop that may make categorisation and word learning more efficient. By 36 months, typically developing (TD) toddlers tend to generalise labels to new category members based on shape: a behaviour known as the *shape bias* (Landau et al., 1988). The emergence of a shape bias appears to coincide with children having between 50 and 150 count nouns in their productive vocabulary (Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999). Analysis of children's early vocabularies suggest that they are dominated by nouns referring to solid objects belonging to shape-based categories (Perry & Samuelson, 2011; Samuelson & Smith, 1999), so children may learn early on that attending to shape is a useful strategy for generalising learned words to new exemplars. It has been suggested that through this early word learning experience children learn to make 'higher order' generalisations (Smith et al., 2002), allowing them to use this shape bias to generalise newly learned words from a single example, supporting faster, less effortful language acquisition.

However, autism is characterised by differences in visual attention (Happé & Frith, 2006) that may affect the development of shape bias, even when children's vocabularies have surpassed the 150 noun threshold (Tek et al., 2008). Although autistic children can generalise based on shape, they also frequently generalise by other perceptual features, such as colour (Hartley & Allen, 2014). Given that the shape bias is associated with faster vocabulary growth in typical development (Gershkoff-Stowe & Smith, 2004), it is useful to understand how and why autistic children may have difficulty accessing it.

However, the exploration of the role of shape bias in autism so far may have been hindered by disagreements regarding the nature of the shape bias, and differences in approaches to testing it experimentally. Broadly, 'shape bias' as a term describes a strategy for generalising novel count nouns on the basis of shape similarity, rather than other perceptual features (e.g. colour, size or material). However, the shape bias has been the source of debate on several fronts. For instance, is a shape bias *for* word learning, or is it a general category learning tool? While classic accounts of shape bias describe how children can use shape similarity to generalise labels (e.g. Landau et al., 1988), critics of this approach claim that this neglects the importance of conceptual knowledge and cannot explain how children learn what words mean (e.g. Bloom, 2000). However, as the following literature review will demonstrate, the wealth of research into the shape bias in typical development does converge on a coherent explanation that has room for both viewpoints. Evidence suggests that 'the shape bias' is not a rigid rule to be applied, but is a flexible tool that children can use as appropriate. It is born out of their past word learning experiences and informed by the structure of their own vocabulary. Furthermore, it is just one of multiple strategies that can be primed as part of an efficient and powerful attentional learning system that is observed when children are engaged in learning what words can refer to and what they mean.

The power of a shape bias as a word learning tool arguably comes from its flexibility. Children can generalise by shape when the situation calls for it, however they can also use alternative perceptual biases if it is appropriate to do such, such as texture for animate creatures (Jones et al., 1991) or deformable substances (Samuelson & Smith, 2000a). So, in addition to developing a shape bias, children must also learn when it is appropriate to use. Because of this, a shape bias can be sensitive to variability in task demands. Furthermore, ‘shape bias’ is not a discrete entity that a child does or does not possess, but merely a behavioural outcome that relies on multiple underlying processes including perception, attention, object recognition, working memory, language processing, communication, and social interaction (see Kucker et al., 2019) Whether or not a shape bias will be exhibited in a given task may be influenced by the demands on these processes in one way or another. Despite our rich understanding of task demands on shape bias in typical development, this same consideration is lacking in research in to shape bias in autism.

Since shape bias can be a powerful word learning tool, there is an obvious appeal to investigating the bias in autism, which can often be characterised by language delays and communication difficulties (Anderson et al., 2007). Furthermore, as we continue to try and unravel the role of attention and social processes in shape bias in typical development, understanding where autistic children may find this challenging can provide some insights into the underlying processes. Early attempts to investigate shape bias in autism suggested that autistic children do not have a shape bias even when they have the underlying language structure that predicts they should do (Tek et al., 2008). However, findings from different researchers have provided inconsistent results. While several studies claim that the use of shape for generalisation may differ in autism, the nature of these differences is not always the same. In this thesis I set out to investigate whether methodological differences in shape bias research conducted with autistic children can explain variability in findings.

Unity of the approach

The overall aim of this thesis was to explore the impact of different task demands on shape bias in autism. Having a shape bias is not a binary state, but a tool that can be deployed when the situation seems to call for it. As it relies on multiple processes (Kucker et al., 2019), any aspect of the experimental method that places a demand on a process that is more difficult for autistic children could inadvertently be detecting the difference in process and not a general shape bias deficit. For instance, tasks that require children to generalise from memory rather than having an example object visible could be detecting differences in memory for shape, rather than a weaker shape bias. Due to the wide variety of methodological differences that can be found in the last 30 years of shape bias research, the current research considers a select number that are pertinent to the current state of autism research in the field. In the current thesis, two variables of interest were selected as a common thread: whether the exemplar of the novel category (referred to as the 'standard') was labelled or not, and whether it was visible for reference while test items were being generalised.

The specificity of the shape bias as a lexical effect has been a long-time source of discussion (Booth & Waxman, 2008). Different researchers have interpreted and applied the term inconsistently, and this inconsistency carries through to the autism literature. For instance, some research with autistic participants uses the difference between labelled and unlabelled trials to measure shape bias (Potrzeba et al., 2015; Tek et al., 2008). Others define shape bias as a 'higher than chance' preference for shape which may be enhanced by labels (Field et al., 2016b). While the interpretation of the importance of labels to shape bias appears to vary, what is consistent is that autistic children perform differently to TD children when items are labelled. In this research we explored the extent of these differences by including both label and no-label trials for each version of the experiment.

The second variable of interest is the visibility of the standard. This was inspired by the findings of Tek et al. (2008), who used both a pointing task with a visible standard, and a screen-based looking time task (called the 'intermodel preferential looking' paradigm) during which the standard was not visible. The additional working memory demands of the IPL task inspired our initial research question, which is whether a requirement to remember the standard results in a weaker shape bias in autism. To address this, each experiment included an online condition, where an image of the standard remained visible for direct comparison, and an offline condition, where generalisation had to be performed from memory.

The final, and perhaps the most significant theme within this thesis is the differing demands of forced-choice and yes/no versions of novel noun generalisation (NNG) tasks. While previous work acknowledges that these tasks differ in the goal for the child – i.e. to find 'the best' example of the category or to judge if each item is a suitable example or not - the shape bias is robust in both of these tasks in neurotypical children (Samuelson et al., 2009). Previous research with autistic children has demonstrated a shape bias when the task is clearly a forced-choice pointing task (Field et al., 2016b; Hartley et al., 2019, 2020; Tek et al., 2008), whereas in a sequential sorting task this is not always the case (Hartley & Allen, 2014; Tovar et al., 2020). Specifically, when there are not multiple items competing for attention at the same time, autistic children appear to over-extend labels. In this research we compare two forced-choice and two yes/no tasks to explore how different instructions interact with the other demands.

Rationale for alternative format

The current studies contribute to telling the same story in three distinct parts, each with their own questions and contributions to the overall thesis. The first paper focuses on the question "Do autistic children have a shape bias?". Using highly contrasting stimuli, such as used in Field et al. (2016b), the task is designed to give autistic children the optimum chance of success. Items are perceptually distinct, it employs a touch-screen task with no visible 'speaker' to attend to socially, and both yes/no and forced-choice variants of the NNG task were used. The study was designed to test whether standard visibility and labels support, or interfere, with shape bias for autistic children. By replicating conditions that have previously encouraged shape bias in autism, then changing those conditions, we can determine if the task demands themselves have a different impact on shape bias for autistic and TD children.

The second paper asks if verbal labels are important to shape-based categorisation, and whether there is evidence that they support generalisation for 'basic' level categories. Specifically, children demonstrate a stronger shape bias for labelled items than they do for unlabelled items (e.g. Landau et al., 1988). While some accounts describe this as a temporary developmental window (Landau et al., 1988), others consider the label effect as a defining characteristic of shape bias (Tek et al., 2008). Here, we use an identical procedure to paper 1, this time using perceptually similar stimuli. In this task, TD children are expected to exhibit a stronger shape bias for labelled stimuli than non-labelled stimuli, which would give the opportunity to compare the effect of labels for autistic children.

The final paper focuses on the underlying mechanisms that could cause a difference in shape bias in autism, specifically testing the hypothesis that autistic children will pay attention to small details of the stimuli at the expense of the global shape due to weak central coherence

(Happé & Frith, 2006). Using the same screen-based game as the other experiments, this study uses stimuli designed to have a small-but-salient feature that could draw attention away from their shape.

Each of these three papers offers an independent and unique insight into how shape bias presents in autism. Yet together, they contribute to a clear argument that each of these methodological choices has an impact on the strength of the shape bias observed for autistic participants. Furthermore, they combine to make a compelling case that autistic children do prioritise shape for categorisation.

Account of construction

This thesis consists of five chapters. Chapter 1 reviews the current state of research into shape bias in autism. In reviewing recent inconsistencies in the findings, we consider whether methodological choices that appear to be suitable for investigating shape bias in TD children are equally suitable for autistic participants.

Chapter 2 includes the first paper of the study: "Task demands affect 'shape bias' for autistic children during novel noun generalisation". This describes two experiments, a forced-choice and yes/no task, investigating when autistic children exhibit a shape bias using 'high contrast' stimuli.

Chapter 3 consists of the second paper: "Shape bias for perceptually similar stimuli during 'online' and 'offline' novel noun generalisation for autistic and typically developing children". In another two experiments, forced-choice and yes/no tasks were repeated using 'low contrast' stimuli. With stimuli that were similar to the standard, these experiments were able to investigate the differing effect of 'standard visibility' and 'labels' on shape bias for autistic and TD children.

Chapter 4 is the final paper of the thesis: "Attention to global shape, not local features, informs offline noun generalisation in autism and typical development". The final experiment tests whether a local attentional bias disrupts shape bias for autistic children with stimuli that include a prominent feature.

In Chapter 5, we discuss the findings of the five experiments and how they contribute to our understanding of how, and when, autistic children generalise by shape when introduced to new categories.

Outline contribution

I am the primary author and lead researcher on all three papers that form this thesis. While my supervisors have been an unending source of advice and support, the conception, design, execution, data analysis and writing of each element of this research project is my own work. Dr Calum Hartley and Dr Katie Twomey have provided written feedback on drafts, recommended reading, and prompted me to question my ideas where appropriate.

Account of test administration procedures

This project was designed so that all participants could complete all five experiments. As the forced-choice and yes/no variants of the study used some of the same stimuli, a testing schedule and counterbalancing strategy was designed to minimise the risk of earlier tasks influencing responses on later ones.

The five experiments were completed over three sessions, each held on different days. These were presented in fixed groups, with two main considerations. Firstly, tasks that used the same stimuli were assigned to different sessions, to ensure a break of at least one day between them. Secondly, only one yes/no task was included in each session as this task had more trials and hence took longer to complete. This resulted in three experimental sessions as follows:

- Session A: Forced-choice with high contrast stimuli; yes/no with low contrast stimuli
- Session B: Forced-choice with low contrast stimuli; yes/no with high contrast stimuli
- Session C: Global shape vs. local feature yes/no task (unique stimuli)

All participants completed all three sessions, one during each visit. A typical visit began with a standardised test, administered in a fixed order: BPVS (1st visit); EVT-2 (2nd visit); Leiter-3 (3rd visit). Following a standardised test, children completed either session A, B or C. The order of these was determined by the following counterbalancing process.

For sessions with two experiments, the order of the high contrast/low contrast studies was counterbalanced. Half of the participants completed a study with high contrast stimuli first. The same order was used for both sessions A and B, meaning that a participant who saw a forced-choice task first on session A would see a yes/no task first on session B, both with the same stimuli. This ensured there would always be at least one task involving different stimuli completed in-between.

Each experiment included the within-participants variable 'standard visibility', which consisted of an online and offline condition. The presentation of these conditions was also counterbalanced, with half of the participants completing the online condition first. To maintain an achievable number of combinations, this order was also held constant for every task that participant completed. There were also label/no label conditions in each experiment. These were presented in a fixed order: no label trials followed by labelled trials. This was due to the possibility that experiencing labelled trials first may prime children to use the same strategy on the subsequent no label trials. Within each condition, the individual trial order was randomised by the computer.

Taking each of these factors into account, there were 6 possible orders that sessions A, B and C could be completed in, within which there were a further 4 orders for the tasks and standard visibility conditions. This resulted in 24 presentation orders for the 5 experiments that met the described criteria. To allocate these orders to participants, each order was entered into a separate row on an Excel spreadsheet and shuffled using a random number generator. Autistic and typically developing children were then assigned a row in order of participation within their group, starting from row 1, so each combination was allocated to at least one participant in each group.

While the intention was for all participants to complete all experiments, disruption to testing caused by Covid-19 meant several participants were unable to do this. Table 1 shows the overlap of participants across the five experiments.

Table 1: Participant overlap across experiments

	Session A		Session B		Session C
	Forced-choice (High contrast)	Yes/no (Low contrast)	Forced-choice (Low contrast)	Yes/no (High contrast)	Local/Global
Total participants	30	29	27	30	34
All sessions done	25	24	22	25	24
Two sessions done	4	1	1	4	3
One session done	1	4	4	1	7

Note: These figures represent participants whose data were included in the final analysis reported in each chapter.

Chapter 1: Literature review

1.1 The Word Learning Problem in Typical Development

It is not uncommon for literature on language development to begin with the philosophical example of Quine's 'gavagai' (1960). Quine imagines a linguist in a foreign land seeing a rabbit running past. Their guide, a native speaker of the local language, says 'gavagai'. How does the linguist know that their native guide is referring to the 'rabbit' and not one of countless other possibilities in the scene? This hypothetical situation is often used to illustrate the challenges children face when learning new words and acquiring categories. However, Quine intended to describe the problem of translating an unknown language into one that is already known, so for young children this does not tell the whole story. The adult linguist already has a strong concept of what a 'rabbit' is, and what other things can and cannot be called by the same name. Consequently, once they identify that 'gavagai' is the local word for 'rabbit', it comes with an existing knowledge of what the that means and where the category boundaries lie. One could make the mistake of using 'gavagai' to refer to a hare, however, the guide would likely still understand the linguist's intention, and perhaps even offer a correction. For infants, identifying the referent is only one part of the problem. They do not begin with a fully developed concept of 'rabbit' waiting for a label; this knowledge of what rabbits are, and what other entities can be referred to by the same word, must also be acquired over time. It would not be very useful to link 'gavagai' to an individual instance of 'rabbit' or to learn nothing about what else can also be called that. Markman (1989) called this the 'problem of induction', referring to children's need to infer what a new word relates to in the moment, and how that can be applied beyond the current situation as generalised learning about the category.

For infants, the difficulty of generalisation can be thought of as at least three problems: “what does the word refer to in this instance?”; “what else can that word refer to in the world?”; and “what does the word *mean*?”. Fortunately, TD children find solutions for these problems, and usually come to accomplish the task of learning new words without much difficulty. Explaining how they do this presents the challenge, and numerous abilities have been identified as potentially necessary for achieving efficient word learning, such as inferring the referential intention of an adult speaker (Akhtar et al., 1996; Baldwin et al., 1996; Bloom, 2000; Keates & Graham, 2008; O’Hanlon & Roberson, 2007), sensitivity to syntactic structure (Bloom, 2000; Hall & Graham, 1999; Landau et al., 1992), and the ability to engage in joint attention (Baldwin, 1991, 1993; Bloom, 2000). Furthermore, associations between a new word onto a novel referent can be accomplished from the first pairing by a process of “fast-mapping” (Carey, 1978, cited in Carey, 2010). With fast-mapping, children are able to make a best guess on what an unknown word refers to, as other contextual information constrains the possibilities. For instance, instructing children to get ‘the chromium one, not the red one’ provides a cue that the new word is a colour (Carey & Bartlett, 1978, cited in Carey, 2010). Nouns may also be fast-mapped onto novel objects by ‘mutual exclusivity’ (Markman, 1989), which predicts that children will assign novel names to objects that they do not already have a label for. While this ‘fast-mapping’ alone does not result in long-term learning of word-object relations (e.g. Horst & Samuelson, 2008), it does allow infants use that information in the moment to guide their attention without having any deeper conceptual knowledge. Constraints such as mutual exclusivity offer a solution to the induction problem by reducing the likely referents of a new word to a manageable number of possibilities (Markman, 1989). However, making a label-object link in the moment is only the first problem. Learning that an individual instance of a furry red block, for example, is called a ‘dax’ is of limited use. It is the ability to generalise that learning beyond the original instance to a broader ‘dax’ category that can be an extremely powerful tool.

Landau et al. (1988) proposed a process through which infants could solve the problem of generalising a label beyond a specific naming event by using perceptual similarities between objects. They suggested that infants begin to associate certain perceptual features as being important to category membership, and so give those features preferential attention. One such property is object shape, which may offer an ideal trade-off between minimum amount of information to be processed with the maximum predictive power for basic level category membership in the English language (Rosch et al., 1976). Landau et al. (1988) suggested that young children may initially be concerned with what adults use a word to refer to, rather than what things actually are at a deeper level, and so demonstrate a ‘shape bias’ when asked to extend labels to new examples. They demonstrated that children are likely to generalise newly learned nouns to objects that are the same shape as the original, while being less likely to generalise to objects that are made from the same material if there are a different shape. Beginning from an assumption that ‘things that look the same are likely to be called the same name’ can provide a useful point of shared understanding while children are still in the process of learning about the category. Even if the child extends incorrectly, for instance calling a ‘zebra’ a ‘horse’, the shape assumption provides an insight into the child’s intention and so communication does not break down. Additionally, this gives the opportunity for the adult to offer a correction so word learning continues to develop.

Landau et al. (1988) demonstrated the ‘shape bias’ in a series of studies varying the magnitude of change between ‘dimensions’ of perceptual similarity: size, shape, and texture. If shape did indeed have a special status as being indicative of category membership, then small changes to shape could discourage participants from believing that objects belong to the same basic category, while large changes to size and texture may be accepted providing that shape remains constant. On the other hand, if shape was not special and just happened to be the most salient change, large changes in size or texture should be more disruptive than minor shape changes. What they found was that both 2 and 3-year-olds were sensitive to the importance of shape, and that the shape bias was stronger when the objects were labelled. Also, the bias was thought to be more fragile in 2-year-olds and became stronger with age. Adults almost exclusively chose same shape items regardless of whether or not they were presented in a word learning context, suggesting that this bias develops as experience with categories increases. Crucially, these results were obtained using novel objects and words that children did not know the meaning of, indicating that infants can use perceptual features for generalisation in the absence of pre-existing conceptual knowledge.

While Landau and colleagues demonstrated that shape was the most important property in their word learning task, this ‘shape bias’ was to become a source of contention. The authors proposed that this bias was learned: as infants notice that the majority of objects they know the names of can be classified by shape, they form an association between object names and shapes. As concluded in Smith et al (2002), young children's attention becomes 'trained' over time to prioritise shape over other perceptual features when a new word is encountered. In this way, the shape bias begins as a word learning strategy that does not require a deeper knowledge of the meaning of the word being learned.

It is important to acknowledge that this is not the only mechanism through which language can be learned. Rather, it represents one of many biases and constraints that appear to be present in children’s word learning. This collection of biases can be thought of as a ‘toolbox’, with no single tool being more important than any other in general - it is the context that determines the usefulness of each tool. For instance, just as you cannot use a screwdriver to hammer a nail, you cannot use a shape bias to learn verbs. However, for the specific word learning situation that infants regularly find themselves in - needing to have a reference system with their caregivers for things in their shared environment - a shape bias can be an extremely useful tool. Hence, it is natural to reflect on the implications of shape bias when language acquisition appears to be disrupted, as is the case for many autistic children (Anderson et al., 2007; Norrelgen et al., 2015). If a shape bias is a tool to support word learning without being necessary for it, one could expect a lack of shape bias to perhaps result in a slower, but not absent, rate of word learning. So, it is possible that autistic children who do have language delays are not benefiting from the use of this tool, either because they do not have it or cannot identify the best time to use it.

1.2 The Word Learning Problem in Autism

One of the diagnostic traits of autism is difficulty initiating and sustaining social communication (World Health Organization, 2022). It is important to acknowledge that language delays or difficulties are not a diagnosis-defining characteristic. While around 25-30% of autistic children are estimated to have little or no spoken language by the age of 9 years, others may have age-appropriate language abilities (Anderson et al., 2007; Norrelgen et al., 2015). However, language delays are often one of the first symptoms noticed by children’s caregivers (Kim et al., 2014), and it is not uncommon for children to have uneven language abilities (e.g. greater word production accompanied by lower word comprehension; Charman et al., 2003; Lazenby et al., 2016). Lower receptive and expressive vocabularies in childhood have also been associated with language outcomes in adulthood for autistic individuals (Mawhood et al., 2000), and functional language at 5 years predicts ‘positive’ outcomes in other areas of adult social life, such as employment and close relationships (Lotter, 1978).

The proportion of autistic children with minimal language has decreased, likely due to a combination of factors including earlier diagnosis and intervention, along with broader diagnostic criteria (Tager-Flusberg & Kasari, 2013). However, the causes of these difficulties and the heterogeneity of language ability within the autistic population remain unclear. Anderson et al. (2007) tracked the language skills of autistic children from 2 to 9 years and identified at least 4 different ‘groups’ of language development patterns, ranging from non-verbal at the age of 9, to rapid progression to above average verbal ability. It may be that different areas of difficulty have different underlying explanations. However, identifying any difference could have important implications for future interventions. Eigisti et al. (2011) suggested that all autistic individuals do have some language differences compared to typical development, and that heterogeneity in ability is due to differences in the areas of language that individuals find particularly challenging. For instance, autistic adults and children without intellectual disability can have differences in pragmatics and reciprocal language use that are not detectable using standard language measures (Colle et al., 2008; Naigles & Tek, 2017; Suh et al., 2017). In this regard, some aspects of language, particularly concerning social communication, may still offer a commonality between autistic individuals.

Despite population heterogeneity, it is evident that a large proportion of autistic children can acquire and use language fluently. Autistic children with typical language skills follow the same trajectory of language development as TD children with regards to syntax and grammar (Tager-Flusberg et al., 1990). Similarly, verbal autistic children have shown similar vocabulary structure and grammatical abilities to non-autistic ‘late talkers’(Ellis Weismer et al., 2011), and a similar pattern of word comprehension and expression to TD children that is merely delayed (Charman et al., 2003). They also have a typical proportion of common nouns in their productive vocabulary by the time they exceed 50 words (Charman et al., 2003). In other words, while language can be delayed in autism, once underway acquisition follows the same pattern as expected in typical development.

Children across the autism spectrum perform well in tasks that use linguistic word learning constraints for identifying the referents of new labels. For instance, they use the principle of mutual exclusivity to map novel labels onto an unknown object (de Marchena et al., 2011; Preissler & Carey, 2005), demonstrate a ‘noun bias’ by classing novel words as nouns rather than verbs (Swensen et al., 2007), and can fast-map labels onto referents (Hartley et al., 2019). These abilities should be useful for mapping novel words onto an individual instance of a referent, as in the example of identifying the rabbit as the ‘gavagai’. However, despite demonstrating strength in these abilities, language delays are common in autism. This implies that there must be other factors that contribute to word learning in typical development that may be more challenging for autistic children. If a shape bias is an important tool for generalising novel nouns to new instances, a reduced ability to use this strategy could impact the efficiency of autistic children’s word learning. This could result in the need to acquire language using slower and more effortful means. To date there is some evidence that this may be the case, which will be discussed in section: Shape bias in autism. However, historical debates around the nature of the shape bias in typical development have result in inconsistent approaches to research into shape bias in autism. Recognising and reconciling these disagreements in the classic shape bias literature is essential to provide a stable basis from which to discuss atypical examples.

1.3 Debates around the Shape Bias

1.3.1 Attentional learning account

Despite its seemingly simple definition, the ‘shape bias’ has been a surprising source of debate. The original explanation offered by Landau et al. (1988), and subsequently named the Attentional Learning Account (ALA) (Smith et al., 1992, 1996), was that shape bias is the result of an attentional process that is optimised for learning count nouns. The authors were open-minded about the reasons for such a bias, suggesting that object shapes may serve as clues to object functions or deeper properties. They also suggest that shape could be a shared point of reference for adults and children, from which children can gain more knowledge. As highlighted by the ‘gavagai’ problem, using the wrong label for something that is perceptually similar at least gives both speakers the best chance of understanding what was meant. Furthermore, it can serve as a foundation for teaching and learning the correct label. Though Landau et al. (1988) discuss why such a bias is useful, the ALA line of research explored the contexts in which children demonstrated it. So how did children know when to generalise by shape and when not to?

Firstly, shape bias is specific to count nouns. For mass nouns which are not ‘countable’, as is the case for non-solids such as gel or sand, a substance bias can be demonstrated (Soja et al., 1991). As non-solid substances are consistently made of the same material, but can change shape, substance and not shape is the most informative cue for generalising mass nouns. Children appear to be able use the most useful attention bias for the broad class of words they are engaged in generalising, however only after they have enough of those words in their vocabulary structure to support it (Perry & Samuelson, 2011; Samuelson & Smith, 1999). With experience, children can learn associations between classes of word and their most predictive, perceivable characteristic.

Secondly, children's attentional biases are sensitive to syntactic cues in the instructions given. When novel words are presented as adjectives, as in "Is this a dax one", children attend to texture (Smith et al., 1992). Similarly, when the higher level category is alluded to, as in "This is a kind of Dax", texture can become more important than shape (Landau et al., 1992). In this account, one possible explanation of the label effect in shape bias research is that non-labeled trials are interpreted as superordinate categorisation tasks, so higher shape variability is accepted. Furthermore, not all categories of count nouns are equally well-organised by shape. For instance, living creatures can move, so some variation in shape can be expected. When eyes are added to items to give them the appearance of being animate, children will tolerate greater changes in shape and generalise labels by texture (Jones et al., 1991). So, the shape bias is also specific to artefacts in this account.

A further question is whether the shape bias begins as a general bias and becomes more specific during development, or whether it only emerges once children have sufficiently developed concepts of different object kinds. In the ALA, the shape bias is also framed as a linguistic effect. Landau et al. (1988) suggested that the shape bias begins as a word learning strategy and becomes generalised to non-lexical tasks during development. This body of research paints a picture of processes that are sensitive to both visual properties of objects and lexical cues from speakers, which are then used to direct attention to the most informative properties based on what has already been learned.

Smith et al. (1996) explicitly tested whether attentional mechanisms underlying shape bias are automatic or consciously controlled. They hypothesised that if children's attention to shape is a conscious, directed strategy, that they would be sensitive to contextual information presented during a task. Participants were shown novel objects with smaller parts attached to them, such as a chicken-wire cylinder with metal clips. A function was demonstrated either for the ‘global’ base shape (e.g. looking through it like a telescope) or the ‘local’ part (e.g. holds a pencil). They found that 3-year-old children used this functional information to make decisions about whether items were similar to each other, however, they only extended labels on the basis of salient perceptual features. Smith et al. (1996) suggested a possible explanation that there are two different processes in effect: that decisions about kinds of things are consciously directed and can make use of the additional information about object function, whereas for label extension, attention is automatically drawn to the salient shape. Furthermore, Landau et al. (1998) found a developmental trend for children taking context into account when generalising. Three-year-old children extended labels the basis of shape alone, while five-year-olds and adults took function into account. Young children learning labels may be more concerned about what kind of things can be called by that word rather than any deeper meaning, as a basis for shared understanding with adults. Real cars, toy cars, pictures of cars, talking cartoon cars, cuddly cars, and car cakes may all be called ‘car’ without any expectation that they all have the same function.

A key claim of the attentional learning account is that it arises from higher order generalisations once children have a sufficiently large vocabulary (Smith et al., 2002). Around 50 count nouns are thought to be enough for children to develop an association between names and shapes, leading them to apply a shape bias to newly encountered words and objects. The logic of this argument is that infants begin with no attention bias and extract all the available information from their environment over time. With exposure, they learn that some perceptual properties are correlated with names more often than others, and this leads them to prioritise shape as a basis for generalisation. Since resources are now directed primarily to the most predictive features, and not expended unnecessarily on less relevant information, the learning process becomes more efficient and children's word learning can accelerate (Samuelson & Smith, 1999). The emergence of a shape-bias appears to precede a 'vocabulary spurt', a period around infants' second year where the rate of word learning increases (e.g. Goldfield & Reznick, 1990). This suggests that it may serve an important function in supporting word learning without being essential.

Smith et al. (2002) demonstrated that shape-bias training can affect the rate of language acquisition. They conducted a training study with 17-month-old infants and found that training children with categories where shape was an informative feature resulted in accelerated vocabulary acquisition outside of the lab, according to parent reports. By contrast, training with categories organised by colour or material, or unlabelled training, did not have a facilitative influence on children's vocabulary development. These results may be evidence that labels are important for guiding attention to the relevant features of objects. Children who learned unlabelled categories were able to make 'first-order' generalisations, meaning they could identify novel exemplars of those categories on the basis of shape. However, only training with named categories facilitated shape-based generalisation for a new category that children had not encountered during training, suggesting the labels played a role in supporting higher level generalisation.

While a shape bias is a strong focus of the ALA, it should be noted that it is not a 'theory of shape bias'. ALA is a theory of word learning, in which automatic attentional processes are tuned through early experience and engaged to support language acquisition. A shape bias is just one such attentional bias, which Landau et al. (1988) suggested could be the 'default' in the absence of any other information. They introduced the notion of 'bundles' that could be activated for particular things (e.g. an ‘animal’ bundle could weight texture more heavily and allow more shape changes than an ‘artefact’ bundle). This attentional learning account, in essence, describes an ongoing process in which the infant brain learns where resources are best directed for any new name it encounters. This attentional system may use contextual cues as a guide (Smith et al., 2010).

However, the ALA is not without its critics, and some researchers have questioned whether a vocabulary spurt even exists (Bloom, 2000). The opposing side of the shape bias debate was dubbed the ‘shape-as-cue’ account.

1.3.2 ‘Shape-as-cue’ Account

The ‘shape-as-cue’ (SAC) account was proposed by Bloom (2000) as an alternative explanation for a shape bias. The principle is that children understand that count nouns refer to *kinds* of objects, and that a shape bias occurs because shape is a useful cue to the ‘deeper’ properties of the object. According to this account, when young children learn new words they are engaging in learning what those words mean, in contrast to the ALA suggestion that children begin by learning what the words can refer to in the world.

In support of the SAC account, Diesendruck and Bloom (2003) demonstrated that 2- and 3-year-old children were just as likely to generalise by shape when they were asked to find another ‘kind of’ a novel object as they were when the object was labelled. They argued that this suggests that children infer that labels refer to object kinds, as an associative link between the label and shape without this inference should result in a stronger shape preference for labelled items. They also did not find differences between 2 and 3-year-olds, suggesting that the shape bias was not becoming stronger with age. The lack of developmental change, and the absence of any effect of label, initially appears to conflict with the findings of Landau et al. (1988). However, large differences in the stimuli may account for these two studies' contrasting findings.

Diesendruck and Bloom (2003) presented children with test objects that matched the standard on either shape, colour or texture, however the colour- and texture-match competitors were a completely different shape from the standard. The only plausible explanation for choosing a competitor as an example of the label would be to map the label onto the material or colour of the object instead of shape. On the other hand, Landau et al.'s (1988) stimuli consisted of test items of varying magnitudes of shape change, and in all cases there was still some overall similarity with the shape of the standard. The question was not whether a child would accept a completely different shape if the texture was correct, but how much change in shape the child would tolerate before they refused to extend the label. Landau et al. (1988) suggested that the presence of a label could cue children into looking for basic level categories exemplars, whereas the use of the more ambiguous phrasing of 'kind of' or 'goes together' could lead children to look for superordinate exemplars. The highly different stimuli used by Diesendruck and Bloom (2003) could be so divergent from the target that they could not be interpreted as belonging to the same superordinate category, leaving the shape-match as the most plausible choice regardless of the presence of a label.

The SAC account also draws on evidence from examples where children do not exclusively use shape for categorisation. For instance, TD children reflect on an artist's referential intentions when naming a drawing. For example, similar looking drawings will be called a 'balloon' or a 'lollipop' based on what it was intended to be during creation, rather than what it looks like (Bloom & Markson, 1998). However, this reluctance to just use shape when there is other information available does not run counter to the ALA's claims. What it suggests is that children are sensitive to multiple cues, including communicative intentions, and that these other cues can override perceptual similarity depending on context.

Booth and Waxman were also critical the ALA's claim of a 'Stroop'-like automatic attentional process, arguing that word learning was 'smart' (Booth & Waxman, 2002b) and sensitive to contextual information (Booth & Waxman, 2002a). Using the same sort of stimuli as the original Landau et al. (1988) study, Booth and Waxman (2002a) used vignettes about objects to manipulate the contextual information provided to 3-year-old children. In one condition, the vignette presented the object in a way that was consistent with it being an animate creature (e.g. a dax with a mummy and daddy that loved it), while the other condition described it as an artefact (e.g. a tool for a spaceship). They found that children's label extensions were affected by the different vignettes, and the story that presented the novel object as animate encouraged extensions on the basis of both texture and shape. They present this as evidence against ALA on the grounds that contextual information was taken into account, calling into question the role of an automatic attentional process.

Booth and Waxman make a valid case for the importance of conceptual information, demonstrating that shape bias can be overridden when there is additional information (Booth et al., 2005; Booth & Waxman, 2002a). Children as young as 1.5-years can generalise by shape for artefacts, and both shape and texture for animates, when they are told a story that provides the appropriate context. They argue that this is evidence against the claim that word learning is purely an automatic perceptual process that is impervious to deeper knowledge. While the importance of conceptual knowledge must be recognised, however, proponents of the ALA never claimed that it should not. Rather, they suggested that shape may initially be more important than conceptual information for very young children, and that this would shift with age. For instance, 2 and 3-year-old children may generalise labels by shape even when they have been given some information about the function of the object, while older children and adults will take the function into greater consideration (Landau et al., 1998).

Cimpian and Markman (2005) went one step further and claimed that a shape bias was nothing more than an experimental demand characteristic. They describe a series of six studies which they claim challenge the existence of a shape bias, as they show children extending labels to objects that are the same taxonomical category over objects that are the same shape. In a forced-choice task, children were presented with a shape-match and a taxonomical match for the standard and asked to generalise a novel label to one of them. To test for demand characteristics, some of the experiments also included a ‘none-of-the-above’ option, so children were not ‘forced’ to choose. They found this option eliminated children’s shape choice preference but not their taxonomical match preference. Cimpian and Markman argued that these results show that children see labels as applying to object kinds, not object shapes, in line with the SAC account. However, these studies all used known objects and asked children to extend a new ‘froggy language’ word for the benefit of a frog puppet. For instance, if the target was a carrot, the shape-match was a screw and the taxonomical match was a potato. As the participants were aged between 3 and 5 years, they would likely be familiar with these items already. The use of familiar items is crucially different to studies taken as evidence for the ALA, which typically use novel objects of which children have no existing taxonomical knowledge. Applying a new word to an item that already has a label is a different task. Children already have a concept of a ‘carrot’, ‘screw’, and ‘potato’, so it is not surprising that they may be reluctant to call a carrot and a screw by the same name when they already know they are unrelated. This scenario is more akin to Quine's 'gavagai' problem of translation. The ALA, on the other hand, proposed a way for new words without pre-existing information to enter the vocabulary, and did not predict that children would prioritise shape over their existing categorical knowledge (Smith & Samuelson, 2006).

1.3.3 Is shape bias specific to lexical tasks?

The specificity of the shape bias as a lexical effect has been another topic of debate. Landau et al. (1988) reported that the shape bias was stronger when objects were labelled than when they were not. However, this effect disappeared as children got older, and shifted to a shape-based generalisation strategy regardless of naming. The researchers suggested that a shape bias is rooted in lexical acquisition. Shapes and labels become associated through repeated pairing, but may become more general as children's understanding of deeper category meaning increases (as suggested by the SAC account). This link between categories and their names is undeniably important (Markman, 1989), and from early development there is evidence of labels supporting perceptual categorisation by drawing attention to similarities (Althaus & Plunkett, 2015) and shaping the categories that children form (Westermann & Mareschal, 2014). Labels may support perceptual categorisation by providing a feature to bind to the perceptual cue (Samuelson et al., 2013), thus supporting the formation of quick perceptual categories without the need for deeper conceptual knowledge. Object shape and label can become bound together. For children, Xu (2002) suggested that labels can serve a role as 'essence placeholders' prior to concept acquisition. This idea is also captured in adult concepts where words have been described as linguistic 'shortcuts' that allow us to manage complex ideas without the need to consciously represent all of that conceptual information every time (Connell, 2019). For instance, as adults can talk about a 'dolphin' without needing to mentally think about everything they know about dolphins, infants can learn what things might be called 'dolphin' before that information is acquired.

While the noun-category linkage is undoubtedly important, what this actually means for a shape bias has varied between studies. In the ALA studies, any prioritisation of shape over other perceptual features for generalisation is described as having a shape bias (Landau et al., 1988). Furthermore, for 3-year-old children, the position of a label in the sentence promotes attention to shape even when children would classify items by material in a non-lexical task, such as for non-rigid objects (Samuelson & Smith, 2000a). However, by the time children are 4 years old they no longer overgeneralise non-rigid objects by shape (Samuelson et al., 2008). This suggests that while children do learn an association between count nouns and shape, over time they realise this is not a strict rule and can apply it more flexibly.

1.3.4 Is this all a misunderstanding?

Much criticism of the ALA stems from differing interpretations of what the theory actually predicts. In short, the SAC account offers a top-down explanation of a shape-bias, where labels apply to object kinds. Infants' interest in shape is a by-product of the fact that an object's shape provides children clues about its 'deeper' ontological nature. SAC presents this as a solution to the ALA claim that word learning is based on purely perceptual associative processes that are impervious to the influence of conceptual knowledge. However, under closer inspection, many of these disagreements appear to be based on assumed claims that were not actually present in the ALA account. For instance, Markson et al. (2008) argue that SAC predicts that shape will be useful for general categorisation whereas ALA predicts that shape is primarily for lexical generalisation. However, original ALA shape bias investigations also included adult participants who generalised by shape with great consistency regardless of whether items were labelled or not, leading authors to conclude that shape bias may initially be a lexical effect that becomes more general with experience (Landau et al., 1988). Furthermore, claims that ALA sees shape bias as impervious to conceptual knowledge (Booth & Waxman, 2002b) seem to overlook ALA evidence that perceptual biases can be adjusted on the basis of existing knowledge. For instance, the ability to generalise by texture when eyes are added to an object (Jones et al., 1991) implies some conceptual knowledge that things that have eyes tend to be alive, and hence employing strategy for classifying living things would be more useful. Landau et al. (1988) in fact suggested that contextual information would influence category judgements and that perceptual features would be important when no other information was available.

One reason for the apparent lack of agreement may simply be that the approaches are looking at slightly different problems at different levels of explanation. While SAC is concerned with how children learn the meaning of words, ALA tries to explain how children learn what words can be used to refer to, which are two different aspects of the word learning problem. As suggested by Quinn and Eimas (1997), categories may begin as perceptual and become conceptual over time. Taking a dual process approach that perceptual and conceptual categorisation are initially separate (Mandler, 2000), with perceptual categories organized by what things look like and conceptual categories based on what things actually are, ALA and SAC may be answering different aspects of the problem. SAC provides an account of how shape can be linked to word meaning and conceptual categories, while ALA focuses on how those percepts develop in the first place. However, the SAC account of shape bias is lacking as a developmental account because it fails to address how that strategy is acquired and how it changes with experience. Linda Smith and colleagues made attempts to clarify the ALA position and find parity between the approaches (Smith et al., 2003; Smith & Samuelson, 2006), however, by this point the apparent conflict was so entrenched in the literature that an alternative approach would be needed to explore the common ground.

1.3.5 Dynamic systems approach

Attempting to find a balance between the SAC and ALA explanations, a Dynamic Systems Theory approach to shape bias (Samuelson & Horst, 2008) acknowledges the role of knowledge and seeks to explain how knowledge is utilised in the moment for the task at hand. Samuelson and Horst claim that ALA characterised knowledge as an “emergent product of the real-time interaction of many components” (2008, p. 210), including the child’s cognitive abilities and task demands. In other words, the response that children give on a generalisation task is affected by multiple factors that are enacting on them in the moment, such as: the knowledge they bring with them; their understanding of the task; where their attention is directed at the time. Furthermore, these factors are not independent, and interact with each other. For instance, children’s existing vocabulary can affect where they direct their attention, but in turn acquiring new words adjusts that underlying vocabulary structure (Smith et al., 2002). In a dynamic system, a shape bias is a behavioural response that arises in the moment when multiple processes converge on that being the most useful bias to apply at the time.

Colunga and Smith (2008) extended the ALA explanation of novel noun generalisation by incorporating dynamic principles, creating a 'D-ALA' model. This model included the ALA idea that attention to the most predictive perceptual feature develops from past experience, and so biases develop slowly over time. However, the authors argued that the response children would give on any given task would not be solely dependent on this long-term experience, and short-term experience within the task would also play a role. Their model considered three 'nested' time scales that could influence whether a shape bias was applied during a task: (1) the long-term knowledge the child brings with them, (2) the information available while completing the task, and (3) the immediate response given. While the model provides an account of how previous knowledge can produce attentional biases, and that these biases are flexible and responsive to the relevant contextual information during the task, it is not presumed that these are the only influences on the final response. A dynamic system would also allow for existing conceptual knowledge to be taken into account, perhaps overriding automatic attentional processes when the situation calls for it. The appeal of this approach is that it considers the interaction between top-down and bottom-up processes, allowing a way for new categories to be acquired before a deeper understanding of the concept is achieved, while still offering a way for that knowledge to inform children's responses on a given task.

Kucker et al. (2019) used shape bias as a case study for how wider theoretical understanding can arise from seemingly contradictory findings. Pitting ALA and SAC against each other as conflicting accounts may be counter-productive to seeing them as different facets of the same concept. Kucker et al. (2019) propose that the last 30 years of shape bias research shows us where a shape bias is placed in the route to word learning mastery, but it is just one stop on the way and is not the only available path. So, the question should not be ‘why do children have a shape bias?’ with one or the other of these explanations being the ‘correct’ answer. Instead, we should be looking at when children show a shape bias, and in what contexts does it seem to be helpful. The observation of a shape bias in a particular task may arise from a combination of factions, including children’s attention in the moment and their existing conceptual knowledge of object kinds. However, with many years of theoretical disagreement in research with typically developing children, attempts to investigate this bias in atypical development has been influenced. If we cannot agree on whether shape bias is a lexical phenomenon in typical development, for instance, how do we make theoretical predictions about children with language delays? We need to clarify our position on these issues in order to make clear predictions about atypical development. As we will see from the existing body of shape bias in autism research, unanswered questions on the TD theories are simply leading to more unanswered questions in ASD theory rather than clarifying the situation.

1.4 Shape bias in autism

An early attempt to investigate shape bias in autism was made by Tek, Jaffery, Fein, and Naigles (2008). Tek and colleagues suggest that difficulties with abstraction in autism may lead to poor shape bias. They conducted a study with 14 autistic children aged between 2 to 3 years and 15 TD children aged 1.5 to 2 years matched on expressive and receptive vocabulary. To investigate the development of a shape bias, the children completed an intermodal preferential looking (IPL) task on four different occasions over a one-year period. The IPL paradigm was a screen-based task in which children were shown a novel object on the screen twice to familiarise them (once on the left and once on the right), followed by two new objects: a shape-match for the exemplar and a colour/pattern match. An accompanying audio track asked ‘which one looks the same?’ (for what they called the ‘NoName’ trials) or ‘Where’s the *dax*?’ (for the ‘Name’ trials), and the children were coded on which side of the screen they looked at in response.

This task differs from the methods that have commonly been used to investigate shape bias in typical development. A typical procedure involves physical objects, with the standard for the category in view on the table. This particular method was modelled after a noun bias task, which the children also completed as part of this study for comparison (Swensen et al., 2007). Tek et al. (2008) suggested that this method would allow for a more implicit measure of shape bias as the children are not required to make a deliberate, conscious decision between the two. For instance, children could think that both objects were plausible examples which would likely result in similar looking times for both candidates. They also defined a clear shape bias measure for this task, defined as ‘the percentage of time the child looked to the same-shape object during the NoName trial…as compared to the Name trial’ (Tek et al., 2008, p. 214). This distinction is important because, firstly, it assumes the shape bias is lexical and so measures the ‘label effect’ rather than a general preference for shape, and secondly, this definition is unique to the IPL method. In principle, using this equation a child exhibiting high preferences for shape in both labelled and unlabelled conditions would not be categorised as showing a shape bias. In practice, for the very young children involved in this study that would be an unlikely scenario, as a TD shape bias in two-year-old children is stronger for labelled items (Landau et al., 1988). However, this raises questions over the extendibility of the measure to older age groups.

For the shape bias, at the first visit (when the children were aged between 18-months to 3 years) neither group showed a preference for the shape-match object in the IPL task. By visit 2, the TD children looked at the shape-match more in the Name condition, and did so for more than 50% of the time, while the autistic group did not. Also, by visit 2, the TD group had surpassed the autistic group’s count-noun vocabulary size despite being matched at the first visit, which introduces some difficulties in comparing the groups. This pattern was replicated at the following two visits, with only the TD group showing a significant shape bias for labelled items. Tek et al. (2008) concluded that the autistic children did not demonstrate a shape bias in the IPL task, despite performing no differently to the TD group in the noun bias task. As the children otherwise had no difficulty with the IPL method when used for the noun bias task, this is taken as evidence that autistic children have difficulty using a shape bias as a word learning strategy.

In line with more traditional methods (e.g. Booth et al., 2005; Diesendruck & Bloom, 2003; Landau et al., 1988), the children in Tek et al (2008) also completed a pointing task using physical versions of the test objects. In this task, when the children were asked to explicitly select one of the items, both TD and autistic choose the shape-match item at a higher-than-chance rate from the very first visit, and the TD group only showed a 'label effect' at visit 3. While the different tasks appeared to yield conflicting results for autistic children, Tek et al (2008) suggested that the shape preference in the pointing task was indicative of a deliberate strategy, not automatic attention. Conversely, a shape bias in the IPL task would be observed if children had learned an association between object shapes and labels that would direct their attention towards the shape-match option in a naming context only. This explanation is consistent with the ALA account and suggests that automatic attention may not be employed by autistic children. If this is the case, an explicit, conscious awareness of the usefulness of shape may be needed to compensate for the lack of an attentional shortcut to word learning.

The implicitness of the children's response is not the only difference between the IPL task and the pointing version. In more conventionally-used NNG tasks using real objects, the standard usually remains visible for reference during generalisation, while in the IPL task only the test objects are visible on the screen during children's decision making. This could make the IPL task more challenging overall, particularly when shape changes between the objects are more subtle. Tek et al. (2012) tested the effect of the similarity of stimuli using this task with typically developing children and found that they looked longer at an 'overall match' when the stimuli were more similar, however, this has not been investigated with autistic children to date. Later studies have acknowledged the possible limitation of the increased memory demand (Potrzeba et al., 2015), but have argued that autistic children's success on similar IPL tasks measuring noun bias (Swensen et al., 2007; Tek et al., 2008) and syntactic bootstrapping (Naigles et al., 2011) is evidence against this being an issue. However, it could be argued that the working memory demands of such tasks are not equivalent to classic NNG tasks because they do not require generalisation. Both tasks involve remembering the same entity that children were familiarised with. There is evidence that TD children can remember a whole object at a younger age than they can remember the shape well enough to generalise to a new example (Perry et al., 2016). Furthermore, evidence of autistic children accurately generalising by shape from memory has used objects with easily differentiated, highly contrasting shapes (Hartley et al., 2019, 2020). It is possible that generalising from memory is more challenging or delayed for autistic children despite being able to perform well on recognition tasks using the IPL paradigm.

The finding that autistic children did not show a shape bias by Tek et al's (2008) definition has been replicated with a larger sample and a wider age range (Potrzeba et al., 2015), and has more recently been used to investigate individual differences in autistic children (Abdelaziz et al., 2018). In this extension, researchers found that autistic children who engaged in more instances of joint attention with a parent during a 30-minute play session were subsequently found to have a stronger shape bias on the IPL task. Conversely, a weaker shape bias was predicted for children who attended passively to their caregiver. They concluded that the weaker shape bias for children who engage in less joint attention may indicate the importance of understanding referential intent, as proposed by the SAC account (Diesendruck & Bloom, 2003). Abdelaziz et al. (2018) do stress, however, that these differences in social interactions alone do not explain shape bias differences. They suggest that both perceptual (as with the ALA) and conceptual (SAC) processes are necessary for a full explanation. In some ways this argument is similar to that of the dynamic systems account, seeing interactions between different processes and individual children's experiences contributing to shape bias performance. However, this explanation has not yet been applied to research in autism. Despite the move in TD literature to integrate conceptual and perceptual processes into a unified explanation, the discourse around shape bias in autism is still entrenched around ALA and SAC debates.

A clear illustration of this comes from Field et al. (2016b), who investigated whether the ALA or SAC account of shape bias was more suitable for explaining the performance of autistic children. Specifically, they argued if autistic children had a weaker shape bias than the other participant groups (typically developing children and children with other developmental disorders matched on receptive vocabulary), this would support the shape-as-cue account by indicating they had difficulty understanding the social intention of the experimenter. In this study the children had a choice of three items: a shape-match, a colour-match, and a texture-match. The objects in this study were real items, made from modified kitchen utensils or other household items. For half of the participants the standard was named, and the other half completed a no name condition where the item was simply referred to as 'nice'. After being introduced to the standard, the experimenter asked the participant to give them 'the other dax' or 'the other one' dependent on naming condition. In this study the differences between the objects were very clear, with obviously different shapes and uniform colours that were highly contrasted. The standard was also visible during testing. This set up should provide an ideal scenario for shape-bias to present itself as there was very little source of uncertainty in the trials.

Field et al. (2016b) found that TD children, autistic children and children with Down's Syndrome all selected the shape-match item at above-chance levels in the named condition, however, only the TD group did so on the no name trials. This result is interpreted as supporting the ALA account for the ASD and DD groups, and SAC for the TD group. What is not clear though is exactly how this conclusion is reached. The TD children did not show the 'label effect', however, there is also evidence that naming only affects shape bias during a short developmental window around 2-3 years old, after which children begin generalising by shape regardless of words (Landau et al., 1988). The lack of a label effect in itself is not incompatible with the ALA account, which makes the suggestion of alternative strategies less convincing as an argument. Abdelaziz et al.'s (2018) proposal that both explanations may contribute to shape bias is more appealing. Under this explanation, both processes could be used by both autistic and TD children, however individual differences could explain differences in the balance between them.

In a separate line of research, Hartley and Allen (2014) hypothesised that autistic children may not be able to identify shape as the most salient feature relevant to category membership, perhaps due to weak central coherence, and consequently generalise newly learned words on the basis of other visual features, like colour, in addition to shape. The participants in their study were autistic children with profound language learning difficulties, and TD children matched on receptive vocabulary. In their task the children played a sorting game. They were taught a novel name for a novel object (e.g. "blicket"), told to put all 'blickets' in one box, and other things into a different box. A mixture of items, and pictures of the items, were used as stimuli to determine if autistic children had more difficulty understanding that the new words also referred to images of the object. The results suggested that while TD children only generalised on the basis of shape, autistic children generalised the new label to items that matched on both shape and colour, just shape, or just colour. This over-extension of labels to irrelevant features suggested the autistic children did not preferentially weigh the shape over other perceptual cues when making their decisions. Hartley and Allen (2014) suggested that the reason could be due to a combination of poor prototype formation and weak central coherence. A detail-focused attentional bias in autism (Happé & Frith, 2006) could prevent children from attending to the global shape of the object and inhibit the formation of a prototype (Klinger & Dawson, 2001). Thus, autistic children might equally weigh different visual dimensions in their mental representations of new concepts, whereas TD children weigh shape more heavily.

Hartley and Allen (2014) has been recently replicated and extended, by Tovar et al. (2020). They suggested that the over-extension observed in the sorting task may not be due to equal weighting of features, but be attributed to a tendency to extend a label to any novel object. Tovar and colleagues also used a sorting task with objects matching the standard on shape or colour, and additionally included a novel item with nothing in common with the standard. This study replicated Hartley and Allen's (2014) finding that autistic children were more likely to include same-coloured items in the category, but exclude items they already know. However, the novel finding was that they were also more likely than chance to include the completely novel item. They suggest that some autistic children exhibit atypical categorisation, generalising on the basis of novelty rather than on the basis of a learned predictive cue like shape. It is possible that these children may rely more heavily on heuristics like mutual exclusivity, which is thought to be a strength in autism (de Marchena et al., 2011), to rule out objects that they already know the name of, and then have difficulty using constraints such as shape bias to form categories efficiently.

Despite the modest number of studies, whether autistic children have a shape bias is unclear. While the results of sorting tasks (Hartley & Allen, 2014; Tovar et al., 2020) and IPL tasks (Abdelaziz et al., 2018; Potrzeba et al., 2015; Tek et al., 2008) suggest that autistic children may lack a shape bias, there is still evidence of a preference for shape over other perceptual features when children are given a forced-choice pointing task (Field et al., 2016b; Tek et al., 2008). A sufficient explanation of shape bias in autism must account for both of these findings. Furthermore, the debates and disagreements that were prevalent in the 1990s and 2000s in the TD literature have led to inconsistencies in approaches and terminology in the autism field. Lack of clarity on the lexical nature has resulted in some studies defining a shape bias as higher shape-matching in the name condition compared to no name (Abdelaziz et al., 2018; Potrzeba et al., 2015; Tek et al., 2008), while others simply measured the proportion of shape-match choices and tested whether it was higher than chance (Field et al., 2016b). Field et al. (2016b) also suggested that shape-match choices in the no name condition as well as the name condition were evidence in favour of the shape-as-cue account. This lack of agreement is clearly problematic, as the TD children in Field et al. (2016b) who choose shape consistently regardless of label, and were classed as being consistent with the SAC account, would be categorised as showing no shape bias using Tek et al.'s (2008) method. This presents challenges when comparing and discussing these findings. It is perhaps unsurprising that there is disagreement concerning whether autistic children exhibit a shape bias when there is disagreement on how to define and measure shape bias. However, by considering the role of task demands and how they may interact to influence the shape bias, we may be able to account for these different findings under a unified, dynamic approach.

1.5 Explaining a shape bias 'deficit'

In the small amount of literature currently available there is reasonable agreement that the shape bias is weaker in autism (Abdelaziz et al., 2018; Field et al., 2016b; Hartley & Allen, 2014; Potrzeba et al., 2015; Tek et al., 2008; Tovar et al., 2020) and some suggestions that autistic children may attend to different object properties, such as function (Field et al., 2016a). However, as the extent and nature of differences are inconsistent across studies, there is more speculation about the source of the difficulty than there is evidence. Here, we consider some of the proposed explanations to determine which are viable candidates for further investigation.

1.5.1 Vocabulary size

Shape bias research in TD populations suggests that vocabulary size is an important factor in predicting when a shape bias is likely to be detected for English-speaking children. Specifically, between 50 to 150 words need to be known, and the majority of these words are likely to be count nouns that can be classified by their shape (Samuelson & Smith, 1999). Furthermore, the structure of children's early vocabularies predicts how well they can learn novel words (Perry et al., 2016); children with higher numbers of shape-based nouns in their vocabulary show better memory for the features and names of new objects. This suggests that it is through repeated experience of words that can be generalised by shape that children learn to give shape preferential attention. Gershkoff-Stowe and Smith (2004) discuss the possibility that a shape bias may contribute to the rate of word learning: attending to shape helps children learn words faster. Children can clearly acquire new words without this bias, however, the process may be more demanding on cognitive resources. However, as seen in Tek et al. (2008), vocabulary size was not a good predictor of a shape bias for autistic children. Similarly, in Field et al. (2016b), what should be a sufficient vocabulary size for TD children did not predict a shape bias in autism. However, the children with the highest vocabulary scores exhibited the strongest shape bias. This suggests that vocabulary is still important, however, this could be attributed to differences in vocabulary structure (e.g. the proportion of shape based count nouns in the vocabulary) (Perry & Samuelson, 2011; Samuelson & Smith, 1999), or the need for more word knowledge than expected to generalise the shape rule. It is difficult to judge the trajectory of shape bias development from Tek et al. (2008). Although the autistic and TD groups were matched on vocabulary at visit 1, by visit 4 12-months later this was no longer the case, with some children in the autism group still having no words at all in their productive vocabulary. So, it is possible that if the autistic children had been compared with a TD group that was still matched on receptive vocabulary, then there could have been less difference between the groups. It could be the case that the shape bias is merely delayed in autism rather than absent. It is, therefore, beneficial to consider the vocabulary size, as well as participants' overall expressive and receptive vocabulary scores.

1.5.2 Atypical prototype formation.

Much of the existing literature suggests a potential link between weak shape bias for categorisation and atypical prototype formation in autism (e.g. Field et al., 2016b; Hartley & Allen, 2014; Tek et al., 2008). The rationale for this link is the finding by Klinger and Dawson (2001) that autistic children who were taught a novel category were able to make rule-based judgments about category membership (e.g. a category containing things that have long legs), whereas they had more difficulty than typically developing children in identifying a previously unseen example that was a constructed 'average' of the rest of the set. While this suggests that autistic children may have a different mechanism for category formation, which could be linked to difficulties with perceptual categorisation, it is unlikely that poor prototype formation could *cause* a weak shape bias. As it is typical to test shape bias with exposure to a single example of the category, followed by immediate presentation of the test items (either immediately after or at the same time), it is questionable whether there is any opportunity for a prototype to be formed. If there are correlations between categorisation using prototypes and shape bias it is possible that they share a common underlying cause, rather than weakness in one causing the other. For this reason, poor prototype formation appears to be a weaker explanation for generalisation issues in the current context.

1.5.3 Weak central coherence

Autistic individuals tend to show an advantage for local feature processing over global features, a characteristic of autism described as 'Weak Central Coherence' (WCC) by Happé and Frith (2006). In contemporary explanations, autistic individuals may have an advantage over TD individuals for processing local details (Booth & Happé, 2018; Happé & Booth, 2008; Happé & Frith, 2006; Plaisted et al., 1999), and they may have an automatic attentional bias towards local elements (Plaisted et al., 1999). For tasks that require global attention and integration, autistic participants may need additional processing time (Booth & Happé, 2018; Van der Hallen et al., 2015), however there is evidence that attention can be directed to the global level when it is a clear requirement of the task (Plaisted et al., 1999).

This local attentional bias has been suggested as an explanation for shape bias deficits (e.g. Field et al., 2016b; Hartley & Allen, 2014), as autistic children may be inclined to focus on small parts of objects at the expense of overall shape. Evidence of attention to local details over global form in autism comes in various forms, such a performance on the Navon hierarchical figures task, which involves grouping together individual elements to perceive a whole, gestalt image (e.g. a letter 'H' made up of small 's's) (Navon, 1977). While TD children are generally faster to respond to the global form and are not prone to inference from the local letters, autistic children show an advantage for responding to the letter at the local level (Plaisted et al., 1999). However, Plaisted and colleagues also found that children are able to shift their attention from the local to the global level when directed to do so. Field et al. (2016b) proposed that this could form an effective intervention if WCC is found to play a role in shape bias difficulties, as giving children instructions on where to attend during word learning could provide support. Because this has not yet been tested as an explanation, and it has the potential to inform beneficial interventions, this is a priority for future investigation.

1.5.4 Theory of Mind

Another explanation for weaker shape bias in autism concerns understanding speakers' referential intent, or as Bloom (2003) describes it, the 'Mindreading' needed for word learning and communication. Of particular relevance to the SAC account of shape bias - where children are engaged in learning the meaning of words - understanding what a speaker intends with a word can be important. As demonstrated by Abdelaziz et al. (2018), engaging in joint attention may be a good predictor of a shape bias. Social cues could provide one contributing source of useful information for children to draw on during word learning. However, this finding does not necessarily reflect a social understanding of referential intent, as joint attention can also support attentional learning. For instance, autistic children are able to use social feedback to guide attention for word learning and generalisation (Hartley et al., 2019); whether the ability to use social feedback as an additional attentional cue implies an ability to infer what the speaker is intending to communicate is unclear. Rather, that ability to infer meaning may instead be linked to the concept of Theory of Mind.

Theory of Mind (ToM) involves understanding the internal state of another person as being different from one's own, and is commonly cited as a difficulty in autism (Baron-Cohen, 2000). ToM difficulties may account for a range of findings such as performance on false-belief tasks (Baron-Cohen, 1989) and differences understanding pragmatics (Colle et al., 2008). It is claimed that ToM could also play a role in word learning as children need to understand what a speaker is saying and make inferences about their state of mind and communicative intent when they speak (Bloom, 2000). In the ALA, difficulties with ToM need not have a direct impact on children's ability to recognise associations between labels, shapes and categories, and to use perceptual cues for generalisation. Put differently, children do not need to know what someone else means in order to notice a link between shapes and names.

The importance of ToM for word learning may therefore be overstated. For instance, Bloom (2000) argues that ToM underlies mutual exclusivity behaviour as children infer that the speaker is unlikely to be referring to the item that has a known different name. However, autistic children commonly do well on tasks of mutual exclusivity (de Marchena et al., 2011; Hartley et al., 2019), and recent evidence suggests that ME-like behaviour can arise from competition between multiple novel objects that children do not know names for (McMurray et al., 2012; Twomey et al., 2016). Hence, there are other avenues to word learning available to children that are not wholly dependent on inferring the intent of a speaker, and this may offer an additional source of information. For the current research we instead focus on explanations that could disrupt the link between perceptual features and labels and consequently impact early associative learning, however, social cues are a factor to consider once we have a better understanding of fundamental processes.

1.6 Methodological differences in shape bias investigation

Previous shape bias research has used a variety of task types, which the Dynamic Systems Account (Samuelson et al., 2009, 2013; Samuelson & Horst, 2007, 2008) proposes will affect how conceptual and perceptual information interact, potentially eliciting a shape bias given the right circumstances. In studies investigating shape bias in autism, different tasks have been used interchangeably on the assumption that because they are all suitable for investigating shape bias in typical development, they will also provide equivalent measures for autistic children. In the current research we explore whether these task variations can explain the differing shape bias findings for autistic children, due to unexpected demands in methodological choices that may disproportionately affect them.

1.6.1 Visibility of the stimuli: 'online' vs. 'offline'

Tek et al. (2008) used two tasks with a potentially important difference. In their intermodal preferential looking paradigm the standard appeared on screen twice for training, but it did not remain visible when the test objects were on screen, so generalisation had to be performed from the memory of the exemplar. Conversely, during the pointing task, the standard was visible on the table while the test objects were presented to children, so comparisons could be made directly rather than from memory. This 'online' version of the shape bias task was also used by Field et al. (2016b), who found that autistic children showed a shape bias. Furthermore, preference for shape was stronger when the objects were labelled, an effect consistent with TD children's performance (e.g. Landau et al., 1988), and was weaker in children with the lowest vocabulary scores. Most research exploring shape bias uses 'online' type tasks (e.g. Jones et al., 1991; Landau et al., 1988; Smith et al., 2002), whereas 'offline' tasks are more common for testing category formation and label extension in younger children using novelty preference looking tasks (e.g. Althaus & Plunkett, 2015; Hupp, 2008; Quinn et al., 1993). However, novelty preference tasks usually involve a long familarisation phase, (e.g 45 seconds in Hupp (2008)), whereas the Tek's implementation of the IPL paradigm uses two 4-second presentations of the standard. As a result, children would be reliant on working memory representations to complete this task.

There is evidence for differences in visual and spatial working memory (WM) in autism, with autistic children and adults performing weaker on some WM tasks (Funabiki & Shiwa, 2018; Zhang et al., 2020). However, these differences may be attributed to the time given to 'memorise' the subject of the task (Funabiki & Shiwa, 2018). Processing time has been shown to be important in other areas, for instance, autistic children perform just as accurately as TD children on certain word learning tasks, just at a slower rate (Hartley et al., 2020). 'Offline' novel noun generalisation (NNG) tasks may disadvantage autistic children by giving insufficient time to process the standard, and hence weaker performance on the IPL task could reflect differences in processing speed and working memory rather than differences in shape bias.

Therefore, one aim of this thesis is to investigate whether working memory dependent generalisation presents more of a challenge for autistic children than their TD peers. While memory-based shape bias tasks are less frequently used in shape bias research they are more reflective of real-life generalisation tasks where another example is often not in sight when children are asked to find objects by name. It is also possible that memory-based tasks will help identify whether more able autistic children are 'hacking out' a solution to shape bias tasks by effortfully directing their attention to the shape. If the same children show a shape bias during the online direct comparison task but not in the offline task, this may suggest they have an intellectual understanding that the shape is useful, but that automatic attentional processes are not supporting the extraction of these details for the memory-based task.

1.6.2 Yes/No vs. Forced-choice tasks

While more recent shape bias research utilises forced-choice tasks, earlier research used yes/no versions. In the yes/no variation of the task, children are presented with test items sequentially and asked to make a decision on each one. This may be done as a label extension task (e.g. 'Is this a dax?') or without labels (e.g. "Is this like this one?") (Landau et al., 1988, 1992, 1998; Smith et al., 1992, 1996). The benefit of the yes/no task is that children are less restricted on their generalisation decisions. They could exclusively accept shape-match items, or they could also accept items with other perceptual similarities, allowing them to form broader categories. In effect, the yes/no task asks children to identify if a test item is a 'good enough' example of the category, so a test item that is a slightly different shape but made from the same material may be accepted as a category member. The drawback of this task is that younger children, particularly below the age of 2 years, can find this more difficult (Landau et al., 1988; Samuelson et al., 2008). Additionally, young children can be biased towards giving 'yes' responses to yes/no questions (Fritzley & Lee, 2003).

For the forced-choice task, children are typically shown two or three items that vary from the standard in a systematic way (e.g. one shape-match, one colour-match, and one texture-match) and are asked a variation of "Which of these is a dax?" (e.g. Booth et al., 2005; Diesendruck & Bloom, 2003; Field et al., 2016b; Landau et al., 1988). The implicit goal of the task is to pick the 'best' example of the category. While children are not explicitly told to do this, the setup of the task with two or three items to choose from gives the impression that there is one 'correct' answer, even if children would accept all the items given the freedom to choose. By restricting the choice to one item, the forced-choice task is less useful for exploring feature combinations for generalisation, such as the importance of both shape and texture for living things (Jones et al., 1991). However, studies that have used both versions of the task in typically developing children have found that they prioritise shape in both (Booth et al., 2005; Landau et al., 1988). Samuelson et al. (2009) suggest that the forced-choice can amplify relatively small differences in feature weightings, so children might judge both shape and texture to be relevant to the category, but shape only needs to have a slightly higher priority over texture to ‘win’ against a texture-match competitor. Arguably, the forced-choice task is still useful for identifying which features children see as the *most* important and enables the shape bias to be detected in younger children who may struggles to understand the yes/no task.

Although both forced-choice and yes/no versions of NNG tasks are effective for observing a shape bias in typical development, the same may not be true for autistic children. Of particular concern is that inconsistent findings in previous research may be explained by differences in the task demands. Evidence that autistic children prioritise shape in generalisation has been generated by forced-choice NNG tasks (Field et al., 2016b; Hartley et al., 2019, 2020; Tek et al., 2008). Conversely, studies showing that autistic children do not exhibit a shape bias used alternative tasks, such as sequential sorting, in which children over-extended the category to non-shape-match items (Hartley & Allen, 2014; Tovar et al., 2020), and IPL, in which there was no clear difference in looking times between shape-match and non-shape match items (Abdelaziz et al., 2018; Potrzeba et al., 2015; Tek et al., 2008). While the IPL task is similar to a forced-choice task insofar as children can only look at one candidate at a time, over the course of each trial they may allocate their looking time between them equally, hence 'no choice' is a valid outcome. The discrepancy in findings between methods could be because the forced-choice task is effective at identifying the most important feature, whereas the flexibility of the alternative tasks could be prone to more interference for autistic children. The existing research into shape bias in autism has not confirmed if these tasks produce equivalent results for this population, and so investigating this possibility is a core aim of papers 1 and 2 in this thesis.

1.6.3 Magnitude of difference between the stimuli

The extensive variability in stimuli utilised in shape bias research can have implications for how the bias is exhibited. It is not sufficient to simply give children the option to generalise to a shape-match item without considering what kind of decisions or strategies could lead children to select alternatives. For instance, if stimuli are very similar in appearance, a small change in shape could be accepted if children judged it to be a good overall match (e.g. a ‘drinking glass’ category may tolerate some variation in shape as long as choice items are made out of glass). However, distractor items that are extremely different from a standard would not represent good overall matches, and choosing them would suggest that a child consider characteristics other than shape to be indicative of category membership. While studies have examined the effects of stimuli differences on shape bias (Landau et al., 1988; Smith et al., 1992; Tek et al., 2012), many others do not consider the implications of those design choices.

Previous studies differ in terms of magnitude of variation between standard and test items. At one end of the scale, early shape bias research used stimuli that changed in a single characteristic, while everything else about the object remained constant (Landau et al., 1988, 1992; Smith et al., 1992). For instance, the shape-change test item matched the standard in colour, size, and texture, and deviated only partially from the original in shape. By changing only one dimension of the test objects, these studies measured how flexible children’s label extension processes were and how accepting of each change they would be. If children accepted that ‘daxes’ could be different textures and sizes, but not different shapes, it could be inferred that shape was judged to be the most important criterion for category membership. Thus, changing only one feature at a time allowed researchers to build up a detailed picture of the relative importance of different perceptual features.

At the other end of this scale are studies where the test stimuli have only one variable in common with the standard, and differ on more than one dimension (Diesendruck & Bloom, 2003; Field et al., 2016b; Graham & Diesendruck, 2010). For instance, Field et al. (2016b) used different objects such as lemon squeezers, sink stoppers and shoe horns to make up their sets, and were covered them in brightly coloured materials (e.g. pink tissue paper or orange sticky tape) to manipulate the appearance of colour and texture. In these cases, the colour and texture-match items differed greatly in shape from the standard, and so label extension to a distractor would suggest that the matching feature was perceived as more relevant to category membership than shape was.

There are also studies that fall between these two extremes, where shape change distractors are neither a minor distortion of the original nor a complete contrast. For example, in Tek et al. (2008), both the standard and the distractor consisted of wooden blocks with two prongs, though the objects' orientation and components differed in appearance. In cases like this it is difficult to say if the selection of distractors indicates a flexible category, where stimuli are accepted because there are more overall similarities (indicating belonging to the same superordinate category), or whether different cues are being used to inform label extension decisions.

The consequences of stimuli variations on shape bias have been examined in previous research. Landau et al. (1988) systematically varied the magnitude of shape changes between standard and test stimuli and found that two- and three-year-old children were less likely to generalise as shape changes increased. Smith et al. (1992) found that young children would accept small differences in shape as still being ‘similar’ in an unlabelled condition, though were far less likely to generalise a label. When shape was completely different, this effect for labelled items reduced even further, from 29% to only 12% acceptance of the object. Tek et al. (2012) also investigated the impact of shape similarity on shape bias using their IPL task, and found that 2-year-olds’ preference for the shape match item was stronger when distractor shapes were less similar to the standard. Furthermore, when distractor items were highly similar to the standard, children looked more at the ‘overall’ match when it was named.

Variability in stimuli may be referred to using several terms, including ‘magnitude of change’, ‘high/low similarity’, or ‘high/low variability’. In the current research we classify stimuli as ‘high contrast’ or ‘low contrast’. When objects have low contrast, picking a non-shape-match distractor item could be interpreted as an ‘overall’ match. That is, although shape might be different, it is not so different that it is implausible that it could belong to the same broad category as the standard. In these cases, accepting a distractor item may happen if the child believes that multiple features are important to category membership (e.g. they believe that both shape and texture are relevant to the category). Therefore, choosing an item that is a “close enough” match on shape that preserves texture might be the best example of the category.

Conversely, we consider high contrast distractor objects as having nothing in common with the standard. The shape, or more specifically the outer contour of the object, is completely different, so generalisation to a distractor item does not suggest an overall match. Instead, non-shape-match choices suggest that the label has been mapped to a feature other than shape. High contrast stimuli were initially more common in studies that hypothesised that children might find another feature useful (e.g. when classifying mass nouns compared to count nouns; Samuelson & Smith, 1999). However, there are also instances where highly different stimuli have been employed by research focusing only on count nouns (e.g. Field et al., 2016b; Hartley & Allen, 2014).

This variation in stimuli is particularly relevant in relation to the different task types. As yes/no tasks are not competitive, children are more accepting of small variations in shape (Landau et al., 1988; Smith et al., 1992). Given a forced-choice task, however, an object with the most similar shape may be preferred as the *best* example regardless of how different the distractors are. This variability has only been considered for TD children, however, research with autistic children has employed a mix of high contrast (Field et al., 2016b; Hartley & Allen, 2014; Tovar et al., 2020) and low contrast (Potrzeba et al., 2015; Tek et al., 2008) stimuli sets. In this thesis we explore how children generalise novel nouns for both high contrast (Paper 1) and low contrast (Paper 2) stimuli sets, using both forced-choice and yes/no task variants.

1.6.4 Audio stimuli

Another difference between studies concerns how questions are phrased to children, particularly if the study includes 'no label' or 'no name' conditions. For labelled items, word positioning in a sentence can provide clues that guide children's beliefs about whether it is a noun, verb, or adjective (Landau et al., 1992; Smith et al., 1992). However, variation in how questions in the no label condition are phrased also places constraints on how children interpret the task. For instance, Diesendruck and Bloom (2003) investigated the effect of different phrasing on shape bias by asking children which test item was the 'same kind' and which one 'goes with', in addition to a naming condition (e.g. "which is also a Patoo?"). They found that 2- and 3-year-old children chose shape-match items in the 'name' and 'kind' conditions, but not when they were asked what 'goes with'. This phrasing is important as it needs to be clear to children that the goal is to find an object that belongs to the same category in order to determine if shape is being used as a cue for categorisation, otherwise the lack of shape bias in a task may be attributed to differing interpretation of the goal rather than reduced attention to shape. Asking children to find ‘one that goes with’ can invite thematic links (e.g. Imai et al., 1994), for instance. ‘Dogs’ and ‘leads’ can ‘go together’, but they do not belong to the same category, and so no shape bias would be expected.

The choice of phrasing used for the 'no name' trials in research with autistic children has deviated from common phrasing in TD research. The majority of autism research uses the phrase “which one looks the same?” (Abdelaziz et al., 2018, p. 2689; Potrzeba et al., 2015, p. 6; Tecoulesco et al., 2021; Tek et al., 2008, p. 212). Alternatively, “can you give me the other one?” has also been used (Field et al., 2016b, p. 3199). The appeal of the ‘kind of’ phrasing is that this frames the task as one of categorisation, whereas the dominant phrasing in autism research asks about the appearance of the object. This may present an issue as there are cases in children’s word learning where things ‘look the same’ without being the same, and as children begin to develop conceptual categories it is possible these questions would be interpreted quite differently. For instance, children will not generalise a novel label for a round cookie to a round CD even though they are the same shape (Cimpian & Markman, 2005), however, a direction to ‘find one that looks the same’ could be taken as an instruction to find perceptual similarities regardless of whether things are the same ‘kind’. The alternative phrasing ‘give me the other one’ is also potentially ambiguous. While ‘the other one’ could refer to the “other one that is the same as the standard”, it could also be a statement used to indicate a contrast (e.g. “not the one like this, the *other* one”).

In many studies, TD children still exhibit a shape bias despite variations in questions. However, we should not presume that autistic children interpret different instructions the same way. In this thesis, the phrasing used is as simple as possible to reduce the risk of alternative interpretations. The phrasing "Is this another one?" (for yes/no task) or "can you find another one?" (for forced choice task) (e.g. Booth & Waxman, 2002a) have been chosen as neutral phrases that do not specifically refer to either 'kinds' or the perceptual appearance of objects. Additionally, the lexical form of the sentence permits the 'one' to be replaced with a name for the label condition while retaining consistency in the rest of the audio (e.g. "Is this another vink?" vs. "Is this another one?").

1.7 About the current research

This literature review has highlighted that a wealth of knowledge about how TD children can acquire, and use, attentional learning strategies flexibly for category name learning has arisen from thorough investigation of the shape bias. By questioning discrepancies in findings and the impact of researcher choices, such as the task type, stimuli, and the phrasing of instructions, may affect whether children exhibit a shape bias in a given task. Explanations like the Dynamic Systems Account show an appreciation of how a child's response is not the result of a binary bias that a child may have unlocked or not, but is the product of processes that are changing moment-by-moment (Kucker et al., 2019; Samuelson et al., 2009, 2013). When viewed as an available process in a dynamic system, it makes sense to ask under what circumstances a shape bias is likely to be activated rather than whether children develop a shape bias.

The current research seeks to update our understanding and conceptualisation of shape bias in autism with insights from the dynamic approach. Instead of asking if autistic children have a shape bias, as the first investigative approach did (Tek et al., 2008), this thesis explores how altering aspects of experimental design may impact whether shape bias is utilised in the moment. In other words, *when* do autistic children exhibit a shape bias? Additionally, if autistic and TD children exhibit differences in shape bias under different task demands, those differences could be revealing about underlying causes of word learning difficulty. The research in this thesis consists of 5 experiments, described across 3 studies, each investigating the impact of a methodological variation on the tendency to prioritise shape when making quick categorisation judgements about novel artifacts.

Study 1, which forms the first paper of this thesis, investigates the role of shape bias in online and offline categorisation with perceptually distinct 'high contrast' stimuli. This study included a forced choice task (experiment 1), in which autistic children have previously demonstrated a shape bias (Field et al., 2016b), and tested whether the visibility of the standard and the presence of labels influenced the shape bias differently for autistic and TD children. This study also tested the same variables with the same stimuli in a yes/no task (experiment 2). There is evidence, using high contrast stimuli, that autistic children may over-generalise to characteristics besides shape (Hartley & Allen, 2014). This research investigates whether this over-generalisation can be explained by the demands of the task by using the same stimuli in both a forced-choice and yes/no NNG task.

Study 2, as described in the second paper of the thesis, uses the same forced-choice (experiment 3) and yes/no (experiment 4) tasks but with perceptually similar 'low contrast' stimuli. While the highly different stimuli of Study 1 are intended to be less challenging and increase the likelihood of observing a shape bias, the 'distractor' items are too different to plausibly belong to the same category. Thus, in Study 2, the shapes are intended to be similar enough that distractor items could be reasonable 'overall' matches. With these low contrast stimuli, we investigated whether the visibility of the standard and the presence of labels influenced the strength of the shape bias when 'good enough' overall matches were alternative choices. Furthermore, we looked for population difference in the impact of these variables on children's generalisation choices.

The final paper of this thesis describes the fifth experiment, which tests whether a local attentional bias due to weak central coherence (Happé & Frith, 2006) may, explain a weaker shape bias in autism. Using a yes/no task with objects that could be classified by either global shape, a local feature, or both, we tested whether autistic children would be more likely to generalise by the local feature than TD children. We also examined if feature-based generalisation was influenced by including labels and varying the visibility of the standard for reference.

Data collection for all 5 experiments was completed at the same time as part of a unified project. The general task and stimuli details are described below, while the full procedure for each experiment is contained within the relevant papers.

1.7.1 Task design

This research utilises variations of forced-choice and yes/no versions of the NNG task. Though most classic shape bias research uses physical 3D objects, for this research we opted to use a touch-screen computer to administer all of the tasks. Despite this being a deviation from more common tasks, computer-based tasks can be easier and more motivating for autistic children to complete (Allen et al., 2016). Furthermore, the reduced social interaction needed to engage with the task could provide a greater opportunity to perform to their potential, as may be the case with other computer administered measures of cognitive processes (e.g. Ozonoff & Strayer, 2001). While the use of real objects with an adult speaker would arguably be closer to a real-world word learning scenario, the aim of this research is to establish whether autistic children can generalise using a shape bias when other competing demands are reduced. The results of this research will establish a baseline from which the impact of less optimal learning environments could be explored further. Secondly, using a screen to display the stimuli allows for computer-generated objects to be used instead of real ones, affording much greater control over how objects vary and match in appearance. Additionally, the ease of generating variations of stimuli allowed for multiple trials with each standard. While it is common for children to see only one shape-match, colour-match, and texture-match alternative for each standard, in this research children completed multiple trials for the same standard with different test items on each trial.

The experimental tasks were all designed using PsychoPy2 (Peirce, 2007), a python-based experiment handler that controlled counterbalancing, trial order, stimuli presentation, randomisation, timings, and recorded all participant choices directly in preparation for data analysis. Each of the 5 experiments was created as a mini-game within a single combined programme. This was to ensure consistency of the randomised name and object pairing for each participant if they took part in all 5 experiments. Thus, if a particular object was labelled a 'zain' in one experiment, it was not possible for a child to hear a different object being called a 'zain' on a different day. Each participant was set up with an individual file at the first session with their own counterbalanced order for completing the studies, and their own word-object pairings.

The forced-choice and yes/no task variants were designed to emulate in-person procedures as closely as possible. For the forced-choice task, children were presented with three objects that shared a characteristic with the standard: a shape-match, a colour-match, and a texture-match. The children had to touch one of the items to make a response. For the yes/no task, each item was displayed with response buttons on the screen: a green tick for 'yes' and a red cross for 'no'. We anticipated that this may be more challenging than giving a verbal or gestural yes/no response, so the procedure included training with the buttons and the option for children to give a verbal response instead if necessary.

1.7.2 Stimuli design

All stimuli were uniquely designed for this research using Blender: open source 3D graphical modelling software (Blender Development Team, 2017). Blender is suited to creating 3D images for animation and games, and can produce texture and lighting effects that look photorealistic. As such, it is ideal for creating stimuli where the colour, material, and shape need to be completely controlled by the researcher. All of the stimuli were created to specifications that have found strong shape bias in TD populations in previous research as follows. First, all stimuli base models resembled non-deformable artifacts made from geometric shapes that had been modified, manipulated, and combined to create novel objects. Objects that appear to be animate (Booth et al., 2005; Jones et al., 1991), non-solid (Soja et al., 1991), or deformable (Samuelson et al., 2008) can result in alternative attentional biases other than shape, which we did not wish to intentionally encourage. Next, 'materials' were created using Blender to emulate a variety of textures that could be applied to any base model: smooth metallic, plushy/furry, rough stone, wood, ceramic, marbled stone and metal foil, matte plastic. These were also generated in a variety of colours.

1.7.2.1 Stimuli for Studies 1 & 2

To investigate the role of online and offline categorisation, both the forced-choice task and the yes/no task use the same stimuli sets. In the forced-choice task, all stimuli in the set were used. The yes/no task used half of the set to limit the length of the experiment. Each set consists of one standard, 12 test items, and two memory check items. The test items each match the standard on one dimension only: shape (x4), colour (x4) or material (x4). As earlier research found no effect of changing size (Landau et al., 1988) unless the size variation was extreme (e.g. 2cm object vs. 2 ft tall object), the display size of each stimuli was held constant across test objects and trials.

1.7.2.1.1 High Contrast Stimuli (sets A-D)

The high contrast stimuli sets are similar in style to Field et al. (2016b): overall object shapes had no designed similarities and contrasting colours were used (as determined by a triadic colour scheme using a colour wheel: e.g. red-yellow-blue). Individual stimuli were coloured with a monochromatic colour scheme, with lighter and darker hues of the base colour used to create visually interesting items (see Figure 1). This high contrast set was intended to provide the optimal conditions for shape bias, with little ambiguity about the group membership. Tek et al. (2012) found that young children may prefer an 'overall match' when stimuli are similar, weakening the shape bias. Therefore, it is important to provide the opportunity for autistic children to demonstrate a shape bias under optimal conditions. These stimuli were used for Study 1, which included forced choice (experiment 1) and yes/no (experiment 2) variants of the NNG task using the same stimuli. The novel labels used with this set were *teep* or *foss*, both selected from the English Lexicon Project (Balota et al., 2007) as plausible English words.

Figure 1: High contrast stimuli sets (A-D) used in experiments 1 and 2.

1.7.2.1.2 Low Contrast Stimuli (sets E-H)

The low contrast sets were also designed to match the standard on one dimension of interest: shape, colour, or texture. An alternative option would have been to change one characteristic of the stimuli and hold the others constant, however, we wished the results of the different studies to be comparable Including a different number of variations would have made it difficult to determine if any contrasting findings were due to the magnitude or number of changes. Hence, we kept the number of changes constant throughout the research, and altered only the extent of the change. With the low contrast stimuli, the different shape distractors have some similarities to make it appear that they could plausibly belong to the same parent category as the standard. For each standard, the test objects were created from the same base shape (cylinder, cone, wedge and torus) and manipulated to introduce shape variation. The colour changes were also less extreme than in the high contrast condition, using an analogous colour scheme (see Figure 2).

The novel labels *vink* and *zain*, from the English Lexicon project (Balota et al., 2007) were paired with this these stimuli sets.

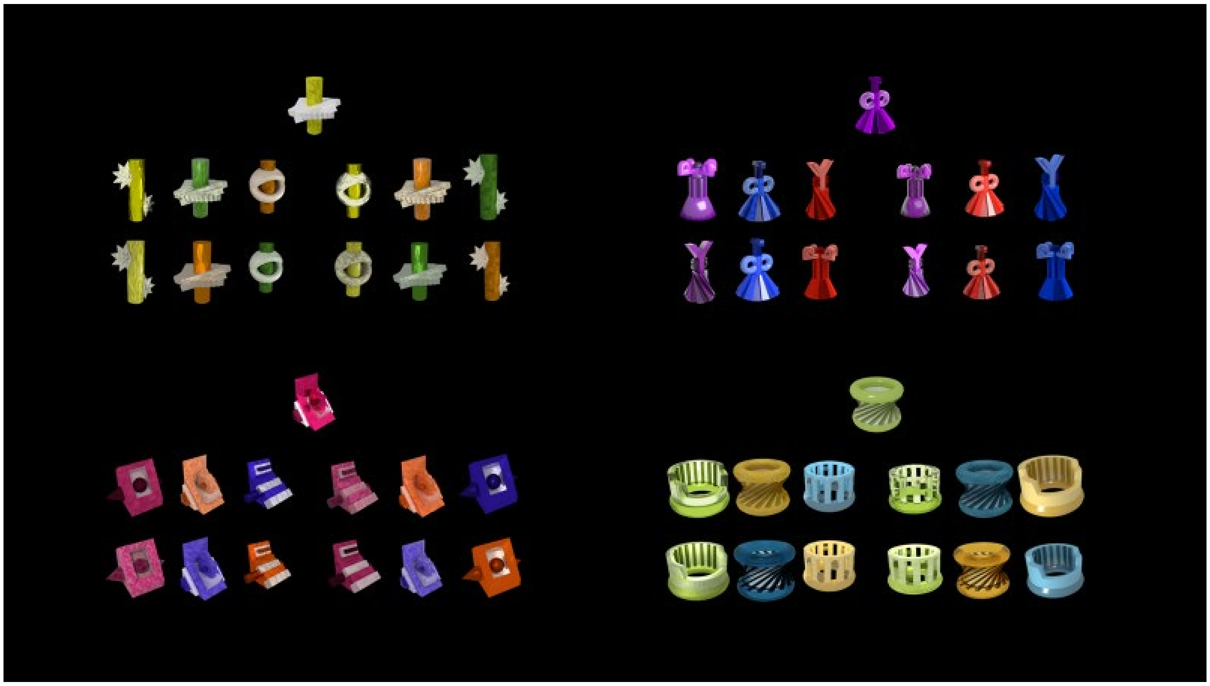


Figure 2: Low contrast stimuli sets (E-H) used for experiments 3 and 4

1.7.2.2 Stimuli for Study 3

To test the hypothesis that autistic children may show a weaker shape bias due to increased attention to irrelevant features of objects, four stimuli sets with a salient local detail were created. Each set consisted of a standard and seven test items. The standard consisted of a novel shaped object with a distinctive contrasting feature. Test objects differed from the standard could differ from the standard in shape, feature, colour, or any combination of the three. For simplicity, material was held constant across all the objects as a matt plastic-type appearance. For each stimuli set there were two features: one which appeared on the Standard and a contrasting feature. The features themselves differed in colour and shape to make them visually distinctive, and different features and colours were used for every condition to avoid cross-contamination. For these stimuli conditions, four test items were a colour match, four were a feature match, and four were a shape match, with each combination appearing once as shown in Figure 3. The novel words *tookie* and *numble* used for the label trials with these stimuli sets, and again were selected from the English Lexicon Project (Balota et al., 2007).

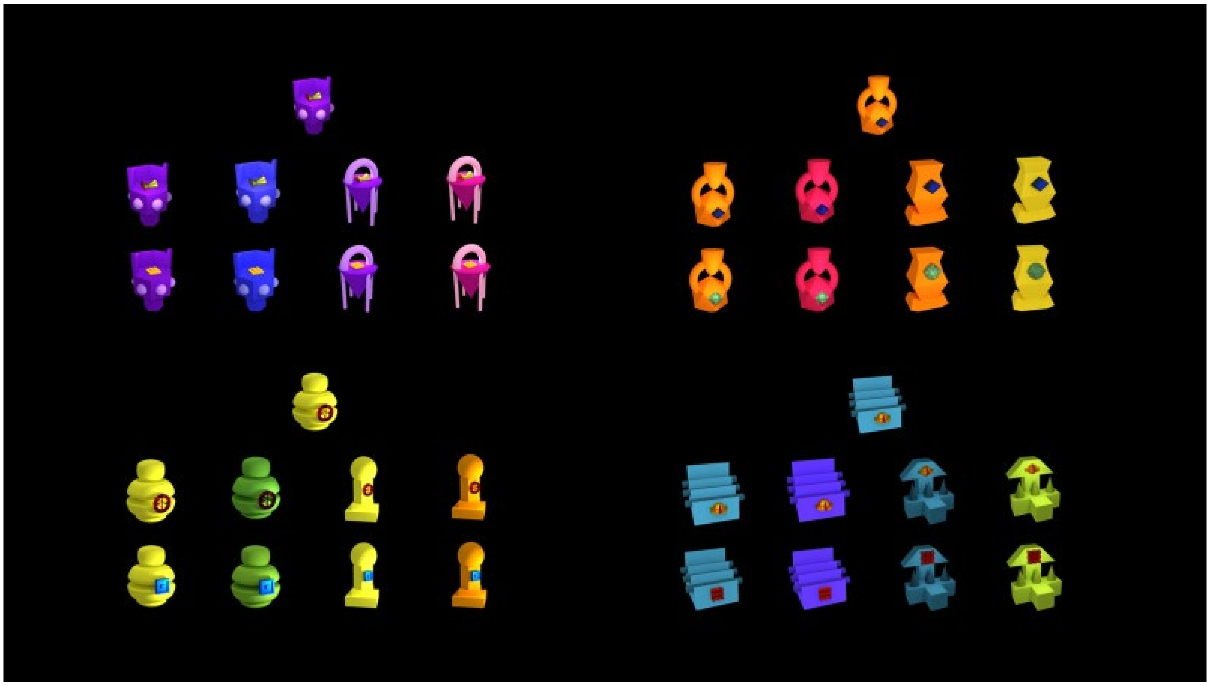


Figure 3: global/local feature stimuli sets (I-L) used in experiment 5

1.7.3 Aims of the research

This research aims to determine which conditions are optimum for a shape bias for autistic children, and which task demands may interfere with it. Throughout each of these experiments we test explicitly whether the requirement to remember the standard during a novel noun generalisation task disadvantages autistic children, and whether labels afford the same benefits as for TD children. By determining which experimental task demands make generalisation more challenging, we can begin to make predictions about the impact taken-for-granted aspects of real-world learning may have on autistic children, and the implications for their language acquisition.

Chapter 2: Task demands affect 'shape bias' for autistic children during novel noun generalisation

2.1 Introduction to Chapter 2

When Tek et al. (2008) first asked "do autistic children have a shape bias?" this appeared to be a simple question, however, it does not yet have a satisfactory answer. In this paper we address this question in relation to two different tasks. Inspired by the method and findings of Field et al. (2016b), this paper provides a starting point from which both autistic and TD participants could exhibit a shape bias by replicating ideal circumstances from previous research. The stimuli used were highly different, easily differentiated shapes, and created in contrasting colours. With these high contrast stimuli, presented in a forced-choice NNG with a visible standard for 'online' comparison, we establish a baseline for a strong shape bias in both autistic and typically developing participants. From there, manipulations to the standard visibility, the presence of labels, and the type of task used, suggest that some experimental task demands interfere with shape bias for autistic children. As a result, this paper shifts the topic of discussion from 'do autistic children have a shape bias?' to "*when* do autistic children *exhibit* a shape bias?".

Author contributions:

Leigh Keating: study design, data collection, statistical analysis, manuscript writing, review.

Calum Hartley: study design, review. *Katherine Twomey*: study design, review.

2.2 Abstract

Recent research suggests that autistic children may not prioritise object shape over other features when generalising novel nouns, as typically developing (TD) children do. The absence of this 'shape bias' may impact children's ability to efficiently add new count nouns to their vocabulary. However, evidence for a shape bias deficit in autism comes from experimental tasks that place differing demands on TD and autistic children, interfering with measurements of underlying word learning ability. This research investigated whether autistic and TD children exhibited a shape bias in forced-choice (Experiment 1) and yes/no (Experiment 2) variations of a novel noun generalisation (NNG) task. On some trials, a standard was visible for reference, while on other trials the standard had to be remembered briefly. A mixed logit model revealed a shape bias for both groups in the forced-choice task. However, in the yes/no task, autistic children were significantly more likely to accept texture and colour-match items that where a different shape. These findings suggest that both TD and autistic children can use shape as a cue for category inclusion, however, only TD children reliably use shape difference to inform category exclusion decisions.

2.3 Introduction

From as early as 3 months old, infants can use perceptual regularities between examples as the basis of their first categories (Quinn et al., 1993). During their first year of life, children learn to link labels to these categories (Fulkerson & Waxman, 2007), and use those labels to define category boundaries (Plunkett et al., 2008). Through these early word learning experiences, and interactions with their caregivers, typically developing (TD) infants begin to recognise the most frequent 'rules' that govern label-object/category mapping. Once recognized, they can apply these rules in novel word encounters, resulting in more efficient word learning (Samuelson & Smith, 1999; Smith et al., 2002). One such word learning heuristic is known as the shape bias (Landau et al., 1988). Children who exhibit a shape bias will generalise a newly learned count noun to other items that are the same shape, and ignore features that are less predictive of category membership such as size or colour. In this way, a shape bias can support vocabulary growth by providing a word learning strategy that can operate from a single example. A growing body of research into shape bias in autism, however, suggests that autistic children may not generalise nouns on the basis of shape to the same degree that TD children do (Abdelaziz et al., 2018; Field et al., 2016b; Hartley & Allen, 2014; Potrzeba et al., 2015; Tek et al., 2008; Tovar et al., 2020).

Autism is characterised by difficulties with social communication (World Health Organisation [ICD-11], 2018), and while research estimates a good proportion of children with this condition can use language fluently, approximately 30% have no spoken language at 9 years old (Anderson et al., 2007). Given that the emergence of shape bias has been associated with the onset of the 'vocabulary spurt' in TD children (Gershkoff-Stowe & Smith, 2004), a shape bias deficit could disrupt the trajectory of language development in autism. However, methodological differences in the way the shape bias is investigated may leave open the possibility that cognitive differences other than shape-based generalisation explain why autistic and TD children respond differently. Much of the recent research with autism uses ‘offline’ generalisation, where the example of the category is presented beforehand and must be remembered during generalisation. In this research we investigate whether the addition of a working memory demand in a novel noun generalisation task contributes to a ‘weaker’ shape bias in autism in two commonly used variants of the NNG task: forced-choice and yes/no. If the strength of the shape bias interacts with experimental manipulations in different ways for TD and autistic participants, this would be indicative of differences in using the shape bias for the task at hand rather than a deficit. Understanding the situations in which autistic children do, or do not, use a shape bias for word learning is an important step towards providing optimum teaching conditions.

In typical development, it is expected that children will exhibit the shape bias in word learning tasks by time they have between 50 and 150 count nouns in their vocabularies (Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999). However, there is evidence of perceptual category learning much earlier in development. From as early as three months old, infants appear to form categories based on statistical regularities in perceptual features that include previously unseen instances (e.g. Behl-Chadha, 1996; Quinn et al., 1993). By the time infants are 10-months-old, there is evidence of labels supporting categorisation as infants begin to categorise items together that share a name despite perceptual differences (Plunkett et al., 2008). Labels also appear to play a role in attentional learning by drawing 12-month-old infants’ attention towards the common perceptual features of items given the same label (Althaus & Plunkett, 2015). While these studies of early attentional learning involve infants being familiarised with many examples of the new category they are learning, allowing them to notice which features objects with the same name have in common, this early experience may provide the foundation from which a shape bias can develop.

For the majority of infants' early categories, the common perceptual feature is the object shape (Samuelson & Smith, 1999). Once children have acquired a sufficiently large vocabulary, they can begin to make higher order generalisations by attending to shape from the first encounter with a novel count noun (Smith et al., 2002). By the time TD children have between 50 and 150 count nouns in their vocabularies, they are likely to exhibit a shape bias in word learning tasks (Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999). As children learn to exploit attention to shape as a strategy, the rate of vocabulary acquisition can increase (Smith et al., 2002), and in turn a larger vocabulary leads to a stronger shape bias, suggesting a bidirectional relationship between the two (Gershkoff-Stowe & Smith, 2004). In other words, the development of shape bias can be thought of as part of the process of learning how to learn novel nouns.

A typical shape bias study involves the use of a novel noun generalisation (NNG) task (see Landau et al., 1988). Participants are introduced to a novel object, known as the 'standard', labelled with a plausible nonsense word (e.g. 'dax'). They are then shown new objects which differ from the standard in a controlled way (e.g. size, shape, colour or texture) and asked to indicate whether they are also a 'dax'. When the aim is to investigate whether the presence of a label affects the strength of the shape bias (e.g. Landau et al., 1988), conditions with non-lexical generalisation are used which ask participants if an item is 'the same kind' or 'goes with' with the standard. The NNG task has two commonly used variants: the 'forced-choice' task and the 'yes/no' task. In the forced-choice task, two or more test objects are presented at the same time and participants are asked to generalise the new category to one of the available options (e.g.

Booth et al., 2005; Diesendruck & Bloom, 2003; Landau et al., 1988). In the yes/no task, items are presented one at a time and participants are asked if each one is an example of the same category as the standard without any competitors being present (Landau et al., 1988, 1992, 1998; Smith et al., 1992, 1996).

A feature that both forced-choice and yes/no NNG tasks have in common is that they both typically use ‘online’ paradigms, where the standard remains visible while children are making their generalisation decisions. This allows for real-time comparisons to be made based on currently available perceptual information, rather than what children can remember about the object. Studies investigating the role of memory on shape bias suggest that learning to remember object shape is also related to vocabulary, and may come later in development. Perry et al. (2016) found that 24-month-olds had difficulty remembering object shapes when the standard was not visible, with no difference between testing immediately or after a 5 minute delay. However, feature memory was related to vocabulary structure, and children with a larger number of shape-based count nouns in their vocabularies were better able to remember shape. Vlach (2016) found that TD children only demonstrate a ‘memory bias’ for shape after a shape bias in ‘online’ NNG tasks had been firmly established, and is not common before children are 4-years-old. This suggests that, early in the process of shape bias development, children benefit from being able to use online comparisons when they have only a single example of the category to generalise from. While attention may be drawn to the most useful feature, this is not enough to remember it in sufficient detail when it is removed from view. However, as shape bias continues to strengthen alongside a growing vocabulary, object memory also becomes more efficient as children can encode the most relevant features and perform ‘offline’ comparisons.

Prioritising attention and memory of object shape may be highly useful for young children's word learning, as object shape provides an ideal trade-off between 'cue validity' and 'cognitive economy' in basic level categories (Rosch et al., 1976). In other words, shape may be both simple to mentally represent and be a good predictor of category membership for a high proportion of words in a young child's vocabulary, so attention to this property over others is an efficient use of cognitive resources. Therefore, a delayed, divergent, or even absent shape bias could have far reaching implications for a child's language development.

The first investigation into shape bias in autism came from Tek et al (2008). Two and 3-year-old autistic children, and TD children matched on receptive and expressive vocabulary, were administered two different tasks designed to detect the presence of a shape bias. One task used an intermodal preferential looking paradigm (IPL): children watched a video display across two screens which introduced them to a novel object, presented with or without a label. This was followed by two test objects: a shape-match for the original and a colour/pattern match with a differing shape. Children's responses were measured by coding which screen they looked at when asked 'which one looks the same?' or 'where's the dax?' The intention was that this task supplied a more implicit measure of shape bias than NNG tasks do, as children were not required to make a deliberate, conscious decision between the two. Children could believe that both were plausible examples, and so record similar looking times for both test items. The second task was a pointing task which followed the same procedure as common forced-choice NNG, where the children were presented with the physical versions of the same objects and instructed to point to their choice.

The children repeated both tasks every four months for total of four visits. For the IPL task, they found no evidence of a shape bias for the autistic children, who did not look at the shape match object for any longer than the other test object, and no benefit for labelling the items. This was despite their vocabulary scores indicating a vocabulary size in excess of 100 count nouns by the fourth visit. In contrast, the TD participants looked at the shape-match item for a significantly higher proportion of the trial, and only when the item was labelled. In the forced-choice task, however, autistic children pointed to the shape match at above-chance levels regardless of whether there was a label or not. The authors concluded that the autistic participants did not demonstrate a shape bias in the IPL task. As autistic children did not differ when measuring other lexical heuristics with this method, such as a noun bias (Swensen et al., 2007; Tek et al., 2008), this is taken as evidence that use of shape bias as a word learning strategy could be the source of their language learning difficulties. Also, the finding that autistic children preferred the shape-match in the pointing task, regardless of whether the standard was labelled, suggests that they may be deliberately choosing the shape without benefiting from an automatic attentional processes. The absence of the enhanced attention to shape for labelled items observed in typical development (Landau et al., 1998; Tek et al., 2008) is also consistent with this explanation.

However, Field et al. (2016b) also conducted a pointing task, which yielded conflicting findings on the effect of labels. Using a physical, forced-choice task with three items to choose from (colour-match, texture-match, and shape-match to the standard), they found that autistic children chose the shape-match item most often, as was the case in Tek et al.'s (2008) pointing task. However, when contrasting labelled items with unlabelled items, Field et al. (2016b) reported autistic children with higher receptive vocabularies exhibited a stronger shape bias on labelled trials. Conversely, they found no effect of label for the TD group. Tek et al. (2008) proposed that their IPL task tapped into automatic attention, so the lack of a preference suggested no such attention, while the success in the pointing task was due to a conscious decision. However, if autistic children were relying on conscious decision to apply a shape bias, they should also be able to do so in a sequential yes/no task that requires an explicit response.

In Hartley and Allen's (2014) shape bias study, autistic and TD children responded to each item individually as they would in a yes/no task. Participants were shown pictures of novel objects and completed a sorting task. Objects that were judged to be the same category (e.g. 'blickets') were to be put into one box and items that were judged to be out of category went into an 'other things' box. In this task, autistic children were found to over-extend labels, sorting items that were colour matches for the standard into the 'blicket' box at a higher rate than the TD control group. The only objects that autistic children reliably excluded from the novel category were known items from a different category.

The fact that four slightly different tasks across three different studies have produced different findings about shape bias in autism may not be surprising. In Tek et al. (2008), the implicit nature of the measurement is not the only difference between the IPL and pointing tasks. In the forced-choice task, the standard remains visible for reference, while in the IPL task only the test objects are visible on screen during the decision making. As noted by Oakes and Kovack-Lesh (2007), when the standard is not present for reference the task does not measure categorisation independent from memory. This means that shape bias tasks with a memory requirement may unwittingly detect differences in object memory for autistic children rather than differences in attention during word learning. There is evidence of weaker visual working memory for autistic individuals (Funabiki & Shiwa, 2018; Zhang et al., 2020), so tasks that require offline comparison could be more challenging. However, there is also evidence to suggest that autistic children perform well in offline generalisation tasks, just at a slower speed than TD children (Hartley et al., 2019, 2020). A possible explanation for increased processing time is that autistic children may process additional visual information about objects, due to an enhanced perceptual capacity (Mottron et al., 2006), however, this could result in less efficient encoding of the most useful information. The difference in shape bias between the two tasks could be influenced by the additional working memory demands of the IPL task compared to the pointing task where the standard remained visible on the table for reference.

Research to-date suggests that the question of whether or not autistic children have a shape bias is not clear cut. While the findings of previous research appear to be contradictory, it is possible that these differences could be accounted for by methodological choices. In the current study we investigate whether task differences affect shape bias in autism. In two studies, we manipulated the visibility of the standard for 'online' or 'offline' comparison, and whether or not the standard was labelled. Both experiments were presented on a touch-screen computer as a series of games, which autistic children can find particularly engaging (Allen et al., 2016).

Experiment 1 describes a forced-choice task with a similar design to Field et al. (2016b), in which children were presented with a novel item – the standard – along with three test items that matched on one characteristic: colour, texture, or shape. The task was to choose one test item as another example of the standard. The experiment included 'online' trials, in which an image of the standard remained visible for reference in line with typical shape bias task with real objects, and 'offline' trials, where the standard was shown briefly then removed before the test items were shown to prevent direct comparison. These different task types also had a labelled condition and an unlabelled condition (within-subjects) to investigate whether shape bias was stronger in a naming context.

We predicted that typically developing children would be more likely to choose the shape-match item than one of the distractors, and that they would do so with a higher reliability than autistic children. Furthermore, we expected a stronger shape bias for labelled items overall, in line with previous findings that both TD (e.g. Landau et al., 1988) and autistic children (Field et al., 2016b) have exhibited an effect of labels in earlier studies. However, the labelling effect may be weaker for autistic children compared to TD children, based on Tek et al.'s (2008) finding that labels did not enhance attention to shape for this population. We also hypothesised that the need to hold a representation of the standard in working memory in the offline condition would impact the shape bias, however, the direction of this difference was difficult to predict. The additional demands on working memory should make this condition more challenging and, potentially, disrupt the shape bias. However, there was also the possibility that automatic attention to shape would allow children to encode the most relevant information and not the irrelevant colour and texture details, thus enhancing generalisation (Hartley et al., 2019, 2020). Finally, we expected higher age-equivalent receptive vocabulary scores to be associated with a stronger preference for the shape match item in both groups.

In Experiment 2 we used the same stimuli and conditions to investigate the effect of labels and visibility of the standard on shape-match acceptance in a yes/no task. This task allowed children to accept or reject each item individually. The advantage of this task is that it allows children to decide whether all, none, or a subset of their choosing, are members of the same category as the standard, which makes it well suited to detecting over-generalisation in autism as found in Hartley and Allen (2014). We hypothesised that autistic participants would accept shape-matches more often than the colour- and texture-match options, but at a lower rate than TD children.

Overall, this study addresses whether the differing demands of the yes/no vs forced choice task types can result in different strength of the shape in autistic children compared to typically developing children, and what affect the memory requirement and addition of labels may have. As shape bias has been demonstrated in typical development in each of these task variants, understanding which task demands pose a particular challenge in autism provides a valuable insight into designing teaching strategies that play to this population's strengths.

2.4 Experiment 1: Online and offline categorisation in autism using a forced-choice task

2.4.1 Method

2.4.1.1 Participants

Participants were 16 autistic children aged between 4 years 7 months and 9 years 7 months (*M* age = 7.17 years; *SD* = 1.45 years), and 14 TD children aged between 2 years 9 months and 4 years 8 months (*M* age = 3.70 years; *SD* = 0.51 years). Autistic participants had received a formal diagnosis prior to participating. The groups were matched on receptive vocabulary as measured with the British Picture Vocabulary Scale (2nd edition) (Dunn et al., 1997). A *t* test was used to compare the BPVS scores for the final samples of children with ASD (*M* = 51.32; *SD* = 16.38; range: 30 to 77) and TD children (*M* = 49; *SD* = 16.38; range: 36 to 65). There was no statistically significant difference in receptive vocabulary scores between the groups, t (25.58) = 0.24, p = .814, two-tailed. Additional standardised measures were taken for expressive vocabulary using the Expressive Vocabulary Test (2nd Edition) (Williams, 2007), and non-verbal cognition using the Leiter-3 cognition battery subtest (Roid et al., 2013). Due to Covid-19 interrupting testing, several participants in each population did not have the opportunity complete these measures (EVT: 7 autistic and 8 TD, Leiter-3: 9 autistic and 8 TD). Sample characteristics for both experiment 1 and 2 are detailed in Table 1.

This research was approved by the Lancaster University FST Research Ethics Committee (REF: FST18017). Participants were all recruited from preschools, and mainstream and specialist schools in the Northwest of England and had signed consent from their caregivers to participate.

A further 12 autistic children involved in the project completed the experimental tasks but were not included in the final samples due to their BPVS scores exceeding 78 months[1] (*n* = 8) or inability/unwillingness to complete the BPVS (*n* = 4). The experimental task scores for these children remain available in the open dataset for future analysis: https://osf.io/sgcex/.

[1] The pre-registered plan was to recruit 60 typically developing children aged between 2-years-old and 5-years-old, and 60 children with an autism spectrum condition diagnosis aged between 4-years-old and 9-years-old, who would take part in all five experiments included in this thesis. However, due to the Covid-19 pandemic, data collection was halted in March 2020 before all of the participants could complete every study. Prior to school closures, a total of 60 children were recruited to the project (40 to the ASD group and 20 TD). Only some of these participants had completed the full three sessions, however, as each session included at least one self-contained study it has still been possible to include the data. Furthermore, our pre-registration stipulated that we would match the groups on mean receptive vocabulary scores by adjusting the age of the typically developing sample to match the children with autism. In order to balance group sizes and vocabulary scores, only children whose receptive vocabulary score was equivalent to 78 months or less were included in the analysis since additional data collection was not possible. The complete anonymised data set, including the children who scored higher than the receptive vocabulary threshold, will be made available on the OSF repository: https://osf.io/sgcex/

Table 1: Sample description for Experiments 1 and 2 (standard deviation and range in parentheses)

Population	Study	N	Chron. Age (years)	BPVS age equiv. (years)	EVT age equiv. (years)	Leiter-3
TD	Experiment 1	14	3.70 (0.51; 2.75-4.67)	4.15 (0.86; 3.0-5.42)	4.79 (0.68; 3.75-5.50)	46.8 (8.89; 34-62)
	Experiment 2	14	3.76 (0.52; 2.75-4.67)	4.27 (0.82; 3.0-5.42)	4.8 (0.67; 3.75-5.50)	46.85 (8.99; 34-62)
ASD	Experiment 1	16	7.17 (1.45; 4.58-9.58)	4.28 (1.37; 2.5-6.42)	3.86 (1.15; 2.17-5.58)	38.7 (13.27; 24-56)
	Experiment 2	15	7.15 (1.19; 5.08-9.08)	3.97 (1.42; 1.67-6.42)	3.88 (1.19; 2.17-5.58)	39.36 (13.21; 24-56)

All participants received stickers throughout testing as rewards. At the end of the final session, they were also given a choice of book to take home with them, along with a debrief sheet for their caregiver. Schools were also given an Amazon gift card as a thank you gift for supporting the research.

2.4.1.2 Design

The study employed a mixed (2x2x2) design. Population was a between-participants variable with 2 levels: typical development/autism. There were two within-participants variables, each with 2 levels: Standard visibility (online/offline) and Labels (label/no-label). In the online condition, the ‘standard’ for the novel category was presented on screen at the same time as the test items, allowing direct online comparison. In the offline condition, the standard and test items were never visible at the same time, hence comparisons had to be completed from memory. The order of completion for standard visibility was counterbalanced: half of the children completed the online condition first. Each level of standard visibility included both label and no label trials.

Trials included an audio instruction directing children to 'look' at the standard and a question asking them to choose a test item. In label trials, this audio provided a name for the item (e.g. "Look! A teep!"). The no-label trials directed attention without overtly naming the object ("Look at that!"). For all participants, no-label trials were completed first. This was based on the principle that labels may encourage children to generalise by shape (Landau et al., 1988), so experiencing the label trials first could influence responses to no-label trials. Participants completed four experimental trials for each condition. Each of the four trials used the same standard, but three different variations of the test items, so no test item was seen twice within the same task (see Figure 1).

2.4.1.3 Materials

All digital stimuli described in this section, including images, raw Blender stimuli files, audio files, and Python code to run the experiment are publicly available on the OSF: https://osf.io/sgcex/.

2.4.1.3.1 Visual stimuli

The visual stimuli were images of novel objects created using Blender: an open-source 3D graphical modelling software capable of producing realistic-looking images for video games and animation (Version 2.78; Blender Development Team, 2017). The images were organised into sets, with each set consisting of one standard and 12 test items (four colour-match; four texture-match; four shape-match), plus two control items that did not match the standard on any dimension (see Figure 1). These 12 test items were grouped into four fixed 'triads' to ensure there was no duplication of colour or material within a trial. In total, four different sets were used, referred to as sets A to D, resulting in 60 item images in total (Figure 2).

2.4.1.3.2 Audio stimuli

Audio stimuli for the study were recorded by a female native British English speaker and edited using Audacity® (Version 2.1.3; Audacity Team, 2017). All vocal recording featured the voice of a female British English native speaker. The wording for each track is detailed in Table 2. The novel words, *teep* and *foss*, were selected as plausible English words from the English Lexicon Project (Balota et al., 2007).

Table 2: Scripts for pre-recorded audio stimuli for forced-choice task

Wording	Phase/condition used in
“Look at that!”	Familiarisation phase: no label
“Look! A teep!”	Familiarisation phase: label
“Look! A foss!”	Familiarisation phase: label
“Can you find another one?”	Test phase: no label
“Can you find another teep?”	Test phase: label
“Can you find another foss?”	Test phase: label
“Perfect”	Practice phase only
“Uh oh! Try again.”	Practice phase only

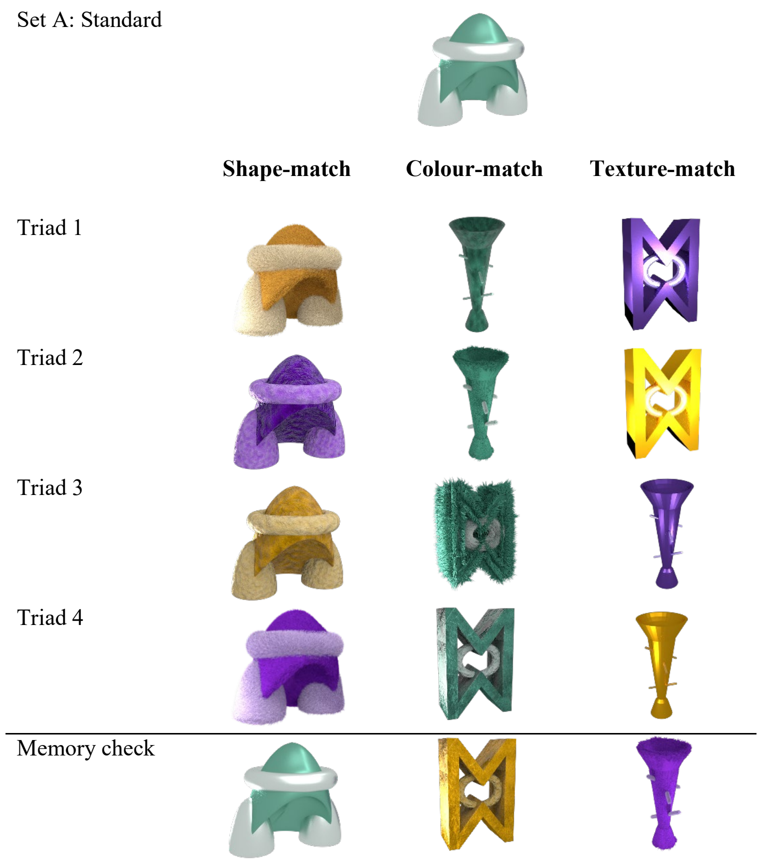


Figure 1: Example complete stimuli set for forced choice task.

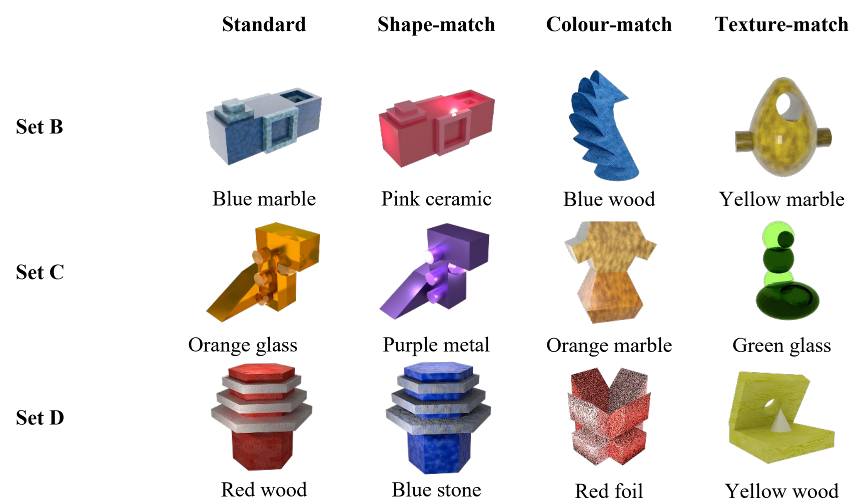


Figure 2: Example groups from stimuli sets B, C and D. The remaining nine items in each set consisted of the same three bases, colours and textures applied in different configurations as see for set A in Figure 1 .

2.4.1.3.3 Presentation software and hardware

The stimuli were all presented to participants in the form of a simple computer-based 'game'. This was created using PsychoPy2 (Peirce, 2007), an open-source Python-based experiment handler. A Microsoft Surface Pro 4 tablet PC with touch-screen functionality was used to run the software. The programme consisted of five 'mini-games', each relating to one experiment (Experiments 3-5 are not reported here). On starting the programme with a new participant, the programme randomly allocated the stimuli sets A to D to one of the four condition combinations. The two labels, *teep* and *foss*, were then randomly assigned to either the online or offline labelled set. These pairing allocations were stored for the participant, along with their current progress.

Timings and stimuli display were controlled by the programme, along with the recording of participant's responses on the touchscreen. The screen layout during the response phase of a trial is illustrated in Figure 3.

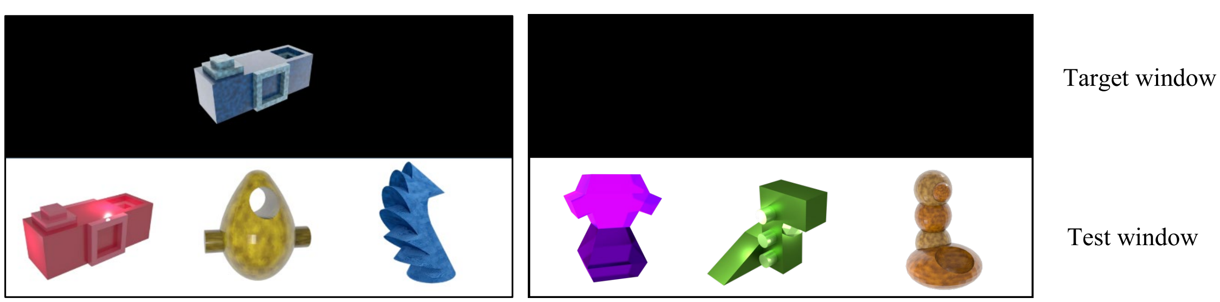


Figure 3: Example of the display layout during an experimental trial in the online (left) and offline (right) condition.

2.4.1.3.4 Button training presentation

An interactive PowerPoint presentation was created for the warm-up task. The presentation consisted of six slides, each with the screen divided horizontally, as in the main experiment (see Figure 3). The first slide displayed a banana in the centre of the test window, with white cross in a red circle to the left, and white tick in a green circle on the right. The target window displayed the *Shoe?* in white text. The cross button was set to trigger the slide transition when tapped. The remaining 5 slides had the following object, text and response pairings: horse – "Horse?" (tick), ball – "Ball?" (tick), horse – "Train?" (cross), banana – "Banana?" (tick), ball – "Elephant?" (cross). Only the correct response was set to trigger the transition.

2.4.1.4 Procedure

Children took part in this experiment as part of a project consisting of five studies. All testing took place in the child's school or nursery, in a separate room or a quiet corner, and a familiar adult was invited to accompany the child. A typical testing session would begin with one of the standardised tests: BPVS (session 1), EVT (session 2) or Leiter-3 (session 3). The current experiment could be completed on any visit due to task order being counter-balanced.

2.4.1.4.1 Warm-up task

Prior to completing the main task, the children completed a short training exercise on how to use the 'yes' and 'no' buttons to respond during the tasks. As the experimenter did not know the task order assigned for each participant in advance, this warm-up was completed first regardless of whether a yes/no or forced-choice task followed. The experimenter informed the child that they were about to play a game on a computer that was going to ask them some questions, so first they were going to practice. The Surface Pro was either stood on the desk in front of the child, or given to them to hold on their lap if they preferred, with first slide (showing a horse) of the button training PowerPoint presentation loaded. The experimenter asked if it was a picture of a shoe. When the child indicated it was not, either verbally or by shaking their head, the experimenter agreed and asked them which button meant 'no'. If the child correctly identified the red button with the cross, they were then encouraged to try pressing it. If they did not know, the experimenter told them which button to use and invited them to try pressing it. When the red button was tapped, the presentation moved on to the next slide. The experimenter asked if the new picture was a ball, which was true, to model a 'yes' response. Participants were then encouraged to try the remaining four slides alone, only receiving reminders from the experimenter if they were unsure of what was expected. If, after completing all six slides, children had not managed to use the buttons correctly three times in a row, the presentation was restarted and they could try again. All participants were able to give three consecutive correct responses within 12 practice slides.

2.4.1.4.2 Forced-choice NNG task.

Following the warm-up task, the experimenter initiated the PsychoPy experiment and returned the tablet to the child. From that point the experiment ran without intervention from the experimenter. The experimenter could intervene to direct attention back to the game, and could press a button on behalf of the child if they gave a clear verbal or gestural response. If a child could not choose an answer and asked for help, the experimenter would advise them to just give what they thought was the best answer. If a child was still unwilling to respond on an individual trial, the experimenter could use a secret code to move on to the next trial and record the response as 'skipped'.

Practice block. The game began with either an online or offline trial, depending on which order had been allocated. Both conditions started with an identical familiarisation phase, which displayed an animation of a known object (a green bed) rotating on the screen accompanied by an audio track: “look at that!”. Once the audio had finished, a ‘next’ button was displayed on screen which would skip to the following phase if pressed. Otherwise, the animation repeated for 30 seconds before moving on automatically. In the online condition, a static image of the same bed was displayed in the target window (see Figure 3 for display layout). After one second, the audio track asked “Can you find another one?”. Following this, images of an identical bed, a blue butterfly, and a gold trophy appeared in the test window. A touch detected on any of the three items would cause a green circle to appear around the chosen item. Tapping any other location on the screen, including the standard, had no effect. Touching the bed, which matched the example, resulted in audio feedback to confirm the correct choice (“Perfect!”), and automatic progression to the next phase. If a distractor was touched, the audio track would play “Uh oh, try again!” and continue to wait for another response. Subsequent incorrect taps would receive the same response, and the programme would not progress to the next part of the game until a correct response had been given.

For trials in the offline condition, the procedure was identical except that the standard was removed from the target window before the test items appeared in the test window.

Experimental block. On detecting a correct response on the practice block, a ‘no-label’ experimental block was initiated for either the online or offline condition, corresponding with the condition that had been practiced. As with the practice trial, this began by playing an animation of a rotating object. For the experimental trials, this object was the standard for the stimuli set that had been randomly allocated to the condition. The no-label condition used the same audio recording as the practice block: “Look at that!”. After one full loop of the animation, the phase could be progressed by pressing the ‘next’ button or waiting for 30 seconds.

The test phase consisted of four trials. The four triads of test objects (e.g. Figure 1) were presented in a random order by the programme, and were positioned randomly in one of three fixed locations in the test window (left, middle, or right: see Figure 3). The phase progressed as described for the practice block, with the exception that no feedback was given. The programme would highlight whichever test item was chosen with a green circle, and wait for 1 second. During this time, no audio played, and the programme would accept a change of answer. The final choice was logged and saved automatically in a .csv file. Following this, the next trial would begin with the next triad for the set, until all four triads had been responded to. After completing four experimental trials, there was a final attention check trial, which presented a copy of the standard along with two distractor objects which differed from the standard on all features.

In the online condition, the standard remained visible in the target window throughout all four trials. In the offline condition, the standard was shown again for one second at the beginning of each trial, and removed prior to the audio track playing or the test items appearing. The visual display for each stage of the programme in both conditions is shown in Figure 4.

Immediately following the completion of the no-label block, the label block for the same standard visibility condition began. A new standard was presented using the same familiarisation procedure, followed by four experimental trials with the new stimuli set. The audio stimuli were the only difference between label and no-label conditions. During the familiarisation phase, the audio instruction included a label: “Look! a teep!”. The audio track for the experimental trials repeated the label: “Can you find another teep?”

Once the no-label and label trials were completed for the first standard visibility condition, the programme started a new practice phase as described above for the condition still to be completed (i.e., online or offline). This showed participants that the parameters of the game had changed prior to completing the next set of experimental trials. For instance, if the online condition had been completed first, the practice phase served as a pre-warning that the standard was not going to remain visible for the next round of the game. The complete procedure was then repeated for the no label, then label conditions, using the two remaining stimuli sets.

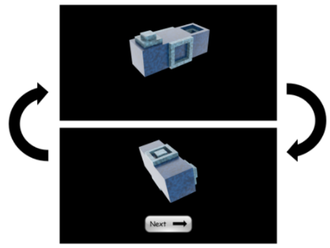
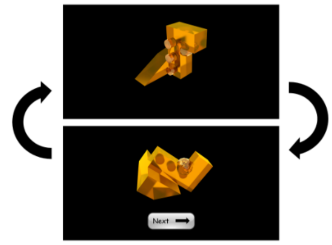
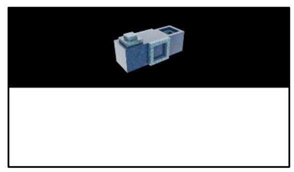
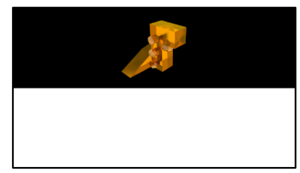
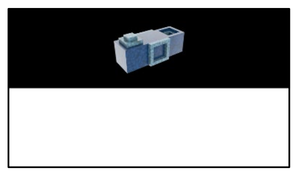
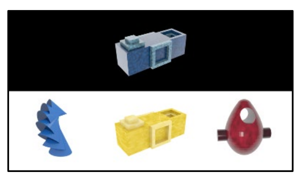
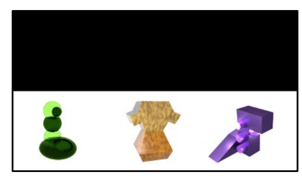
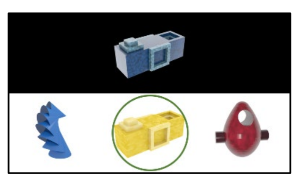
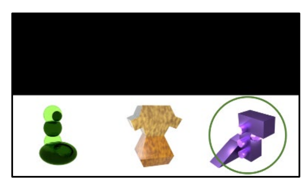
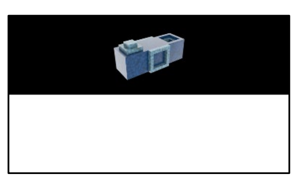
Stages and timings	**Audio track (label/no label conditions)**	**Online block visual display**	**Offline block visual display**
Familiarisation phase: Animation of Standard rotating for up to 30 seconds	"Look! A teep!" *or* "Look at that!"		
Test phase Stage 1: 1 second display time	*No audio*		
Stage 2: Audio presentation (approx. 3 seconds)	"Can you find another teep?" *or* "Can you find another one?"		
Stage 3: wait for participant response	*No audio*		
Stage 4: Response confirmation displayed for 1 second	*No audio*		
Repeat from stage 1 for next triad in the stimuli set	*No audio*		

Figure 4: Visual display for each stage of the forced-choice game in online and offline conditions

2.4.2 Results

2.4.2.1 Shape-match acceptance against chance

To establish whether the children showed an overall shape bias, we tested whether the likelihood of choosing the shape-match was greater than expected by chance (0.33). A two-tailed, one-sample *t*-test revealed that both the TD and ASD groups selected the shape-match test item significantly more frequently than chance across conditions and trial types (TD: $M = .76$, $t(13) = 5.64$, $p < .001$, $d = 1.65$; ASD: $M = .62$, $t(15) = 3.48$, $p = .003$, $d = 0.88$). Both groups were also chose texture-match and colour-match items significantly less than expected by chance: texture-match (TD: $M = .10$, $t(13) = -6.70$, $p < .001$, $d = 1.76$; ASD: $M = .17$, $t(15) = -4.11$, $p < .001$, $d = 1.00$); colour-match (TD: $M = .14$, $t(13) = -5.02$, $p < .001$, $d = 1.32$; ASD: $M = .21$, $t(15) = -2.18$, $p = .046$, $d = 0.53$). These results indicate a clear shape bias for both autistic and TD participants in this task (see Figure 5).

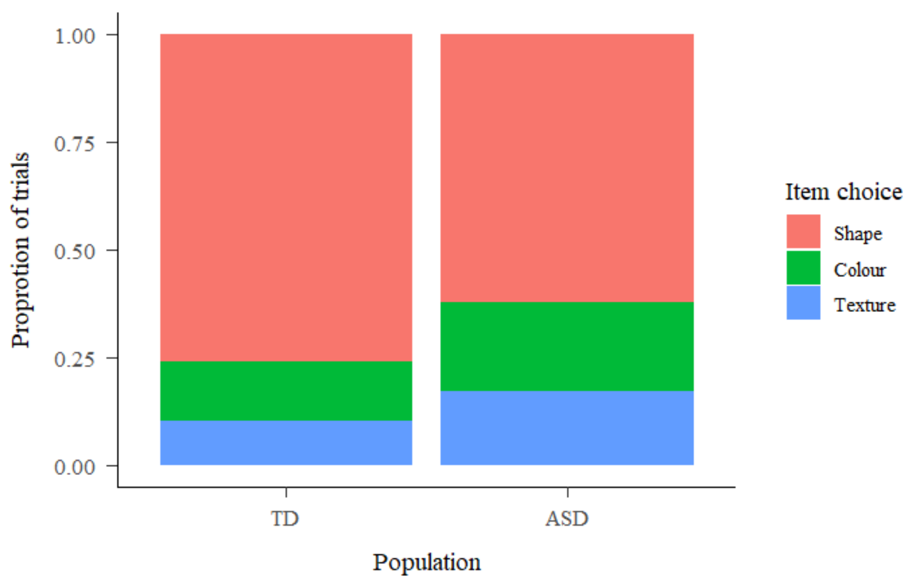


Figure 5: Proportion of forced-choice trials with shape-match, colour-match and texture-match responses for autistic and typically developing (TD) participants

2.4.2.2 Effect of population, labels and task type

All modelling was completed using RStudio Version 1.3.1073, utilising the “lme4” library (Bates et al., 2015), and post hoc comparisons were conducted using the “emmeans” package (Lenth, 2022). To investigate whether the magnitude of shape bias differed as a result of population differences, the visibility of the standard, or the presence of labels, we first created a binary outcome measure of whether responses were consistent with shape bias or not. Responses for each trial were coded as ‘shape match selected’ (1) or ‘other match selected’ (0). Choosing either the texture- or colour-match items resulted in a 0 coding. These responses were submitted to generalised linear mixed-effects models for binomially distributed outcomes (also known as a ‘mixed logit model’), which included the following fixed effects: population (contrast coded; typical development: -0.5, autism: 0.5); standard visibility (contrast coded; online: -0.5, offline: 0.5); labels (contrast coded; no-label: -0.5, label: 0.5). Random intercepts for participant and item set were included, along with random slopes for standard visibility by-participant, as determined by fitting a maximal effects structure and reducing complexity until the model converged (Barr et al., 2013). Including fixed effects and the planned interaction terms (population x labels; population x standard visibility; labels x standard visibility) did not significantly improve fit in comparison with a baseline model containing only the random effects ($\chi^2 = 5.29$, $p = .507$).

2.4.2.3 Effect of receptive vocabulary

To investigate the hypothesis that the strength of the shape bias would increase in line with children’s receptive vocabulary, the mean-centred age equivalent in months as scored on the BPVS was entered into a mixed logit model as a fixed effect. The change in deviance between the fixed effect model and the baseline was compared with a likelihood ratio test (Bates et al., 2015), which showed a significant improvement of model fit when vocabulary was added as a predictor ($\chi^2(1) = 7.22$, $p = .007$). To test whether this association held for TD and autistic participants, an interaction for receptive vocabulary and population was added, however, this interaction was not significant and did not improve fit in comparison with the model containing vocabulary as a fixed effect ($\chi^2(2) = 5.50$, $p = .064$). Table 3 shows the output of the best-fitting model.

Table 3. Results of mixed logit model of effect of receptive vocabulary on shape-match choices

Effect	Group	Term	β	SE	*z*	*p*	
Fixed		Intercept	1.51	0.47	3.23	.001	**
		Receptive vocabulary	0.08	0.03	2.73	.006	**
			Variance	SD	Corr		
Random	Participant	(Intercept)	3.91	1.98			
		Standard visibility	1.77	1.33	-0.29		
	Stimuli set	(Intercept)	0.21	0.45			
AIC	454.6			logLik	-221.3		
BIC	479.6			Deviance	442.6		
				df residual	470		

Number of obs: 476, participants: 30, stimuli sets: 4
Significance indicators: $p < .05$*; $p < .01$**; $p < .001$***

Greater receptive vocabulary was associated with increased likelihood of choosing the shape-match object. For each 1-month increase in receptive vocabulary, the model estimated an approximately 8% increase in probability of choosing the shape-match over one of the two distractors, indicating that shape bias was stronger for children with higher vocabulary abilities.

The model did not detect a difference between populations (see Figure 6).

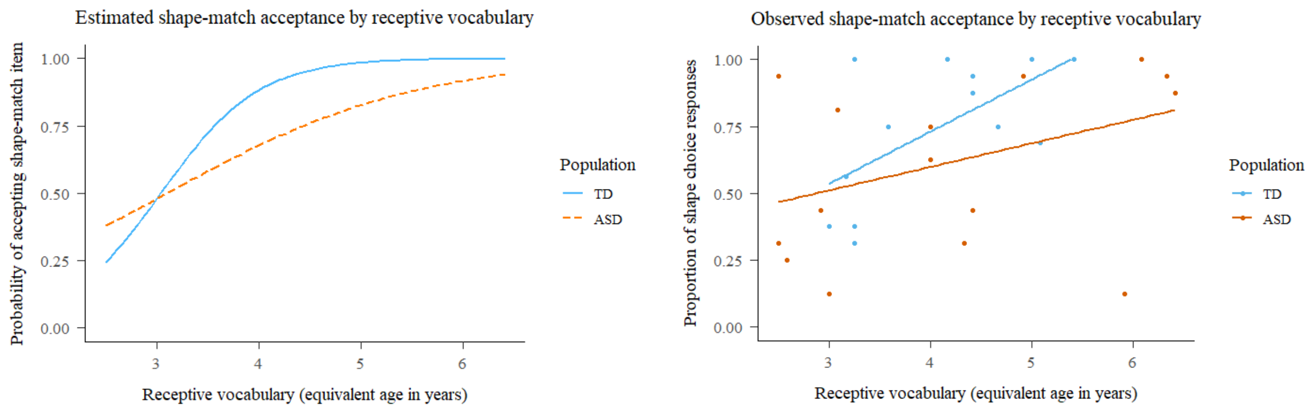


Figure 6: Relationship between receptive vocabulary and shape-match choices for autistic and TD children as (a) predicted by the model, and (b) measured as proportion of shape-match choices.

2.4.3 Discussion

Study 1 investigated whether population, standard visibility (online/offline), or the presence of labels affected the strength of shape bias in a forced-choice task. Our results did not find a difference between TD and autistic children in any condition: both populations showed a clear preference for shape-match items regardless of labels or standard visibility. We hypothesised a stronger shape preference for labelled items compared to non-labelled items, however, our results provided no support for this, nor did we detect any difference between online and offline trials.

Higher receptive vocabulary scores were associated with a higher likelihood of choosing the shape-match object across groups. For TD children, this finding is consistent with research linking shape bias with vocabulary (e.g. Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999). Gershkoff-Stowe and Smith (2004) found that the number of nouns 1-year-old infants had in their vocabularies strongly predicted attention to shape. Tek et al. (2008) also tested the development of shape bias over a 12-month period at 4-month intervals, and found shape preference on a forced-choice task increased at each visit. The relationship between vocabulary and attention to shape persists into later development, as Field et al. (2016b) found TD 2- to 7-year-old children with a higher verbal mental age also exhibited a stronger shape bias. Our finding that vocabulary predicts children's shape bias adds further support for the link between vocabulary and shape bias in typical development.

Our results relating to the performance of autistic children, however, differ from previous findings. Tek et al (2008) found relatively little change in shape preference over 12-months for their autistic participants, and argued that they did not exhibit a shape bias despite having a sufficiently well-developed vocabulary. Field et al. (2016b) also only found evidence of shape bias in autistic children with a higher verbal mental age – the mean equivalent of a typical 6.5-year-old – and not when the mean verbal mental age (VMA) was around 3.5 years. Those findings pointed to a delay in autism, and the possibility that a larger vocabulary might be required for the children to learn that shape is a useful cue. The current findings differ in that we found no significant difference by population, and there was no indication that autistic children with lower receptive vocabulary had a disproportionately weaker shape bias than TD children with comparable vocabulary. We found that higher receptive vocabulary predicted stronger shape bias for both autistic and TD participants.

The discrepancy with Field et al (2016b) may be explained by their decision to divide the sample by age equivalent vocabulary score for analysis. The chronological ages of the autistic participants covered a large range: 4- to 17-years-old. The impact of dividing the group into two based on VMA meant the lower scoring group consisted of children with far greater disparity between their age and language abilities than the high VMA group did. When the developmental groups were compared without dividing them by vocabulary, the autism group chose a similar proportion of shape match items. Hence, the lack of shape bias observed in that case may be due to the severity of the language difficulties of that subset of children. It does not necessarily follow that younger autistic children with similar VMA scores, but less disparity, would also not show a shape bias. Our findings suggest that that the link between vocabulary size and generalisation on the basis of object shape is present in both autism and typical development. Overall, shape bias was as expected in line with receptive vocabulary. However, further research is necessary to determine differences in the strength of this link and the impact of the magnitude of the disparity between chronological age and VMA.

Our primary hypothesis was that strength of shape bias would vary by standard visibility, as the online trials allowed for direct comparison whereas offline trials relied on comparison from memory only. Our results provided no support for this hypothesis. This prediction was based on the observation that autistic children in Tek et al. (2008) did not show a shape bias during an intermodal preferential looking paradigm (IPL) task. This task uses two screens to display the stimuli: the standard appears for several seconds on one screen, then the other, then the two test items are each displayed on a screen so the looking time to each side can be recorded. In this procedure, the standard is not visible on screen at the same time as the two test objects. Despite not observing a shape bias in the IPL task, autistic children did prefer the shape-match in a pointing version of the task where they had a visible standard for reference. This was not the only difference between the two methods, and we did not hypothesise that standard visibility would be the only source of differing shape bias in this task.

One difference between the present study and Tek et al. (2008) was the age of participants. Vlach (2016) found that shape bias predicted memory for shape in children by the time they reached four years of age. They suggested that as attention becomes tuned to shape, encoding object shape into memory also becomes more efficient. However, while four-year-olds showed better memory for object shape, two- and three-year-olds did not remember it any better than colour, even when there was no delay between the presentations. As the majority of children in Tek et al. (2008) were below four years old by the end of the study, it is possible that task type is more influential earlier in development when shape encoding is less effective, and that older children in the current research had a better memory for object shape from which to make comparisons.

Another difference between studies concerns the overall similarity of stimuli used. In contrast to the highly different items used for the current research, the objects used in Tek et al. (2008) had more perceptual similarities. In a later study, Tek et al. (2012) found that typically developing 2-year-olds only showed a shape bias in the IPL task when shapes were highly different. However, when the colour-match distractor was a similar shape to the standard, children looked more to that than the perfect shape match. In this case, the distractor was thought to a better ‘overall match’ as the shape was close enough to belong to the same category. The high contrast in the shapes used in the current study may have compensated for the additional difficulty of having to remember the object in the offline task. As the shapes were so different, there was less risk of misidentification due to a memory lapse. This is also consistent with the findings of Hartley et al. (2019, 2020), showing that autistic children were either equally or more successful at generalising by shape in an offline task with highly contrasting stimuli, even after a 5 minute delay.

Labels were not a significant predictor in our model, and so our hypothesis that shape bias would be stronger for labelled items was not supported. We also found no evidence of population differences for this variable. Of the previous studies that have investigated shape-based generalisation in autism using a forced-choice task (Field et al., 2016b; Hartley et al., 2019, 2020; Tek et al., 2008), only two of those included a label manipulation. The first was Tek et al. (2008), whose findings were similar to our own. In the pointing task, autistic children showed a preference for shape-based generalisation for both named and unnamed items, as was also the case in the current research. In contrast, Field et al. (2016b), found that autistic children made significantly more shape-match choices when standards were labelled. Furthermore, in the unlabelled condition, the proportion of shape-match choices was not greater than chance. Thus, a clear shape bias was observed for labelled items only for their autistic participants.

While it is possible that we failed to detect an effect of labels due to our research having a smaller sample size, this is unlikely based on the observed results. The proportion of trials that autistic children selected the shape-match item was nearly identical in both conditions: 0.621 (no-label) and 0.625 (label). As our findings are in agreement with the results of Tek et al.'s (2008) pointing task, we suggest the anomalous finding of Field et al (2016b) may reflect an additional challenge in their unlabelled trials that was not present in the other studies. One possibility is the way that the question was asked on the unlabelled trials. For TD children, the wording of unlabelled trials has been shown to affect whether they treat them as a generalisation task or a different kind of grouping task, such as thematic grouping (Diesendruck & Bloom, 2003). Field et al. (2016b) used the phrase "Can you give me the other one?", which may be a more ambiguous request than in Tek et al. (2008) ("Which one looks the same?") or the current study ("Can you find another one?"). Although this explanation is speculative, there are good reasons to question whether autistic and TD children interpret instructions in the same way. Difficulties with the pragmatics of language can be a characteristic of autism, even for children and adults who score highly on other vocabulary measures (e.g. Naigles & Tek, 2017). Field et al.'s finding that TD children showed a strong shape bias for both labelled and unlabelled conditions, while autistic children struggled, could reflect a different interpretation of the meaning of the instructions rather than an underlying difference in shape bias per se. This possibility should be considered with future research.

As we found no effect of labels, and no significant population differences, this study also did not provide any evidence for a label effect for the TD participants. This was consistent with Field et al. (2016b), who also found a strong shape bias in TD children in both labelled and unlabelled conditions. Tek at al. (2008) also found no difference for labels at three out of four time points, only measuring a significant difference at visit 3 when the children were around 28-months-old in the study. The lack of a label effect may be explained by differences in the stimuli. Previous research that has reliably found a label effect in TD children (e.g. Landau et al., 1988) used perceptually similar stimuli that only deviated slightly from the standard, and only on one dimension (shape, material, or size). If labels strengthen shape bias by cuing attention to shape, the advantage may only be noticeable when shapes are similar. For the highly contrasting object shapes used in the current study, shape differences may be obvious enough that an attentional cue is not necessary. Therefore, our results suggest that, for highly contrasting stimuli in a forced-choice task, both autistic and TD children exhibit a shape bias in lexical and non-lexical generalisation.

In Experiment 1, autistic children demonstrated a shape bias in all conditions. Furthermore, shape bias increased with receptive vocabulary and did not significantly differ from typical development. The task was designed to be similar to Field et al. (2016b), partly to aid with comparison, but also to maximise the chances of children demonstrating a shape bias. Firstly, the stimuli were designed to be visually distinctive, so the shape-match test item was the only obvious match for the standard if children were applying a shape bias. There was no similarity that could lead children to conclude that a distractor might be a ‘close enough’ match to the shape. Secondly, the forced choice response meant that children only needed to identify the best example from the available options. There was no requirement to make judgments about the other two objects. As a result, a child who suspected that all three items in the array were potentially “vinks”, but successfully identified the most “vink-like” of the three, would achieve the same score as a child who would exclusively generalise by shape. In this way, a slight perceptual bias can become exaggerated in a forced-choice task (Samuelson et al., 2009). However, in a yes/no NNG task, children must make an active decision to include or exclude the distractor items from the category. As a result, we may observe additional biases in autism that would be suppressed by the ‘winner takes all’ nature of the forced-choice task. To the best of our knowledge, the yes/no NNG task has not previously been used to investigate shape bias in autism. However, in research with a sequential sorting task, which also allows individual decisions on each item, autistic children are more likely to generalise new words to distractors in addition to shape-match items (Hartley & Allen, 2014; Tovar et al., 2020). In Experiment 2, we presented the same stimuli as in Experiment 1 in a yes/no task to investigate whether autistic children would still demonstrate a shape bias when the task also required decisions to be made on the distractor items.

2.5 Experiment 2: Does an 'online' vs 'offline' referent affect shape bias for autistic children in a yes/no task?

2.5.1 Method

2.5.1.1 Participants

Participants were 15 autistic children (*M* age = 7.15 years; *SD* = 1.19 years) and 14 typically developing children (*M* age = 3.76 years; *SD* = 0.52 years), all recruited from specialist, primary or nursery schools in the Northwest as in Experiment 1, groups were matched on receptive vocabulary. The BPVS scores of the autistic children (*M* = 47.69; *SD* = 16.99) and TD children (*M* = 51.3; *SD* = 9.86) did not significantly differ, t (22.94) = 0.63, p = .536, two-tailed.

2.5.1.2 Design

The design was mixed (2x2x2), as described for experiment 1: between-participants variable: Population (TD/ASD); within-participants variables: Standard visibility (online/offline) and Labels (label/no label). For participants who completed both experiments, the counterbalanced order of the standard visibility conditions was kept consistent. If children experienced the online condition first in Experiment 1, this would be the first condition for them in Experiment 2. The manipulation of standard visibility and label conditions were the same as described for Experiment 1.

2.5.1.3 Materials

2.5.1.3.1 Visual stimuli

The yes/no task used the same objects as the forced-choice task in Experiment 1 to ensure that any contrasting findings were due to differences between tasks rather than differences between stimuli. If the same children respond differently to the same stimuli when they are presented in a different task, that is compelling evidence that the task itself is a source of variance. Half of the stimuli set was used to reduce the length of time it would take to complete the study. Showing every item sequentially would require children to concentrate for much longer in comparison to the forced-choice task, where they were presented with three objects at a time. Triads 1 and 2 from each set were used so that children would still see the shape-match item in two different textures and colours (see Figure 1). Each participant was assigned a random object/label pairing at the beginning of the study, which was consistent for both Experiment 1 and 2. So, if a participant had already seen set A referred to as a 'vink' in one task (either forced-choice or yes/no dependent on the counter-balanced order), set A would also be a 'vink' in the alternative task. Each image set consisted of one standard, six test items (two colour-match; two texture-match; two shape-match), and one control item that did not match the standard.

2.5.1.3.2 Audio stimuli

The same recordings from Experiment 1 were used for Experiment 2, with the exception of the test phase instruction. The alternative phrasing "Is this another…" was used to be consistent with a yes or no response, and was also recorded by a female native British English speaker using Audacity® (Version 2.1.3; Audacity Team, 2017). The wording for each track is detailed in Table 4.

Table 4: Scripts for pre-recorded audio stimuli for yes/no task

Wording	Phase/condition used in
"Look at that!"	Familiarisation phase: no label
"Look! A teep!"	Familiarisation phase: label
"Look! A foss!"	Familiarisation phase: label
"Is this another one?"	Test phase: no label
"Is this a teep?"	Test phase: label
"Is this a foss?"	Test phase: label
"Perfect"	Practice phase only
"Uh oh! Try again."	Practice phase only

2.5.1.3.3 Presentation software and hardware

This experiment ran on the same Surface Pro, using the same PsychoPy2 programme as Experiment 1. The assignment of stimuli sets to conditions was carried over from the previous experiment for consistency, so if a participant had previously seen set 'A' assigned the 'teep' label in the offline condition, it would still be a 'teep' in the offline condition in this task.

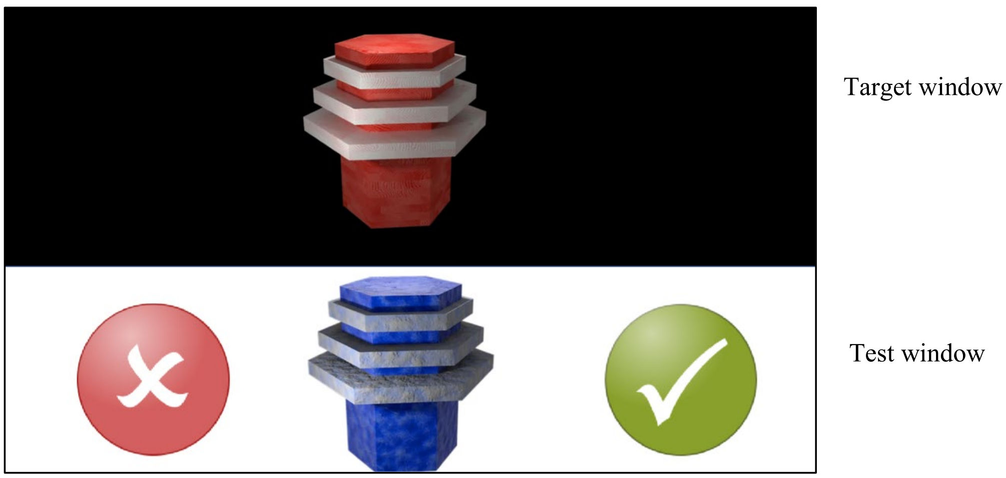


Figure 7: Visual display during a yes/no trial (online condition)

2.5.1.4 Procedure

2.5.1.4.1 Warm-up task

Prior to beginning the study, the children had the opportunity to practice using the 'tick' and 'cross' buttons to give yes or no answers as described in Experiment 1. The warm-up task was repeated before using the computer on any follow-up sessions to ensure that children remembered how to use the interface to respond.

2.5.1.4.2 Yes/No NNG task

The procedure for initiating the experiment and interacting with children while completing the computer mediated task was identical for to the procedure of Experiment 1.

Practice block. The practice block was the same as described in experiment 1, with amendments as described below. Figure 6 provides an overview of a trial in each condition.

Following the familiarisation phase, the ‘green bed’ image was displayed in the target window for 1 second. In the online condition, the image remained visible as an exact copy of the image appeared in the test window. In the offline condition, the standard was removed from view after the second elapsed. All other aspects of the block were the same for both conditions. The audio track asked “Is this another one?” On completion of the audio track, the tick and cross response buttons appeared on either side of the test window (see Figure 7). Audio feedback was given for correct and incorrect answers as in Experiment 1. On recording a correct response, the programme proceeded to a ‘no’ practice trial with the blue butterfly as the test item paired with the green bed as a standard. After a correct ‘no’ response, the experimental block began automatically.

Experimental block. As in Experiment 1, the no-label condition was presented first. The programme procedure was the same as the practice block, but with novel stimuli and no feedback on the response. Following a familiarisation phase for the standard, nine unlabelled experimental trials were presented for either the online or offline condition. Eight of these were randomised and displayed the six test objects and two exact copies of the standard to allow a balanced number of ‘accept’ decisions on the basis of any one characteristic. The programme recorded whether the ‘tick’ or ‘cross’ button had been tapped on each trial. The final trial was always a non-match on any dimension, intended as an attention check.

Following the no label condition trials, a new standard was familiarised for the label condition. Then nine label trials were presented. Only the audio track differed between label and no label conditions: “Is this another teep/foss?” or “Is this another one?”. Once all of the items had been responded to in the label condition, the programme returned to a new practice trial for the standard visibility condition still to be completed. Once both were completed, the game could be exited. Figure 8 illustrates the procedure for an example trial.

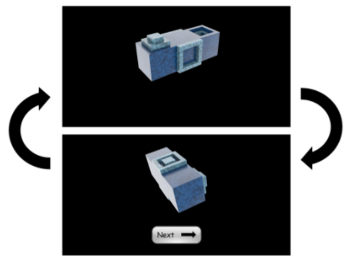
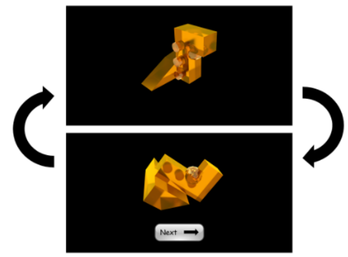
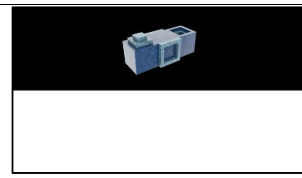
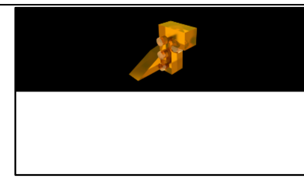
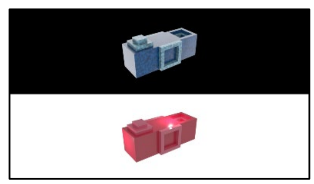
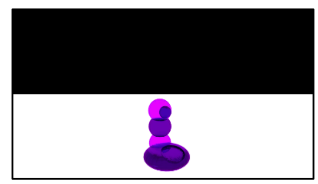
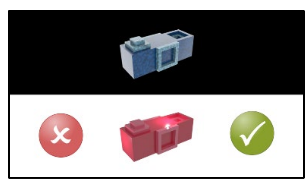
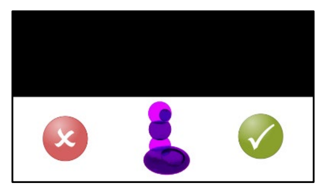
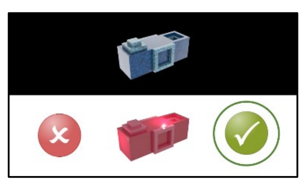
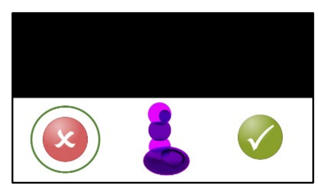
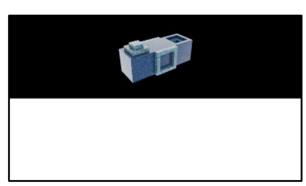
Stages and timings	Audio track (label/no label conditions)	Online block visual display	Offline block visual display
Familiarisation phase: Animation of Standard rotating for up to 30 seconds	"Look! A teep!" *or* "Look at that!"		
Test phase stage 1: 1 second display time	No *audio*		
Stage 2: Audio presentation (approx. 3 seconds)	"Is this another teep?" *or* "Is this another one?"		
Stage 3: wait for participant response	*No audio*		
Stage 4: Response confirmation displayed for 1 second	*No audio*		
Repeat from stage 1 for next triad in the stimuli set	*No audio*		

Figure 8: Visual display for experiment phases for yes/no task

2.5.2 Results

2.5.2.1 Data analysis

We tested the effect of population, standard visibility and labels on shape bias using a yes/no NNG task. Children's responses to each trial, whether they selected the 'tick' or 'cross' button, were coded as '1' for an acceptance (tick) decision and '0' for a rejection. To interpret whether responses were consistent with a shape preference, we also specified an item variable: the 'Standard Congruent Characteristic' (SCC). This variable indicated which feature the test item had in common with the standard on each trial (colour/texture/shape). A mixed logit model containing a fixed effect of SCC (reference category: shape), along with random intercepts for participant and item set, was created as a baseline model from which to compare the experimental variables. SCC in this case can be interpreted as an indicator of shape bias, with higher acceptance of shape-match SCC items indicating a stronger bias, hence it was included in the baseline model. Experimental variables had to improve the model over and above the item characteristics to be shown to have an effect on shape bias strength. For instance, a main effect of population would suggest that one group was more likely to accept items than the other, but without the SCC interaction we would not know if that represented a higher acceptance of shape-match items or distractors. Fixed effects of population, standard visibility, and labels were contrast coded as described in Experiment 1 and added to the baseline model along with the three-way interaction terms (SCC x Population x Standard visibility, SCC x Populations x Labels, SCC x Standard visibility x Labels). The full model was a significantly better fit to the observed data than the baseline model ($\chi^2(18) = 36.97$, $p = .005$), however, none of the 3-way interactions reached significance. To simplify the model, the 3-way interactions were excluded; there was no detriment to goodness of fit ($\chi^2(9) = 10.70$, $p = .297$). There were significant 2-way interactions in the model, so it was not simplified any further. Mean-centred receptive vocabulary was added to the best fitting model, which significantly improved the fit ($\chi^2(12) = 64.96$, $p < .001$). Table 5 shows the model structure and output.

Table 5: Mixed logit model estimating probability of accepting a test item in a 'yes/no' task by standard congruent characteristic (shape, texture or colour), population, standard visibility, label condition and receptive vocabulary

Effect	Group	Term	*β*	*SE*	*z*	*p*	
Fixed		(Intercept)	1.42	0.37	3.87	< .001	***
		SCC (Colour)	-2.11	0.29	-7.17	< .001	***
		SCC (Texture)	-1.96	0.28	-6.95	< .001	***
		Receptive vocabulary	0.08	0.03	2.64	.008	**
		Population	0.48	0.70	0.68	.497	
		Standard visibility	-1.15	0.40	-2.84	.005	**
		Label condition	0.46	0.38	1.21	.226	
		SCC (Colour) x Receptive vocabulary	-0.15	0.02	-6.24	< .001	***
		SCC (Texture) x Receptive vocabulary	-0.12	0.02	-5.16	< .001	***
		SCC (Colour) x Population	1.12	0.56	1.99	.046	*
		SCC (Texture) x Population	0.92	0.54	1.71	.088	.
		SCC (Colour) x Standard visibility	1.43	0.54	2.64	.008	**
		SCC (Texture) x Standard visibility	1.19	0.53	2.27	.023	*
		SCC (Colour) x Label condition	-1.05	0.53	-2.01	.045	*
		SCC (Texture) x Label condition	-0.48	0.51	-0.95	.344	
		Population x Receptive vocabulary	0.00	0.06	-0.07	.945	
		Receptive vocabulary x Standard visibility	-0.03	0.03	-0.93	.353	
		Receptive vocabulary x Label condition	-0.02	0.03	-0.58	.561	
		SCC (Colour) x Population x Receptive vocabulary	0.16	0.05	3.26	.001	**
		SCC (Texture) x Population x Receptive vocabulary	0.05	0.04	1.08	.281	
		SCC (Colour) x Receptive vocabulary x Standard visibility	0.00	0.04	0.11	.914	
		SCC (Texture) x Receptive vocabulary x Standard visibility	0.09	0.04	2.39	.017	*
		SCC (Colour) x Receptive vocabulary x Label condition	0.01	0.04	0.31	.758	
		SCC (Texture) x Receptive vocabulary x Label condition	0.02	0.04	0.46	.645	
			Variance	SD			
Random	Participant	(Intercept)	2.29	1.51			
	Stimuli Set	(Intercept)	0.02	0.15			
AIC	704.6				logLik	-332.3	
BIC	794.8				Deviance	664.6	
					df residual	652.0	

Number of obs: 672, participants: 29, stimuli sets: 4
Significance indicators: p>.05*; p>.01**; p>.001***

2.5.2.2 Item acceptance consistent with shape bias

The model estimated that SCC was a significant predictor of item acceptance. Overall, items that were a shape-match had an approximately 80% chance of receiving a ‘tick’ response, while chances of acceptance were significantly lower for both colour- and texture-match items. Therefore, participants’ responses were consistent with showing a shape bias (see Figure 9 for observed data). To confirm a shape bias for both groups, post hoc comparisons of each level of SCC were conducted on the model estimates for autistic and TD children. As shown in Table 6, autistic and typically developing participants were both significantly more likely to accept shape-match items compared to non-shape-match items. Furthermore, there was no significant difference in likelihood of accepting colour-match or texture-match items. Thus, both groups showed a clear shape bias in the yes/no task.

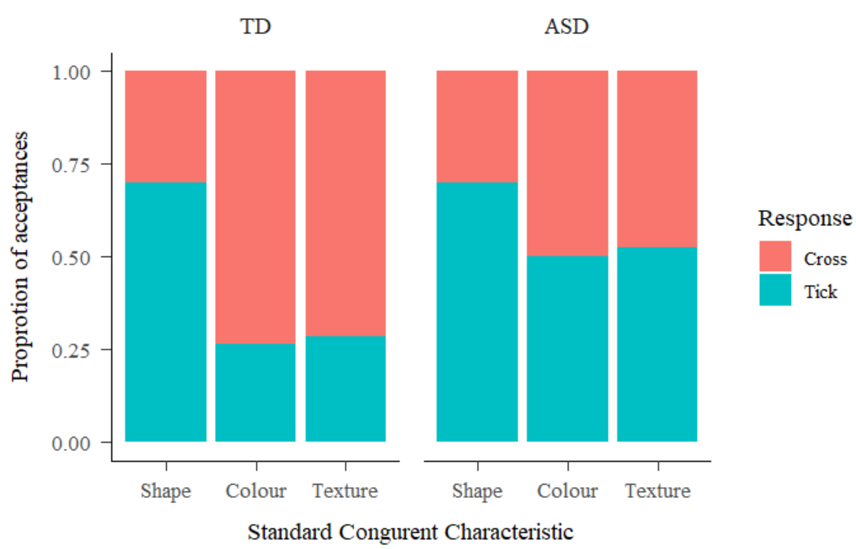


Figure 9: Observed mean proportion of 'tick' and 'cross' responses to shape-match, colour-match and texture-match items by population

Table 6: Post hoc comparisons of probability of accepting items by each level of SCC for TD and autistic participants

	TD		Autism	
Contrast	*z*	*p*	*z*	*p*
Shape/Colour	6.244	<.001	4.018	<.001
Shape/Texture	6.167	<.001	3.882	<.001
Colour/Texture	-0.659	.787	-0.141	.989

2.5.2.3 Effect of population on probability of exhibiting shape bias

The model showed a significant 2-way interaction between population and SCC. To investigate the population differences in the strength of the shape bias, post hoc comparisons were used to compare estimates between the autistic and TD participants. Autistic children had a significantly higher probability of accepting both colour-match items (ASD $M = 0.53$; TD $M = 0.18$; $z = 2.31$, $p = .021$) and texture-match items (ASD $M = 0.54$; TD $M = 0.22$; $z = 2.07$, $p = .039$). Conversely, the probability of accepting shape-match items did not differ significantly between populations (ASD $M = 0.84$; TD $M = 0.76$; $z = 0.68$, $p = .497$). While autistic and TD children were similar when generalising the category by shape, autistic children were also more likely to accept the distractor items that *did not* match on shape (see Figure 10).

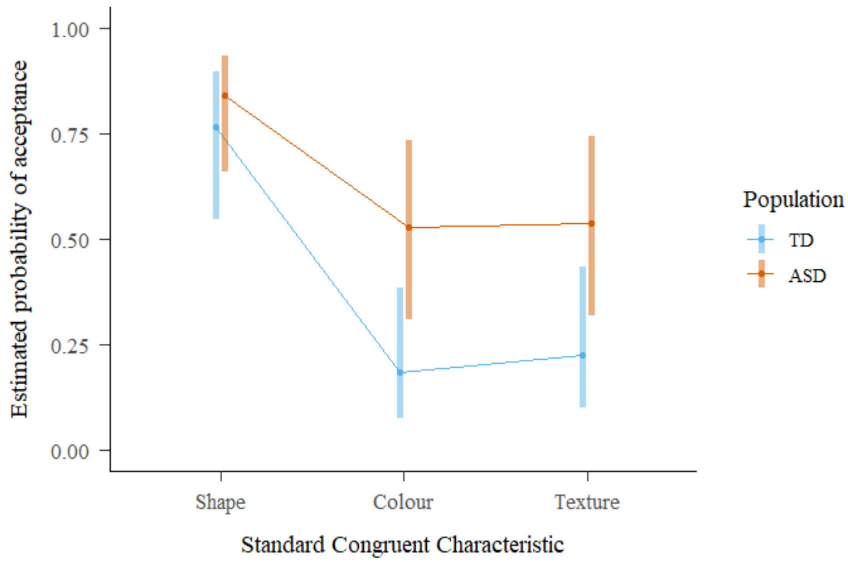


Figure 10: Estimated probability of item acceptance by standard congruent characteristic by population (left). Error bars indicate 95% confidence intervals.

The three-way interaction between SCC, population, and receptive vocabulary equivalent age was also significant. When the SCC was a colour-match to the standard, the odds of accepting the item increased with receptive vocabulary more than expected for autistic participants. The model estimates are illustrated in Figure 11. For shape-match and texture-match items, the relationship between receptive vocabulary and item acceptance did not significantly differ by population: higher vocabulary scores were associated with more shape-match acceptance and more texture-match rejection. Higher receptive vocabulary was only associated with greater odds of colour-match rejection for the TD participants, however. This pattern is consistent with an association between shape bias and receptive vocabulary in typical development, as the direction of the relationship for the different levels of SCC is suggests an increased preference for generalising by shape. For autistic children, however, colour-match items do not clearly follow this pattern.

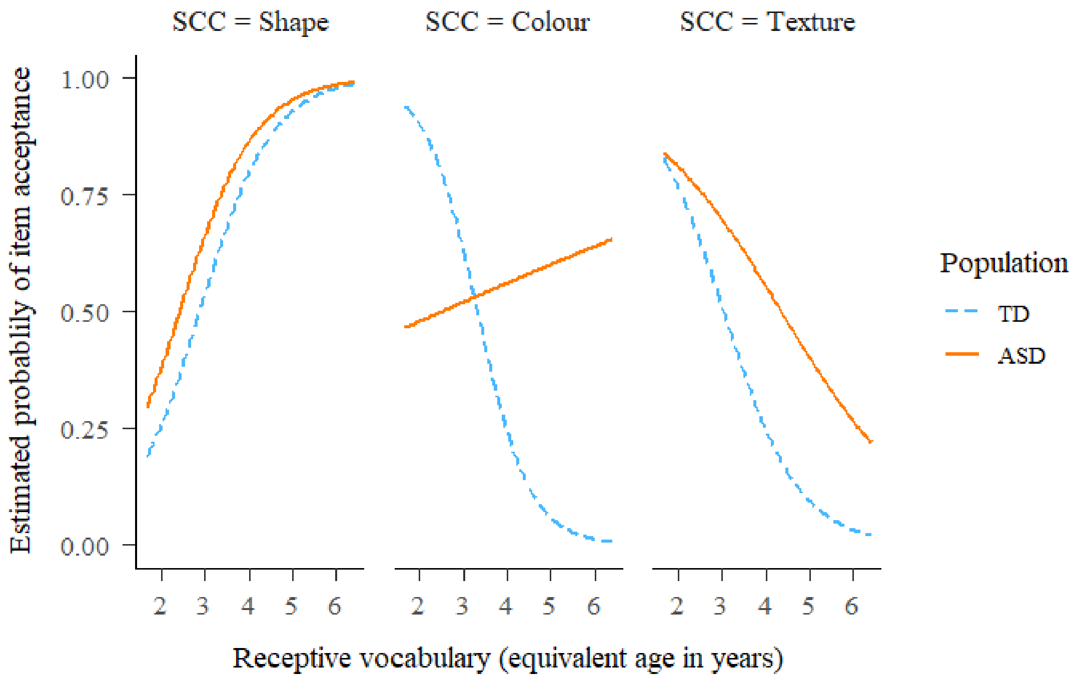


Figure 11: Estimated probability of accepting shape-match, colour-match and texture-match SCC items as predicted by receptive vocabulary by population

2.5.2.4 Effect of standard visibility on probability of exhibiting shape bias

Standard visibility was a significant predictor in the model. The odds of shape-match acceptance were lower in the offline condition compared to the online condition. The interaction term with SCC was also significant, suggesting that the effect of standard visibility differed for the colour- and texture-match items. Post hoc comparisons indicated that the probability of acceptance was only significantly different for the shape-match SCC items in online ($M = 0.88$) and offline ($M = 0.70$) trials ($z = 2.84$, $p = .005$). The comparisons for colour-match ($z = 0.74$, $p = .458$) and texture-match ($z = 0.14$, $p = .893$) revealed no significant difference for standard visibility (see Figure 12). Note that as there was no significant interaction with population, this comparison was conducted on the groups combined. Overall, the probability of accepting shape-match items decreased in the offline condition when the generalisation was based on memory.

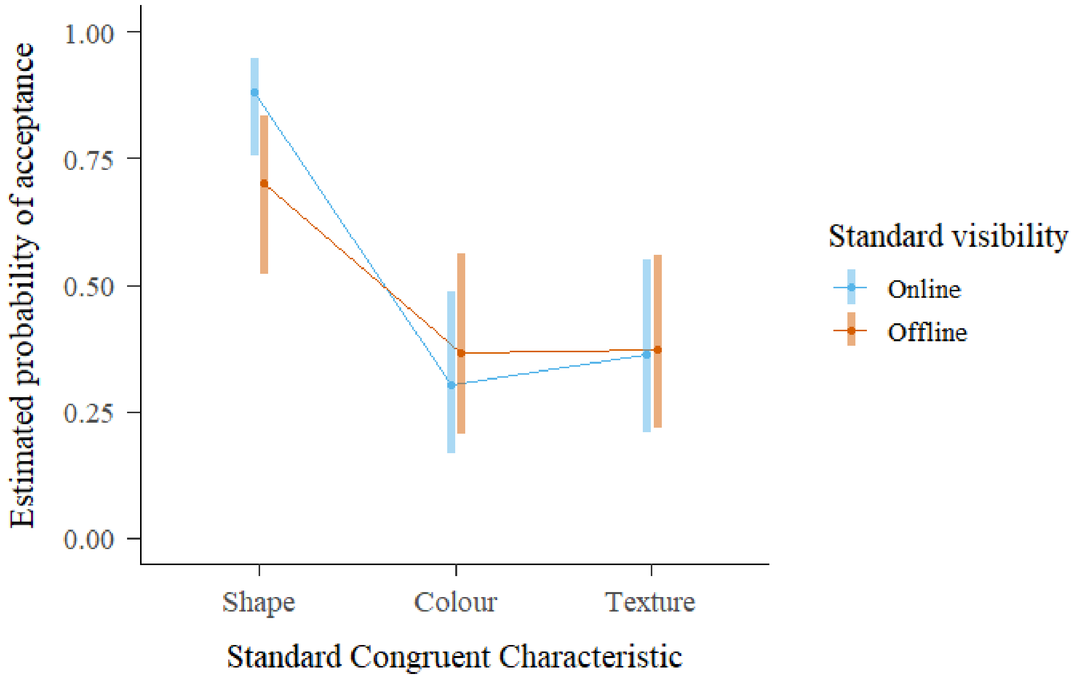


Figure 12: Estimated probability of accepting shape-match, colour-match and texture-match items in online vs. offline trials

The interaction term for SCC x Standard visibility x Receptive vocabulary was also significant in the model. The association between receptive vocabulary age equivalent and item acceptance differed between online and offline trials for texture-match items only. Figure 13 shows receptive vocabulary as less predictive of responses to texture-match items in the offline condition. Overall, higher receptive vocabulary scores were associated with a greater likelihood of shape bias consistent responses for all three levels of SCC: shape-match (i.e. increasing probability of acceptance), and colour-match and texture-match (i.e. decreasing probability of acceptance) for online and offline trials. However, the effect was lessened for texture-match items when the standard was not visible, perhaps because texture-match items were less likely to be accepted even at the younger age equivalents for vocabulary.

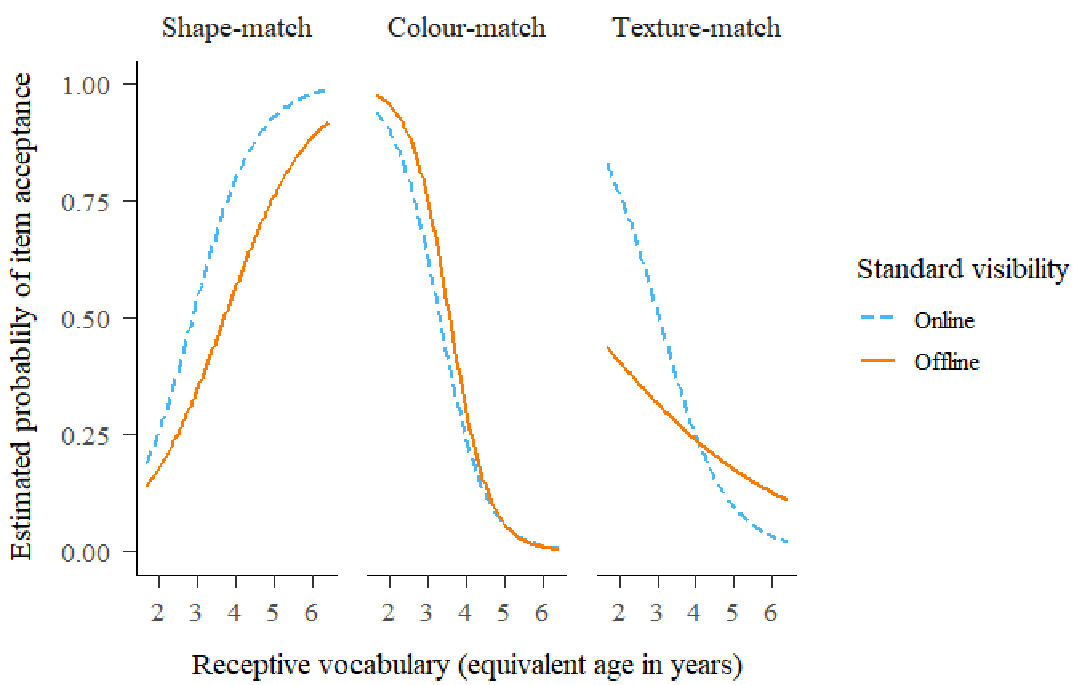


Figure 13: Estimated probability of accepting shape-match, colour-match and texture-match items by receptive vocabulary in online vs. offline trials

2.5.2.5 Effect of labels on probability of exhibiting shape bias

The label condition interacted significantly with SCC. Specifically, the difference in acceptance probabilities for shape-match and colour-match objects was larger for label trials than no label trials, indicating labels were predictive of responses that were consistent with shape bias (Figure 14). However, post hoc comparisons found no significant difference in probability for shape-match items (z = 1.209, p = .227) or colour-match items (z = 1.628, p = .104).

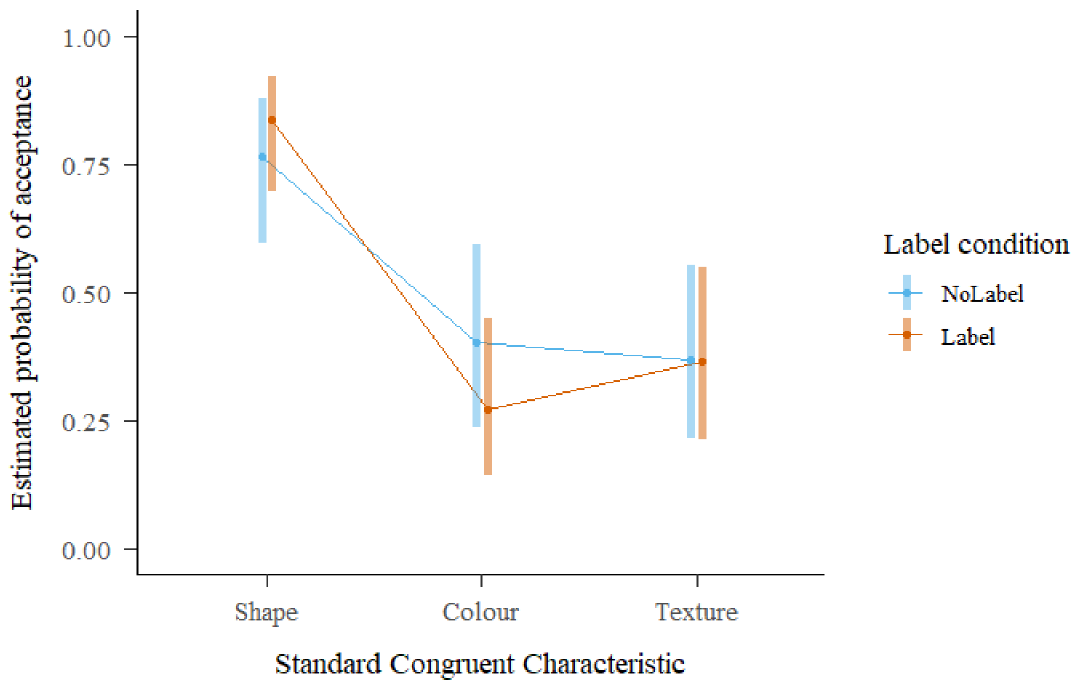


Figure 14: Estimated probability of shape-match, colour-match and texture-match acceptance by label condition

2.5.3 Discussion

Experiment 2 investigated how the presence of labels, and the visibility of the standard, influenced the shape bias in TD and autistic children using a sequential yes/no task. Each test item had one standard congruent characteristic (SCC) - shape, texture, or colour - which was used as a predictor to measure shape bias. While Experiment 1 showed that both groups could generalise by shape, the yes/no task determined if shape was prioritised exclusively, or whether children would generalise by additional characteristics when given the opportunity. We hypothesised that the TD children would be more likely to accept the shape-match SCC items and reject the alternative distractors than the autistic children. An additional hypothesis was that labels and standard visibility would interact with SCC to influence the probability of making a shape bias consistent choice. We found support for an effect of standard visibility, and partial support for our predictions regarding population and labels.

As predicted, TD children were significantly more likely to reject colour- and texture-match distractors than autistic children. However, we found no evidence that autistic children were less likely to generalise by shape than TD children. Both groups had similarly high estimated probabilities of accepting shape-match objects, and both were significantly more likely generalise to a shape-match than a non-shape-match distractor. This finding suggests that autistic children do extend category labels on the basis of shape. However, they may be more reluctant to reject novel distractors and be prone to making incorrect over-generalisations. This was also the case in Hartley and Allen (2014), where autistic children generalised labels to both shape and colour-match items. Our results both replicate the colour finding, and demonstrate that over-generalisation also applies to texture-match items. Overall, our yes/no task suggests that TD children are more likely to exclude an item when the shape deviates from the standard, whereas this is not necessarily the case for autistic children. In other words, the group difference lies in the use of shape as a criterion for category exclusion, not inclusion. It is possible that children with autism are less willing to rule out potential category members on the basis of shape difference alone due to the uncertainly of the category scope from a single example.

For both autistic and TD children, higher receptive vocabulary scores were associated with higher shape-match acceptance. Population had no detectable effect on the strength of the relationship between vocabulary and acceptance probabilities for items that were a shape-match for the standard. For TD children, receptive vocabulary was also associated with lower likelihood of accepting colour- and texture-match distractors. This profile of results resembles a classic shape bias, in that shape-matches are strongly preferred and the alternatives are rejected.

Furthermore, the strength of the bias increases along with receptive vocabulary. This would be expected if autistic children's existing vocabulary supports the development of shape bias, as is thought to be the case for TD children (Perry & Samuelson, 2011; Smith et al., 2002).

The interaction between SCC, population and receptive vocabulary indicated a significant difference between populations for colour-match items only. While autistic children were significantly more likely than TD children to accept texture-match items overall, both for both groups higher vocabulary scores were associated with increased likelihood of rejecting them. The current study included only children whose vocabulary scores were under 6.5 years equivalent, so it could be expected that this accuracy would continue to improve for higher vocabulary levels. For the colour-match items, however, higher receptive vocabulary did not result in higher rejection rates for the autistic participants. Children were more likely to make a 'false positive' inclusion decision when the distractor was the same colour as the standard, regardless of vocabulary scores.

This over-generalisation of colour-match items is consistent with previous findings. In sequential sorting tasks, autistic children have also been found to include same-coloured, non-shape-match objects in a novel category (Hartley & Allen, 2014; Tovar et al., 2020). Those studies did not include texture-match items, so it is possible that this over-generalisation at higher vocabulary levels is specific to the colour characteristic. It has been suggested that colours have a facilitatory effect on processing for autistic participants that is not present in typical development, so may be more easily encoded by autistic individuals (Brian et al., 2003). If there is stronger colour encoding in autism this could be advantageous in situations where the colour is relevant, but equally could be a disadvantage if it draws attention when it is superfluous to the task.

The results revealed two interesting findings relating to standard visibility. Firstly, being able to see the standard only had a significant effect on acceptance probability for shape-match items. While shape-match SCC items had a high likelihood of being accepted in both online and offline conditions, this likelihood was highest when the standard was visible for direct comparison. The second finding was that receptive vocabulary and standard visibility only interacted with SCC when the matching characteristic was texture. The children with younger age equivalent vocabulary scores were less likely to accept texture-match items when they could not see the standard, hence the already low acceptance rates did not decrease much further for higher levels of vocabulary. This was not the case for colour-match items, which had the same relationship with receptive vocabulary for both the online and offline condition. Over-generalisation to texture-match items was only observed for lower receptive vocabularies when the object could be seen. This finding could be explained by a weaker encoding of texture into memory, compared to shape and colour characteristics. This pattern would be consistent with children not encoding texture into memory as well as the other features, so that information may not be as available in the offline condition.

As the offline condition relies on memory for relevant features to successfully categorise, attention differences may interfere with a shape bias in autism. Weak central coherence, a preference for attending to local details over global ones (see Happé & Frith, 2006), has been suggested as a potential explanation for shape bias differences in autism. Oakes and Kovack-Lesh (2007) note that online categorisation requires only attention for direct comparison, whereas offline categorisation relies on encoding. Therefore, any interference to shape bias from a local attentional bias may be more apparent when the task cannot be completed by direct comparison.

In our task, autistic children showed a preference to extend by shape at above-chance levels, while the distractor responses were at chance levels. Thus, they did show a shape bias insofar as shape was preferred over the other potential visual cues. However, as autistic children's responses to the different-shaped distractor items was not as consistent as their acceptance of same-shape items, this suggests population differences making exclusion decisions. If this is reflective of real-life learning situations, it may be challenging for autistic children to make use of that preference for shape to support effective word learning.

2.6 General discussion

This research is the first to systematically investigate the effect of experimental demands on the shape bias in both autism and typical development. We aimed to discover whether conflicting results in the existing literature could be explained by methodological differences, and our results suggest that they can. When a forced-choice method was used, autistic children showed a clear shape bias that was comparable to that demonstrated by TD children matched on receptive vocabulary. In contrast, the yes/no task elicited key differences between the populations; while TD children still demonstrated a strong shape bias, autistic children over-generalised to colour and texture-match items. Furthermore, while the relationship between shape match acceptance and vocabulary was similar across populations and tasks, the link between vocabulary and shape-difference rejection was weaker for autistic children when considering colour-match items.

The difference in findings generated by forced-choice and yes/no tasks has been acknowledged in typical development (e.g. Booth et al., 2005; Landau et al., 1988; Samuelson et al., 2009), and while it is recognised that these tasks answer slightly different questions concerning categorisation (i.e. "which is the best example?" or "is this an acceptable example?"), the shape bias has nevertheless proved to be robust in both. The current research supports this view for TD children. However, this was not the case for autistic children, raising the possibility that these tasks tap into different processes.

Our findings demonstrate that methodological differences may explain the disparate results from previous research investigating the shape bias in autism. Despite the majority of our participants taking part in both experiments, the shape bias was only ‘weaker’ for autistic participants when each item had to be considered individually in the yes/no task. Under these conditions, autistic children were likely to generalise by shape in addition to generalising by colour and texture. This over-generalisation in the yes/no task was also observed in Hartley and Allen’s (2014) sorting task. They suggested this may be due to autistic children equally weighting the importance of available perceptual features for generalisation, as opposed to the weighting shape more heavily as in typical development. While this would explain the results of our yes/no task, it does not explain why autistic children showed a strong shape bias in the forced-choice task. Furthermore, in Tovar et al.’s (2020) replication of the sorting task, they found that autistic participants were also likely to over-generalise to novel objects that did not match the standard on any dimension. The equal feature weighting explanation cannot account for the findings of this subsequent research.

Our findings from both tasks suggest that shape carries a greater weighting than colour or texture during categorisation for both developmental groups, and that this weighting increases with vocabulary size. While over-extension to non-shape-items was apparent in autistic children, shape was still preferred over and above the alternatives. The high rates of shape choice in the forced-choice task suggest that children with autism can, and do, use shape as a primary cue for category membership. In other words, if the question is “which is the best example?” both TD and autistic groups choose the shape-match option. This adds to a growing body of evidence that autistic children can generalise by shape when asked to choose one object from an array (Field et al., 2016b; Hartley et al., 2019, 2020; Tek et al., 2008). Additionally, our findings suggest that the shape bias in autism increases in line with receptive vocabulary, reflecting the same pattern as in TD children. While shape bias may be delayed, as proposed by Field et al. (2016b), it nevertheless appears commensurate with their equivalent receptive vocabulary age.

Our results are in keeping with Tek et al. (2008), in that the shape bias contrasted for our autistic participants in different experimental tasks. However, the intermodal preferential looking task (IPL) used by Tek and colleagues is not directly analogous to the yes/no task. Like the forced-choice task, the implicit question in the IPL is "which is the best example?", with two candidates competing for attention. The commonality with the yes/no task is the freedom not to choose: both items, or neither, may be acceptable answers, and are indicated by a roughly equal looking time at each. If autistic children do give shape a greater weighting, this scenario perhaps ought to invoke greater attention to shape. However, it did not. The autistic children in Tek et al. (2008) were younger than the current study (aged 2 to 3) and were receiving an intensive programme of therapy. It is plausible that the weaker shape bias observed in that case merely reflects the proposed delay. However, the same children chose shape-match items at a greater rate than chance in the forced-choice task. This suggests that, even at this younger age, the children recognised shape as the most important feature when encouraged to make a choice. This preference was not as strong as observed in the current study, however, this is consistent with our finding that the different task types may contribute to the differing shape bias performance for the children.

If the IPL task occupies a space between the yes/no and forced-choice tasks, it could provide an additional piece of the puzzle. The opportunity to choose the best example from a shape-match and a distractor may not be sufficient to encourage attention to shape in autism. It may be necessary to also have no opportunity to choose the distractors. In a forced-choice task, generalisation on the basis of shape can occur without the need to make a decision about the distractors. One of the objects merely has to be 'more likely' as a candidate to win over the competition. However, in the yes/no task and the IPL task, the distractors must also be interpreted as *not* likely category members. For the yes/no task this is explicit in the 'No' response, whereas successful attention to the target in the IPL task at a higher-than-chance rate also requires inhibited attention to the distractor. Thus, deciding what things *are not* may be key to the differences in shape bias for TD and autistic children.

How children decide whether a candidate is *not* a new category member has not been extensively discussed within the shape bias literature. In models of children's responses to a yes/no task, 'reject' decisions have been represented as the absence of an 'accept' decision (Samuelson et al., 2009). So, the finding that some children do accept shape as an important criterion for category inclusion, while simultaneously accepting highly different shapes as potential members of the same category, was an unanticipated finding. It also suggests the need to consider the possibility that shape bias serves a dual function: facilitating both category inclusion and category exclusion. If so, the differing use of shape for autistic and TD children may be exclusive to the task of identifying what things are *not*.

It has been proposed that heightened attention to small perceptual details at the expense of holistic meaning (Happé & Frith, 2006) may contribute to a weaker shape bias in autism, as attention gets drawn away from the overall shape towards other features (Tek et al., 2008).

However, in both the current experiments, autistic children did not significantly differ from the TD group in their acceptance of shape-match objects. This provides no evidence that a local detail processing bias interfered with perception or encoding of global shape. Even when the target had to be held in memory briefly, shape acceptance remained high. If local attention disrupts generalisation, it may only do so in cases of category exclusion. Increased attention to colour and texture, while retaining attention to shape as the most important cue, may account for over-generalisation in our yes/no task.

Autistic children's increased attention to multiple features would fit with the 'Enhanced Perceptual Functioning' model (Mottron et al., 2006), which proposes that some selective attentional processes are not automatic in autism because they do not need to be. Enhanced perception means that autistic children can possibly afford to be less economical with their attention, resulting in additional information being processed and encoded. For instance, autistic participants have demonstrated enhanced ability to discriminate between stimuli (Plaisted et al., 1998), and benefited from stronger encoding of colour in cases where that information was inhibited for neurotypical participants (Brian et al., 2003). Remington et al. (2009) found evidence for a greater capacity to process perceptual information in autism, though not without limits. They demonstrated that autistic adults could ignore distractors in the visual field when the perceptual load was sufficiently high, suggesting that extraneous information is only processed when there are available resources to do so.

While an enhanced perceptual capacity may be an advantage in many situations, reduced inhibition of irrelevant perceptual information can also be a source of distraction (Adams & Jarrold, 2012), which could account for the different task performance in the current study. If there are sufficient perceptual resources to attend to multiple features, extraneous information could also be encoded without interfering with the chances of choosing shape, which is recognised as the most predictive feature. In typical development, alternatively, irrelevant features may be inhibited during encoding and given less attention in the first place. In a forced-choice task, both of these possibilities would result in the same outcome - a preference for shape as the most predictive feature - as seen in the current study. For the yes/no task, however, reduced inhibition of task irrelevant features could cause interference. The current findings are consistent with a pattern of distractor inhibition in typical development, with any item that is not a shape-match likely to be rejected as a category member. With an enhanced perceptual capacity, autistic children may be more reluctant to rule out possible category members on the basis of shape simply because they do not need to be so conservative with their attentional resources. This could allow them to maintain richer mental representations of the item that include colour and texture, or enable them to consider multiple potential theories about the categorisation ‘rule’ until they have more information. This was also suggested as an explanation for the findings of Hartley et al. (2019). Autistic children exhibited superior shape-based generalisation to TD children after a 5-minute delay, which suggests a better memory for the object being generalised from. Reluctance to exclude items children are unsure about could also account for the equal looking times found in Tek et al. (2008), as children may encode both as potential category members. With increased experience, autistic children may then learn to rule out distractors through top-down word learning constraints.

2.6.1 Limitations

One limitation of the current research is the smaller-than-planned sample size, due to data collection being interrupted by school closures during the Covid-19 pandemic. While small sample sizes are not uncommon in developmental research, this is recognised as an issue leading to underpowered studies that fail to replicate (Davis-Kean & Ellis, 2019). This issue is mitigated to a certain extent by analysing the binary response to each trial rather than an aggregated score for each participant, increasing statistical power by modelling more data points. We also limited the model to the pre-registered interactions. However, replication of these findings with a larger sample size is recommended.

2.6.2 Conclusion

In conclusion, across two studies with contrasting methodologies, we observed that a sample of autistic children prioritised shape as the most important feature for category inclusion, generalising novel labels in a manner that is consistent with having a shape bias. Furthermore, the strength of shape bias was appropriate for the children's receptive vocabulary age equivalent. However, by altering the task demands so that participants had to evaluate distractor items in a yes/no task, autistic children did not consistently use shape to inform their category exclusion decisions as the TD participants did. This suggests that shape bias may play a dual role in categorisation, influencing both inclusion and exclusion decisions, but in different ways. Furthermore, the finding that task demands have differing impacts on the presentation of shape bias for TD and autistic children has implications for how the investigation of shape bias is approached. Research to date may underestimate the shape bias of autistic children due to the choice of task. If these children can utilise the bias for word learning given the ideal circumstances, this has implications of what teaching strategies are likely to be most effective.

Chapter 3: Shape bias for perceptually similar stimuli during 'online' and 'offline' novel noun generalisation for autistic and typically developing children

3.1 Introduction to Chapter 3

In chapter 2, we found that autistic children exhibited a shape bias that was indistinguishable from typical development when asked to generalise using a forced-choice NNG task. However, in the yes/no task that included many of the same children, shape bias was disrupted. We argued that differences between the tasks could be explained in terms of shape-based category *exclusion* decisions, as shape-match inclusion decisions were equally high for both groups. Contrary to our predictions, we failed to find an effect of labelling or an interaction between population and standard visibility on the strength of shape bias. In this chapter, we explore the possibility that task demands impact children's shape bias when they are asked to categorise stimuli sets that are perceptually similar. We suggest that these 'low contrast' stimuli are more likely to reveal differences in shape bias across the standard visibility conditions, as children could make the plausible assumption that they are all members of the same superordinate category. Furthermore, increasing the similarity of stimuli should consequently increase the difficulty of identifying perfect shape-matches from memory in the offline condition, and give us a better opportunity to observe population differences if they are present.

Author contributions:

Leigh Keating: study design, data collection, statistical analysis, manuscript writing, review.

Calum Hartley: study design, review. *Katherine Twomey*: study design, review.

3.2 Abstract

There is emerging evidence that autistic children are able generalise novel nouns by shape-similarity just as well as typically developing (TD) children, however, these populations may differ in their use of shape-differences when ruling out potential category members. This research investigates whether task demands influence the strength of 'shape bias' in forced-choice (experiment 1) and yes/no (experiment 2) novel noun generalisation (NNG) tasks for autistic and TD participants when stimuli are perceptually similar enough to belong to the same broad category. The effect of naming the 'standard' for the category, along with standard visibility for direct comparison during generalisation, were varied to test their impact on shape-based inclusion and exclusion decisions. In experiment 1, effects of labelling and standard visibility significantly differed between populations. Autistic children exhibited a weaker, but not absent, shape bias in the 'offline' condition compared to online trials where the standard was visible. For TD children, labels enhanced the strength of shape bias. In experiment 2, only the TD children demonstrated a shape bias, which was weaker in the offline condition. Autistic children accepted items regardless of which feature they had in common with the standard. This suggests that while autistic children prioritise shape for category inclusion, labelling may not spontaneously cue their attention to shape. Less efficient attentional processes could also account for the benefit of having visible examples for reference, which has implications for ideal word learning conditions in autism.

3.3 Introduction

For typically developing (TD) children, the ability to quickly learn what a new word refers to is an important skill, supported by a range of biases (Bloom, 2000). One of the earliest to develop is the 'shape bias' (Landau et al., 1988; Smith et al., 2002). By the time children have learned their first 50-150 count nouns (Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999), TD children usually generalise newly encountered labels to novel objects based on similarity of shape, rather than other perceptual commonalities such as the colour or texture (e.g. Landau et al., 1988; Smith et al., 2002). The shape bias is particularly useful for novel count nouns, which are well-organised by shape in English-speaking children's early vocabularies (Samuelson & Smith, 1999), and may support faster vocabulary growth by allowing infants be more efficient with their attentional resources (Smith et al., 2002). This opens the possibility that differences in the use of the shape bias could contribute to language delays, which commonly occur in autistic children (Anderson et al., 2007).

While several studies have found differences in autistic children's shape bias (Abdelaziz et al., 2018; Field et al., 2016b; Hartley & Allen, 2014; Keating et al., Chapter 2; Potrzeba et al., 2015; Tek et al., 2008; Tovar et al., 2020), there is emerging evidence that these differences are not consistent across experimental tasks (Keating et al., Chapter 2). Both autistic and TD children exhibit shape bias in novel noun generalisation (NNG) tasks that present a forced-choice scenario between competitors (Field et al., 2016b; Keating et al., Chapter 2; Tek et al., 2008). However, when given a task that requires a categorisation decision on each item individually, autistic children are prone to over-generalise and include items of a different shape (Hartley & Allen, 2014; Keating et al., Chapter 2; Tovar et al., 2020), and so shape bias appears disrupted. This suggests that the shape bias is not something that children do or do not have, but rather a tool that may or may not be used depending on the task at hand. The question to be asked then is 'under what circumstances do autistic children show a shape bias?' In this research we investigated whether shape-based generalisation is affected by labelling for autistic children when the stimuli are similar enough to belong to a common superordinate category. Additionally, we investigated impact of having a visible standard during generalisation for the different populations, as autistic and TD children may remember different information about the object when direct comparison is not possible.

TD children commonly exhibit a shape bias in variety of NNG tasks. These tasks always begin by introducing the child to a novel object, which may be uniquely designed for the study or be an existing item that is unfamiliar. This object represents the 'standard' for a hypothetical category, and may be presented with a label (e.g. 'look at the dax') or an unlabelled direction (e.g. 'look at this one'). Different versions of the NNG task offer different methods for children to indicate which 'test items' they would generalise the label to (or include in the unlabelled category) from a set of stimuli that vary from the standard in a controlled way. In a forced-choice NNG, this involves choosing one item from an array of two or three items that may vary in shape, texture, colour, or size (e.g. Booth et al., 2005; Diesendruck & Bloom, 2003; Landau et al., 1988). Alternatively, in a yes/no NNG task, each test item is presented individually and children are asked if each one is another example of the standard when there are no competitors present (Landau et al., 1988, 1992, 1998; Smith et al., 1992, 1996). For items that appear to be rigid artifacts (see Jones et al., 1991; Samuelson & Smith, 2000a for biases for alternative noun classes), TD children prefer to generalise by shape over other perceptual features in both forced-choice and yes/no versions of the NNG task, indicating that shape is seen as the most informative cue to category membership in the absence of any conceptual information to draw on.

While TD children prioritise shape both when generalising labels and in trials where the standard is unnamed, there is evidence that shape bias is stronger when items are labelled, particularly during earlier stages of language development (Landau et al., 1988; Smith et al., 1996, 2002). Consequently, it has been suggested that shape bias develops as infants detect correlations between object names and object shapes for count nouns in their early vocabularies, and so allocate more attention to this feature when new nouns are encountered (Smith et al., 2002). This attentional pull towards object shape in labelling contexts is proposed to be an automatic mechanism rather than a conscious strategy (Smith et al., 1996). Adults and older children, however, appear to consistently use shape as a basis for generalisation even for unnamed categories (Field et al., 2016b; Landau et al., 1988). Additionally, more mature learners can override the shape bias if there is available information that might be more relevant for category membership, such as object function (Smith et al., 1996). However, there are still circumstances where automatically allocating attentional resources to object shape could provide an advantage. Studies where TD children have shown a stronger shape bias for labelled items tend to use test items that have some overlap of shape with the standard (Landau et al., 1988; Tek et al., 2008). Conversely studies that have found no effect of labels with TD children used test items and standards with highly distinctive shapes (Field et al., 2016b; Keating et al., Chapter 2).

In our previous study (Keating et al., Chapter 2), we investigated shape bias in autism and TD with both a forced-choice and yes/no variant of an NNG task across two experiments. Each experiment also included a label manipulation (with label and no-label trials) and a condition for ‘standard visibility’, which varied whether or not an example of the standard was on screen for comparison while children were responding (online or offline conditions). We found no effect of labels for either the autistic or TD children, however our study specifically used stimuli that were 'high contrast'. We suggested that an effect of labels may not be observed if there is enough difference between the shapes to easily differentiate them without the benefit of automatic attention. To compare shape bias for what they termed 'high-similarity' and 'low-similarity' stimuli, Tek et al. (2012) used an intermodal preferential looking (IPL) paradigm. This method is intended to provide an implicit measure of shape bias, inferred from the proportion of time children look at one screen displaying a test object over another. They found that when objects where labelled, TD children exhibited much stronger shape bias for objects that were 'low' in shape similarity. However, when shape similarity was 'high', children looked more to an 'overall match' (i.e. the same colour and material as the standard along with some similarity in shape) compared to a differently coloured, perfect shape-match competitor. These findings do not support our suggestion that the label effect could be greater for stimuli that are more similar, however, this may be explained task differences. The IPL paradigm is an offline task requiring label generalisation to be completed based on a representation of the standard held briefly in working memory. By contrast, in an online task the standard is visible throughout the trial enabling direct visual comparison with test items (e.g. Jones et al., 1991; Landau et al., 1988; Smith et al., 2002). The children in Tek et al. (2008) were 24-months-old on average, and there is evidence that children this age do not remember the shape of an object any better than chance, despite being able to remember the original object itself (Perry et al., 2016). Therefore, at this young age, labels may provide support for the newly developing shape bias in offline tasks in a way that was not necessary for the older participants in Keating et al. (Chapter 2). In that study we also included 'online' and 'offline' generalisation conditions by manipulating whether the standard was visible for direct comparison while children were responding to the test items, and found evidence for weaker shape bias overall in the offline condition. Interestingly, this effect was only significant in the yes/no task, and only for items that were a shape-match, suggesting that exclusion decisions for different-shaped items were unaffected.

While the strength of the shape bias in typical development can be influenced by labelling and stimuli variability, a preference for shape is nevertheless robust across different NNG tasks involving rigid artefacts. For autistic children, however, their responses are not as consistent and may be disproportionately affected by methodological differences. In a variety of tasks, autistic children appear to have a weaker, or delayed shape bias in comparison to TD children. In Keating et al. (Chapter 2), autistic children were prone to over-generalising labels to non-shape match items in a yes/no NNG task, while TD children were likely to exclude them. The same findings have been observed in sequential sorting tasks, which are similar to the yes/no NNG in that children are presented with items individually and asked to indicate whether they are referents for a newly-heard word. In this task, autistic children are likely to generalise novel labels to unfamiliar objects regardless of shape similarity (Hartley & Allen, 2014), including items that do not match the standard on any dimension (Tovar et al., 2020). Evidence from IPL tasks with autistic children also indicate a disruption in shape bias (Abdelaziz et al., 2018; Potrzeba et al., 2015; Tek et al., 2008). However, these tasks that have identified differences in autistic children's shape bias are non-competitive; a single test item is presented and children must make an accept or reject decision on each one individually.

However, when given a forced-choice task, there is strong evidence that autistic children can generalise by shape just as accurately as typically developing children (Field et al., 2016b; Hartley et al., 2019, 2020; Keating et al., Chapter 2; Tek et al., 2008). When Tek et al. (2008) presented their stimuli in a forced-choice NNG, the autistic participants chose the shape-match a higher proportion of the time than the alternative. Using highly contrasting stimuli in a forced-choice task, Field et al. (2016b) also found a shape bias for autistic participants. In our previous research, (Keating et al., Chapter 2), we presented the same stimuli sets in both yes/no and forced-choice NNG tasks with nearly identical samples of autistic children. While their responses were indistinguishable from the TD participants in the forced-choice task, and showed a strong shape bias, there was notable disruption in the yes/no version. Like the TD children, autistic children were significantly more likely to accept shape-match items than texture- or colour-match items. However, they were also significantly less likely to reject the distractors than the TD group. Furthermore, we found that children's shape-based responses overall were affected by the ability to see the standard, but only for shape-match items. Yes or no decisions on non-shape-match items did not change between online and offline conditions. We argued that this discrepancy could be explained by a difference in the use of shape when making *exclusion* decisions, but not *inclusion* decisions. The similar shape preference exhibited by TD children in both tasks may indicate the use of shape similarity as cue for category inclusion, but additionally the use of shape difference to rule out test items and inform their category rejection decisions. However, for the autistic participants, the disparity in shape bias between the two tasks suggested that rejecting items that were not a shape match was more of a challenge, resulting in over-generalisation.

A further difference in shape bias for autistic children is that labels may not support attention to shape in this population to the same extent as for TD children. The majority of research to date has failed to find evidence for a label effect with autistic participants (Keating et al., Chapter 2; Potrzeba et al., 2015; Tek et al., 2008). In the one study reporting that autistic children showed a stronger shape bias in labelled trials (Field et al., 2016b), TD participants did not. In this forced-choice NNG task with high contrast stimuli, TD children chose shape-match competitors with such high consistency in both labelled and unlabelled conditions that there was no difference between them. In this regard, the performance of the autistic children was still 'atypical' despite them demonstrating a shape bias. With the exception of this anomalous finding, the label effect in TD children may be explained by automatic attention to shape in a naming context (Smith et al., 1996). However, attentional differences in autism, such as 'weak central coherence', could disrupt this process (see Happé & Frith, 2006). While neurotypical individuals may pay attention to 'global' shape first in a perceptual task (Navon, 1977), autistic children show an attentional bias towards local details (Plaisted et al., 1999). This could affect children's responses on shape bias task if their attention is drawn away from shape towards the details like the texture. This raises the possibility that labels may not facilitate attention to shape in autism in the same way that they appear to in typical development.

The heterogeneity in extant evidence highlights the importance of considering the contexts in which autistic and neurotypical children use a shape bias to inform their categorisation decisions. However, there are many questions unanswered, particularly concerning the influence of labelling on the shape bias in autism. The current research aims to address these questions by following the same procedure as Keating et al. (Chapter 2), except using 'low contrast' stimuli sets. Having previously demonstrated that task demands can impact autistic children differently, the primary objective of this study was to test our hypothesis that the relative 'ease' of distinguishing and categorising 'high contrast' stimuli reduced potential population differences, plus effects of labelling and standard visibility. Varying these task demands will allow us to explore whether there are group differences in the strength of the shape bias as a result of labels, and whether this indicates underlying differences in how autistic and TD children use shape for generalisation.

In experiment 1, we used a forced-choice task to investigate whether autistic children and TD children matched on receptive vocabulary were sensitive to the presence of labels when the generalisation task involved selecting the best example of the category from a choice of three potential referents. We included an 'online' block, where the standard was visible at all times, and an 'offline' block, where the standard was removed prior to presenting the test items. Given the lack of evidence for an effect of labelling in autism to date, we hypothesised that TD participants would show a greater benefit from items being labelled. For the effect of standard visibility, we tested two contrasting hypotheses that were both compatible with the potential for enhanced processing in autism (Mottron et al., 2006). Hypothesis 1 was that shape match choices would be more likely for autistic children in the offline condition, as enhanced perceptual processing may result in a stronger memory for object shape and give an advantage over TD participants. Conversely, autistic children may encode a greater amount of less relevant detail that could distract from shape. Hypothesis 2, therefore, predicts shape-match choices will be less likely for autistic participants in the offline condition. While our previous findings suggested that standard visibility makes a difference to shape bias overall, but did not detect population differences, we hypothesised that low contrast stimuli could reveal underlying differences that the high contrast version did not. This research will further our understanding of how task demands affect autistic and TD children's use of attentional biases, and perhaps highlight which demands disadvantage autistic children in ways that could have implications for their real-world learning.

3.4 Experiment 1: Does standard visibility impact shape bias in autism for low contrast objects in a forced-choice task?

3.4.1 Method

3.4.1.1 Participants

Participants were 14 autistic children aged between 5 years 1 month and 9 years 1 month (*M* age = 7.20 years; *SD* = 1.23 years), and 13 TD children aged between 2 years 9 months and 4 years 8 months (*M* age = 3.80 years; *SD* = 0.53 years). The children were recruited from specialist schools, mainstream primary schools and preschools in the North West of England. Autistic participants had previously received a diagnosis by a qualified clinical or educational psychologist, using standardised diagnostic measures and professional judgement. The groups were matched on receptive vocabulary as measured by the British Picture Vocabulary Scale (BPVS), 2nd Edition (Dunn et al., 1997). The BPVS age equivalent scores for the autistic participants (*M* = 3.85 years; *SD* = 1.33 years) and TD participants (*M* = 4.35 years; *SD* = 0.80 years) did not significantly differ, t (21.63) = 1.16, p = .259, two-tailed. Additional standardised measures were taken for expressive vocabulary, using the Expressive Vocabulary Test (2nd Edition; EVT; (Williams, 2007), and non-verbal cognition, using the raw score from the Leiter-3 cognition battery of subtests (Roid et al., 2013). Due to Covid-19 interrupting testing, several participants in each population did not have the opportunity complete these measures (EVT: 7 autistic and 8 TD; Leiter-3: 9 autistic and 8 TD). Available data is reported in Table 1.

To ensure the groups had comparable receptive vocabulary levels, an equivalent age score of 6.5 years was established as a cut of for inclusion in the current analysis. This was due to Covid restrictions preventing recruitment of additional TD participants to match the ability of the autistic group. This resulted in data for 8 autistic children who completed this task being excluded from the final analysis. Additional data exclusions were for inability to complete the BPVS (*n* = 3 autistic) and not completing all experimental conditions (*n* = 1 TD). The task results for these children remain available in the data set on OSF for future analysis: https://osf.io/sgcex/. This research was approved by the Lancaster University FST Research Ethics Committee (REF: FST18017) and a signed consent form was obtained from each participant's parent or guardian prior to them taking part in the study. Children received sticker rewards throughout the session and chose a book to take home at the end of the final session. Educational settings were given an Amazon gift card as a thank you gesture.

3.4.1.2 Design

The experiment was a mixed (2x2x2) design. Population was the between-participants factor with 2 levels (typical development/autism). There were also 2 within-participants factors, each with 2 levels: Standard visibility (Online/Offline) and Labels (Label/No-Label). Standard visibility was counterbalanced, with half of the participants completing the online condition first.

All participants experienced the no-label trials first to reduce the risk of interference from the label trials. The rationale was that if labels provide additional cues to generalise by shape and to form basic level categorises, the presence of labels on earlier trials could influence children's strategy on non-labelled trials.

Of the children who took part in experiment 1, 24 of them also took part in experiment 2 (12 TD and 12 ASD). Task order was also counterbalanced, so half of the participants completed the forced-choice task first. Participants who also completed experiment 2 did so on a different day, typically between 1 to 7 days apart.

Table 1: Sample description for Experiment 1 and 2 (standard deviation and range in parentheses)

Population	Study	N	Chron. Age (years)	BPVS age equiv. (years)	EVT age equiv. (years)	Leiter-3
TD	Experiment 1	13	3.80 (0.51; 2.75-4.67)	4.35 (0.80; 3.00-5.42)	4.79 (0.68; 3.75-5.50)	46.83 (8.98; 34-62)
	Experiment 2	14	3.69 (0.51; 2.75-4.67)	4.14 (0.86; 3.0-5.42)	4.79 (0.68; 3.75-5.50)	46.64 (8.94; 34-62)
ASD	Experiment 1	14	7.19 (1.23; 5.08-9.08)	3.85 (1.33; 1.67-6.42)	3.79 (1.19; 2.17-5.58)	38.65 (13.38; 24-56)
	Experiment 2	16	7.18 (1.46; 4.58-9.58)	4.25 (1.37; 2.5-6.42)	3.86 (1.14; 2.17-5.58)	38.85 (13.26; 24-56)

3.4.1.3 Materials

3.4.1.3.1 Visual stimuli

The stimuli for this study were images of novel 3D objects, digitally created using Blender graphical modelling software (Blender Development Team, 2017). Each set consisted of 15 objects: one standard, four colour-matches, four texture-matches, and four shape-matches, plus two additional distractors that did not match the standard on any dimension. To create perceptual commonalities between the object shapes, the same base shape was used for each item in a set (torus, wedge, cylinder, cone). In Blender, the base shapes were manipulated to create three models with distinct shape variations, while maintaining overall similarity. One model was selected as the standard, and the other two were used as non-shape-match test items. Colour and texture combinations were applied to the base models to create four groups of three test items, each containing one colour-match, one texture-match and one shape-match to the standard (see Figure 1). A total of four stimuli sets were created for the experiment.

Images of known items were used for practice trials and a warm-up task. These were all generated using Microsoft Paint 3D using available images from the 3D library. Stock images used were: a green bed, a blue butterfly, a gold trophy, a brown horse, a yellow banana, and an orange basketball. An animated video clip of the bed rotating was created in MS PowerPoint.

Figure 1: Example stimuli set for the forced choice task.

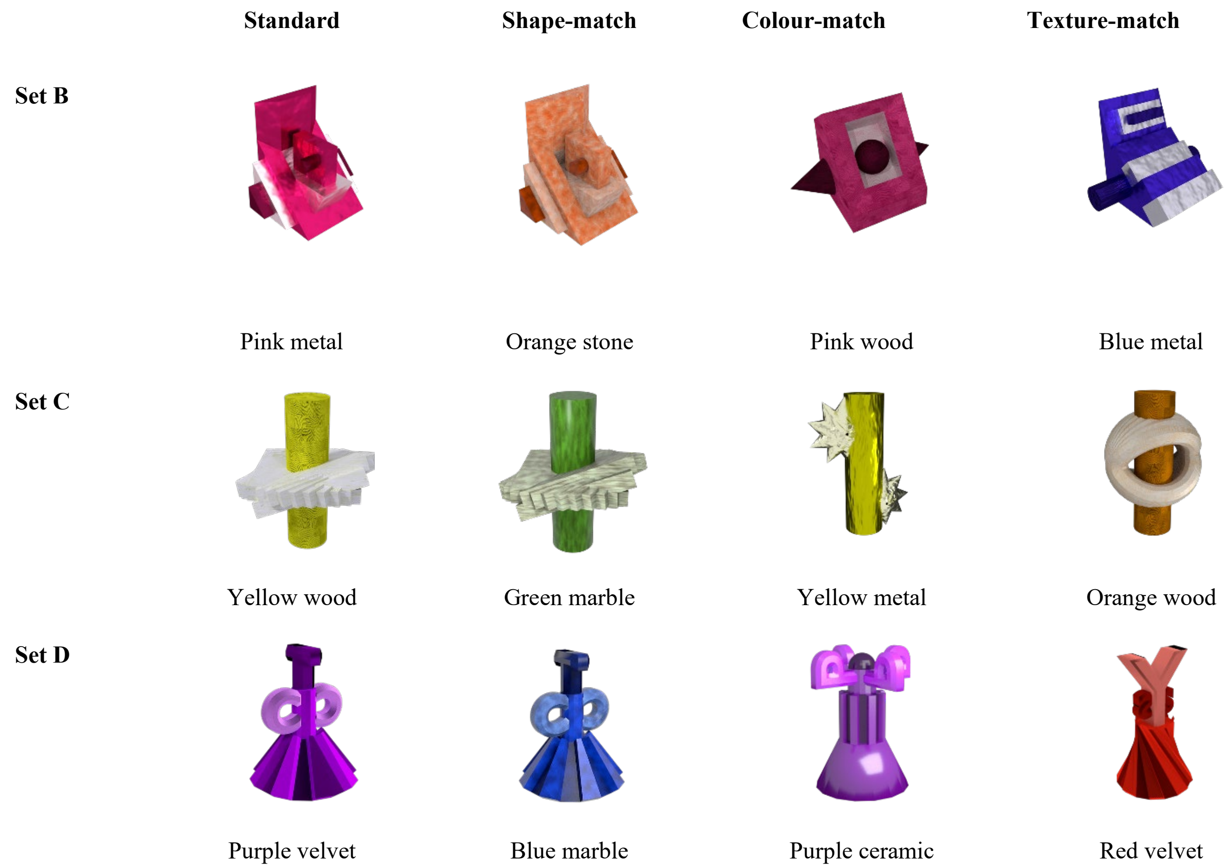


Figure 2: One example trial from stimuli sets B, C and D illustrating which colour and texture combinations were applied.

3.4.1.3.2 Audio stimuli

Audio stimuli were recorded using Audacity® (Version 2.1.3; Audacity Team, 2017) and featured a British, female voice using child directed speech. The scripts of the vocal recordings are detailed in Table 2. For the no-label versions, the recordings were the same as those used in Keating et al. (in prep; see Chapter 2). The novel words 'vink' and 'zain' were both chosen from the English Lexicon Project (Balota et al., 2007) as plausible novel words with an equal number of syllables.

Table 2: Scripts for pre-recorded audio stimuli for forced-choice task

Wording	Phase/condition used in
"Look at that!"	Familiarisation phase: no-label
"Look! A vink!"	Familiarisation phase: label
"Look! A zain!"	Familiarisation phase: label
"Can you find another one?"	Test phase: no-label
"Can you find another vink?"	Test phase: label
"Can you find another zain?"	Test phase: label
"Perfect"	Practice phase only
"Uh oh! Try again."	Practice phase only

3.4.1.3.3 Presentation software and hardware

The experiment was created and presented using PsychoPy2 (Peirce, 2007), an open source Python-based experiment handler, running on a Surface Pro 4 tablet PC. This was presented as a 'game' to the children, and included five mini-games each relating to an experimental task in a large project. Two of the tasks are described in a previous paper (Keating et al., Chapter 2) and the final task is to be described in an upcoming work. Stimuli sets A-D were randomly allocated to online/offline and label/no-label conditions by the programme. These pairings were stored and held constant for each participant across experiments 1 and 2.

The programme controlled the timings and display of all practice and experimental trials. It also recorded the participant's choice of item based on their response on the touch screen. Figure 3 shows a typical layout of the screen while waiting for a response. 'Tapping' one of the three items in the test window would be logged as a response, whereas tapping anywhere else on the screen had no effect.

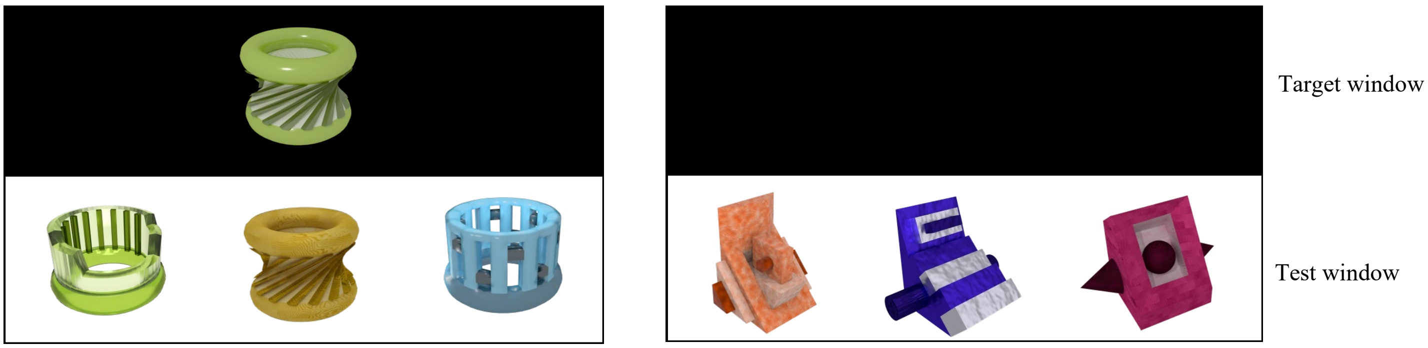


Figure 3: Example of the visual display on screen during an online trial (left) and offline trial (right)

3.4.1.3.4 Button training presentation

An interactive MS PowerPoint presentation was created for a warm-up task. This consisted of six slides and used the same layout at the experimental tasks. The first slide showed an image of a banana in the centre of the test window, situated in the bottom half of the screen. A red circle with a white cross was positioned to the left, while a green circle with a white tick was positioned to the right. The top half of the screen displayed the word "Shoe?" in text. The red 'cross' button was set as the trigger to move to the next slide. The following five slides showed the picture-word pairings: horse – "Horse?", ball – "Ball?", horse – "Train?", banana – "Banana?", ball – "Elephant?". Only clicking the correct button ('tick' for matches, 'cross' for mismatches) triggered transition to the next slide.

3.4.1.4 Procedure

Testing took place in the child's usual educational setting, in a separate room or a quiet corner of their classroom. A familiar adult was invited to accompany them for the duration of the session. Usual procedure was to begin the session with a standardised test (BPVS, EVT-2 or Leiter-3), then complete the computer-based tasks. This experiment was one of five that children could complete as part of the overall project and could be completed on any visit dependent on the counterbalanced order they had been assigned.

3.4.1.4.1 Warm-up task

Before completing the computerised experimental task, children completed a short training exercise to familiarise them with the response buttons used for the yes/no task. First, the experimenter explained the purpose of the task to the child (e.g: "In a few minutes we're going to play a game on the computer. It's going to ask you some questions which you need to answer 'yes' or 'no' to. So, before we start, I'm going to show you the buttons you need to press to answer and we can practice with them together. OK?"). The experimenter then started the button training PowerPoint presentation on the Surface Pro, and either set it in a standing position on a table in front of the child or let them hold it in their lap. In both cases, the Surface Pro was kept in a tablet configuration (i.e. with no keyboard attached) for the duration of the session. With the first slide on display, which showed a banana, the experimenter asked "Is this a horse?". When the child indicated 'no' with a verbal response or a head shake, the experimenter agreed and asked them which button they thought meant 'no'. If they chose appropriately, the experimenter gave verbal praise and invited them to tap the button on the screen. If they chose incorrectly, the experimenter corrected them and encouraged them to try the other button. The correct response triggered the slideshow to progress to the next slide, which modelled a 'yes' response with a correct pairing. If children were unable to give correct responses on three consecutive slides, the presentation was restarted from the beginning. All children were able to meet this standard by the end of their second attempt.

3.4.1.4.2 The Forced-Choice NNG task

Once they had successfully completed the warm-up phase, children were told that they were going to play a game on the computer with some funny things they might not have seen before. The experimenter initiated the game, which controlled the procedure for the rest of the experiment as follows.

The experiment began with a practice block for either the online or offline condition, using known items to demonstrate to children what the task involved and how to input their responses. Both conditions had the same familiarisation phase, in which participants saw an animation of a green bed rotating on the screen accompanied by an audio track that said “Look at that!” Once the audio had finished, a ‘next’ button appeared allowing children to skip to the next phase, otherwise the animation repeated for 30 seconds before moving on automatically.

In the online condition, a static image of the same bed was displayed in the centre of the target window in the top half of the screen. After 1 second, an audio track asked “Can you find another one?” and images of an identical bed, a blue butterfly, and a gold trophy were displayed in the test window below. Whichever item the child touched was highlighted with a green circle, regardless of accuracy, as visual confirmation of their choice. If they chose the bed, audio feedback (“Perfect!”) confirmed the correct response and the programme progressed to the experimental trials. If they chose one of the distractors, an audio track played “Uh oh, try again!” and children were allowed to enter another answer. This repeated for each incorrect response until the correct item was selected. Tapping anywhere else on the screen, including the standard, had no effect.

For trials in the offline condition, the procedure was identical except that the standard was removed from the target window after 1 second. This was prior to the second audio beginning or any test items appearing in the test window, so the standard and test items were never visible at the same time.

Following the practice block there were two blocks of experimental trials: one block of no-label trials followed by one block of label trials, each corresponding to the standard visibility condition that had just been practiced. The no-label experimental block began with a familiarisation phase, exactly as described for the practice trial, with a rotating animation of the standard from the randomly allocated stimuli set. Following familiarisation with the standard, a test phase consisting of four experimental trials and an attention check trial began. The procedure for each test trial was the same as the practice trial, with the exception that no accuracy feedback was given. Each experimental trial presented one of the four stimuli triads for the relevant set (consisting of one shape-match, colour-match, and texture-match), with all four triads presented in a random order. The position of each item in the test window (left, middle, or right) was also randomly determined. Children could touch one of the items to indicate their choice, which was then highlighted with a green circle for 1 second. During this time, the programme would accept a change of answer and record the final choice only. Once the response to the first trial had been recorded, a trial automatically began for the next triad, until all four had been completed. In the online condition, the standard remained in place throughout the trials. In the offline condition, the standard reappeared for 1 second at the start of each new trial, but disappeared before the new test items were displayed. The final trial was always an attention check including one exact match for the standard and two distractors.

After the no-label condition had been completed the label condition began with another familiarisation phase for a different stimuli set. The procedure was as described for the practice block, with an alternative audio track naming the standard (e.g. “Look! A vink!”). Four experimental trials followed for each stimuli triad in set, as for the no-label block, with an adjustment to the audio instruction only (e.g. “Can you find another vink?”).

Once the participant had completed both the label conditions for either the online or offline condition, the procedure restarted from the known item practice block to show that the visibility of the standard would be different for the next stage of the game. The whole sequence repeated for the remaining condition, online or offline, with a new set of stimuli each for no-label and then label blocks. Figure 4 illustrates the procedure for an example trial.

If the child asked for help, the experimenter encouraged them to just give what they thought was the best answer. If a participant was still unable to choose, a secret code could be used to progress to the next trial and record the response as ‘skipped’ so it could be excluded from analysis.

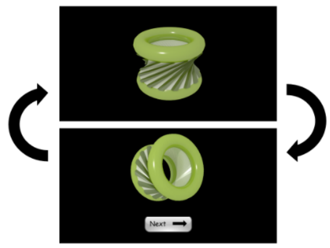
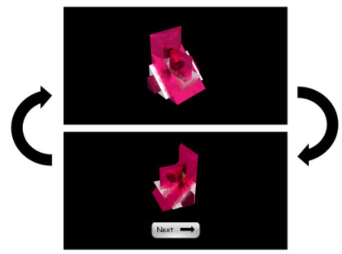
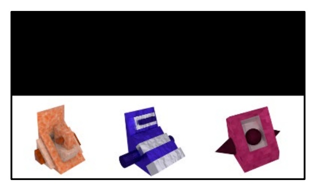
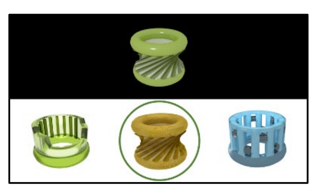
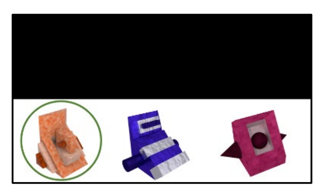
Stages and timings	Audio track (label/no label conditions)	Online block visual display	Offline block visual display
Familiarisation phase: Animation of Standard rotating for up to 30 seconds	"Look! A vink" *or* "Look at that!"		
Test phase Stage 1: 1 second display time	*No audio*		
Stage 2: Audio presentation (approx. 3 seconds)	"Can you find another vink?" *or* "Can you find another one?"		
Stage 3: wait for participant response	*No audio*		
Stage 4: Response confirmation displayed for 1 second	*No audio*		
Repeat from stage 1 for next triad in the stimuli set	*No audio*		

Figure 4: Procedure for an example trial showing audio and visual display for each stage

3.4.2 Results

3.4.2.1 Shape match acceptance against chance

To test the hypothesis that both autistic and TD children would show a shape bias, as defined by a preference for the shape-match item at above chance levels, we compared the proportion of shape-match choices children made against chance (0.33). A two-tailed, one-sample *t*-test showed that both autistic and TD groups selected shape-match test items significantly more often than expected by chance across conditions and trial types (ASD: *M* = 0.48, *t* (13) = 2.18, *p* = .048, *d* = 0.58; TD: *M* = 0.59, *t* (12) = 2.90, *p* = .013, *d* = 0.81). Hence, both populations exhibited a shape bias in this task (see Figure 5). By contrast, both groups chose the texture-match items significantly less often than chance (ASD: *M* = 0.19, *t* (13) = -3.97, *p* = .002, *d* = 1.04; TD: *M* = 0.13, *t* (12) = -4.98, *p* < .001, *d* = 1.36). However, the proportion of colour-match choices did not differ significantly from chance for either group (ASD: *M* = 0.32, *t* (13) = -0.25, *p* = .810, *d* = 0.05; TD: *M* = 0.27, *t* (12) = -0.86, *p* = .406, *d* = 0.23).

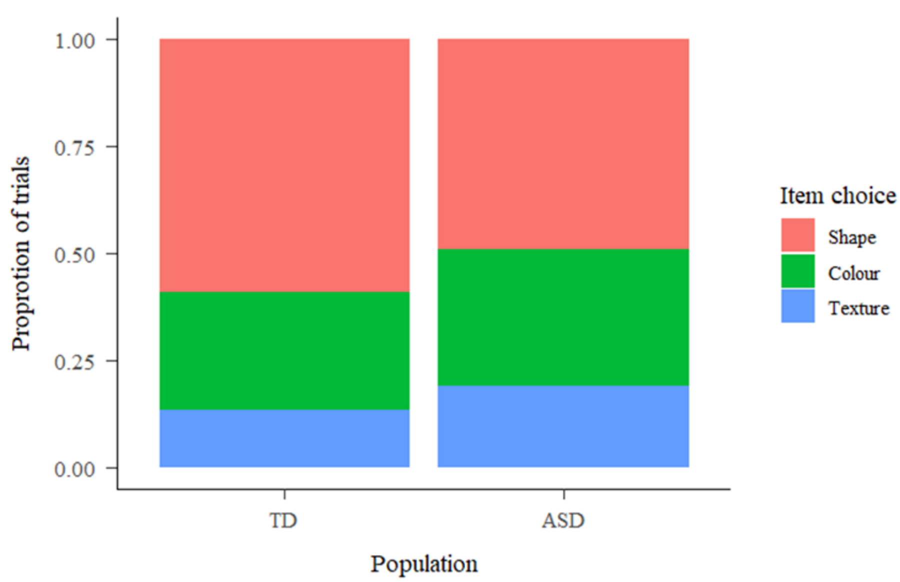


Figure 5: Proportion of forced-choice trials with shape-match, colour-match and texture-match responses for autistic and typically developing (TD) participants

3.4.2.2 Effect of population, standard visibility, and labelling

To investigate the effect of population, standard visibility, and labelling on the strength of shape bias, shape bias consistency scores were calculated for each trial. If the chosen item was the shape-match, the trial was coded as 'shape' (1), whereas selecting colour- or texture-match distractors was coded as 'other' (0) for shape bias consistency. This binary outcome was used as the dependent variable in generalised linear mixed models for binomially distributed outcomes (also referred to as mixed logit models). The independent variables were contrast coded as follows: Population (TD: -0.5, ASD: 0.5); Standard visibility (online: -0.5, offline: 0.5); Label condition (label: -0.5, no-label: 0.5). Modelling was completed in RStudio 2022.02.2, Build 485, utilising the "glmer" function of the "lme4" library (Bates et al., 2015), and post hoc comparisons were conducted with the "emmeans" package (Lenth, 2022).

We began by creating a baseline model containing only random intercepts for participants and stimuli sets and a random slope of labels by participant. The random effects structure was determined by initially fitting a maximal effects structure and reducing complexity until the model converged (Barr et al., 2013). Fixed effects of population, standard visibility, and labelling, along with two- and three-way interaction terms, were added to the baseline model as per our pre-registered analysis plan: https://osf.io/69rxf. A likelihood ratio test on the change in deviance showed that the full model was a significantly better fit to the data than the baseline model ($\chi^2(7) = 14.1$, $p = .049$). The model was simplified by removing the non-significant 3-way interaction and the Labels x Standard visibility interaction with no significant reduction in fit ($\chi^2(2) = 2.63$, $p = .269$). Mean-centred receptive vocabulary score was added as a fixed effect, but did not significantly improve fit ($\chi^2(6) = 7.59$, $p = .270$), and was therefore excluded from the final model (Table 3).

Table 3: Mixed logit model estimates for Experiment 1: forced-choice task

Effect	Group	Term	β	SE	*z*	*p*	
Fixed		Intercept	0.43	0.38	1.13	.259	
		Population	-0.98	0.76	-1.29	.198	
		Labels	0.57	0.39	1.45	.147	
		Standard visibility	-0.19	0.24	-0.80	.425	
		Population x Labels	-1.58	0.78	-2.01	.044	*
		Population x Standard Visibility	-1.03	0.48	-2.17	.030	*
			Variance	SD	Corr		
Random	Participant	(Intercept)	3.25	1.80			
		Labels	1.89	1.38	0.88		
AIC	502.9			logLik	-242.5		
BIC	539.5			Deviance	484.9		
				df residual	419		

Number of obs: 428, participants: 27, stimuli sets: 4
Significance indicators: p>.05*; p>.01**; p>.001***

Main effects of population, labels, and standard visibility were not significant predictors in the model. The odds of accepting a shape-match item over a non-shape-match item (colour or texture) was not found to differ for autistic or TD participants. Children were also no more likely to accept shape-match items that were labelled, or when the standard was visible. However, population significantly interacted with both labelling and standard visibility effects, suggesting that these predictors affected autistic and TD children's shape bias differently.

Tukey's post hoc comparisons of the z scores were conducted using the R package "emmeans" (Lenth, 2022). These indicated a significant difference in probability (P) of choosing the shape-match for TD participants between the label ($P = 0.83$) and no-label ($P = 0.56$) trials ($z = -2.22$, p

= .026). For autistic participants, there was no significant difference in shape item selection between label ($P = 0.46$) and no-label trials ($P = 0.51$; $z = 0.44$, $p = .657$). These findings suggest that TD children were more likely to generalise based on shape in the label condition, suggesting that labels may enhance the shape bias for this population. However, labelling did not facilitate shape-based generalisation for autistic children (see Figure 6).

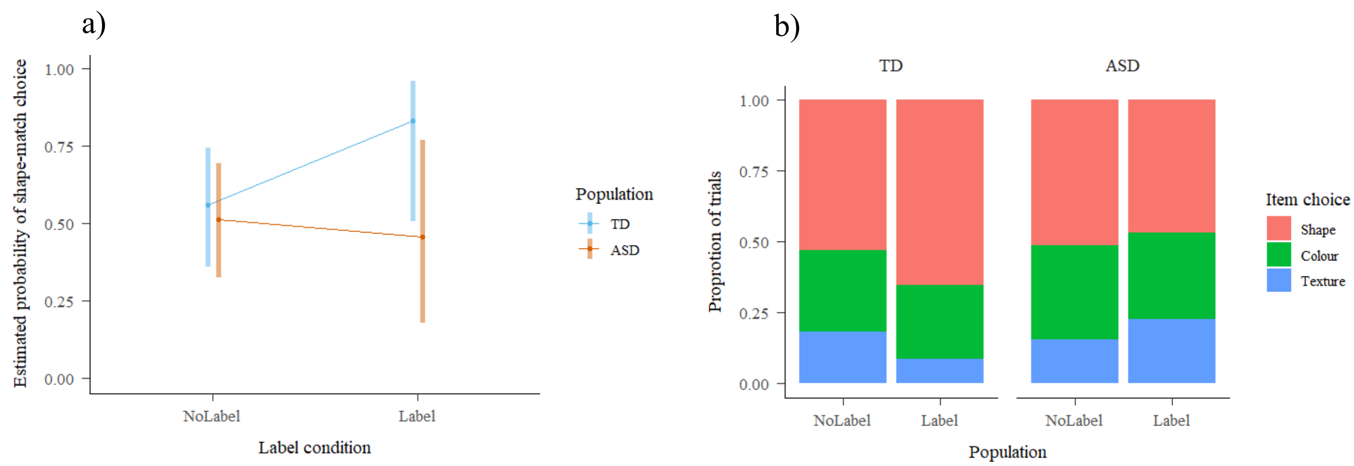


Figure 6: a) Estimated effect of labels on shape-match choices for autistic and TD participants, and b) observed proportion shape, colour and texture-match choices for label and no-label trials

For TD children, there was no difference in shape-match selection between standard visibility conditions (online: $P = 0.68$, offline: $P = 0.75$, $z = -0.92$, $p = .355$). However, for autistic children, the probability of choosing shape-match items was significantly reduced in the offline condition ($P = 0.40$) compared to the online condition ($P = 0.573$): $z = 2.24$, $p = .025$. While both groups showed similar shape preference when the standard was visible, the offline condition only resulted in a weaker shape bias for the autistic participants (see Figure 7).

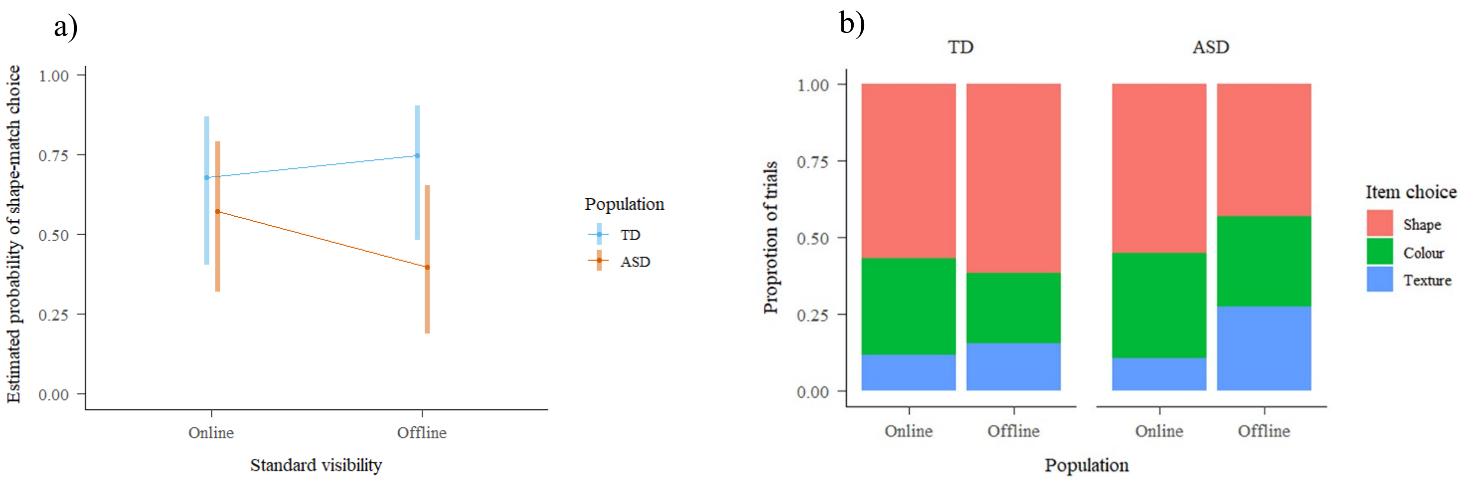


Figure 7: a) Estimated effect of standard visibility on shape-match choices for autistic and TD participants, and b) observed proportion shape, colour and texture-match choices for online and offline trials

3.4.2.3 Likelihood of choosing texture-match or colour-match alternatives

The lower odds of shape-match choices in the offline and no-label conditions raised a question over which items children were likely to choose instead in these trials. In a follow-up analysis, we created new binary outcome variables for texture-match consistency (texture-match = 1, others = 0) and colour-match consistency (colour-match = 1, others = 0). These were modelled as outcomes in two mixed logit models with fixed effects of population, standard visibility, labelling, plus two-way interaction terms with population, and a random intercept for participants.

When colour-match consistency was modelled as the outcome, none of the predictors were significant, and the model did not improve on a baseline model containing only random effects ($\chi^2(5) = 4.03$, $p = .545$). For texture-match consistency, standard visibility and the interaction term for Population x Labels were both significant predictors, yielding a significant improvement in fit compared to the baseline model ($\chi^2(5) = 20.05$, $p = .001$) (See Table 4).

Table 4: Model estimates of texture-match choices in experiment 1 (forced-choice task)

Effect	Group	Term	β	SE	*z*	*p*	
Fixed		(Intercept)	-1.97	0.24	-8.25	<.001	***
		Population	0.50	0.45	1.10	.272	
		Labels	-0.22	0.29	-0.76	.447	
		Standard visibility	0.82	0.29	2.84	.005	**
		Population x Labels	1.44	0.58	2.50	.012	*
		Population x Standard Visibility	0.89	0.58	1.55	.121	
			Variance	SD	Corr		
Random	Participant	(Intercept)	0.70	0.84			
AIC	364.9			logLik	-175.5		
BIC	393.4			Deviance	350.9		
				df residual	421		

Number of obs: 428, participants: 27, stimuli sets: 4
Significance indicators: p>.05*; p>.01**; p>.001***

The model estimated a main effect of standard visibility. Post hoc comparisons confirmed a significantly higher probability of choosing the texture-match item in offline trials (*P* = 0.17) compared to online trials (*P* = 0.08) overall. This suggests that children were more likely to choose texture-match items when the standard had to be remembered, and no difference was detected between autistic and TD children. For the effect of labels, texture-match choices were significantly more likely in the no-label condition for TD participants only (label: *P* = 0.06, no-label: *P* = 0.15, *z* = 2.12, *p* = .034). There was no significant labelling effect on texture-match selection for autistic children (label: *P* = 0.19, no-label: *P* = 0.12, *z* = -1.366, *p* = .172) (see Figure 8).

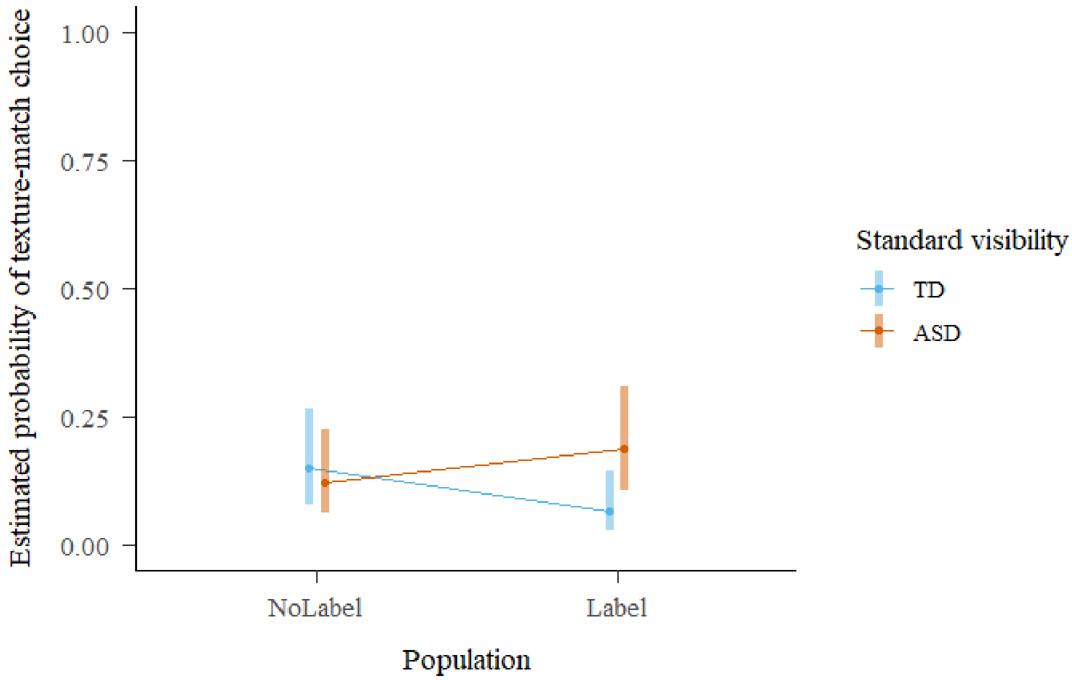


Figure 8: Effect of labels on texture-match choices for autistic and TD participants

Overall, shape-match choices were most likely when the standard remained visible, and when items were labelled. For autistic children only, shape bias was disrupted in the offline condition. Conversely, only the TD children showed a stronger shape bias for labelled stimuli. In both cases, a reduction in shape-match choices was accompanied by an increase in texture-match choices, but not colour-match choices. While colour-match choices were made, there was no evidence that the probability of doing so was related to the predictors in the model.

3.4.3 Discussion

Experiment 1 investigated whether standard visibility and labelling objects influenced TD and autistic children's preference for shape in a forced-choice NNG task when stimuli were a similarly shaped. We predicted that children would be more likely to choose shape-match items than texture- or colour-match options, and this prediction was supported for both groups. However, we found no support for our hypothesis that shape bias would be weaker for autistic participants. For the effect of labels, we expected shape bias to be stronger when items were named. This was only supported for the TD participants, who had a higher probability of choosing the shape in the label condition compared to the no-label condition, while labelling had no influence on autistic children. Additionally, in the offline condition, shape bias was weaker for autistic participants but not TD children. For both standard visibility and labels, increased preference for shape was accompanied by a decrease in texture-match choices. Interestingly, while colour-match choices were just as common as texture-match choices overall, they were unaffected by task manipulations. Therefore, it appears that any benefit from labels or standard visibility were exclusive to the shape-texture distinction for both groups.

The results of the current experiment add to a growing body of evidence that autistic children exhibit a shape bias when given a choice of between two or three potential new category members (Field et al., 2016b; Hartley et al., 2019, 2020; Keating et al., Chapter 2; Tek et al., 2008). Our findings suggest that autistic children prioritise shape over other perceptual features in an NNG task when there is competition between test items. As the forced-choice task presents a 'winner takes all' scenario (Samuelson et al., 2009), we had previously suggested that this finding shows that shape matches are the most likely winners for both autistic and TD (Keating et al., Chapter 2). However, the stimuli used in our research were purposefully designed to have highly different shapes that would be easy to distinguish from each other. This left open the possibility that group differences could emerge when shapes were more similar, as has been observed in previous research using perceptually-similar stimuli (Potrzeba et al., 2015; Tek et al., 2008). The findings of the current study suggest that autistic children still prioritise shape over texture and colour, even when stimuli share similarities in shape, as TD children also do.

While both autistic and TD children prioritised shape overall, and had similar odds of choosing the shape-match in the online condition, only the autistic participants showed lower probability of picking shape matches when the standard had to be remembered. This suggests that the offline condition presented a barrier to the autistic participants that did not adversely affect the TD participants. In typical development, there is evidence that shape bias is accompanied by a memory bias, allowing prioritised encoding of item shape over irrelevant perceptual features. Vlach et al. (2016) demonstrated that TD 2-year-olds who already exhibited a shape bias also had a better memory for object shape than colour or size. Although they could still remember other features with good accuracy immediately after presentation, shape recall was more accurate and the only feature that was still remembered after a 5-minute delay. If a prioritised memory for item shape preserves shape bias in the offline condition for TD children, the weaker shape bias observed in the autistic group may be a result of differences in feature encoding to memory.

At present, we lack data on how this shape memory bias manifests in autism, however, there is evidence that visual (Funabiki & Shiwa, 2018) and spatial (Zhang et al., 2020) working memory generally may be weaker, despite apparent advantages in visual discrimination of ‘online’ complex stimuli (Plaisted et al., 1998). A weaker working memory representation of item shape could account for the weaker shape bias observed in the offline condition of the present study. While there is evidence of intact shape-based generalisation for autistic children in other word learning tasks with a memory requirement (Hartley et al., 2019, 2020; Keating et al., Chapter 2), the stimuli in those studies were highly contrasting. Therefore, our current findings support our supposition that the offline task may only interfere with shape bias when the distractor shapes are similar to the standard. With greater similarity, more accurate encoding of shape would be required to recognise which test item was the same when comparing from memory. Conversely, with highly contrasting stimuli, one would not need such an accurate mental representation to be successful at the task as the distractors bear no resemblance to the standard.

While the offline condition predicted a reduction in shape-match choices for autistic children, this condition also predicted an increase in texture-match choices for children overall. When participants had to make the comparison from memory, though shape-match choices were still the most likely response, children had an elevated likelihood of choosing the texture-match option. This pattern is similar to that found by Tek et al. (2012) in TD infants using an offline intermodal preferential looking task. They found that TD infants preferred a named object that was the same colour and material as the standard over a competitor that was a perfect shape match, but only when the object was a good ‘overall-match’ due to having *some* similarity in shape. For items that had ‘low similarity’ in shape, the perfect shape-match was still preferred. The findings of the current study could also be interpreted as an increase in ‘overall-match’ preference, due to the high shape similarity making the texture-match a plausible category member. However, our finding that texture-match choices were less likely when the standard was visible for reference suggests that this may be explained by what children are remembering about the item, rather than a strategy to select an ‘overall-match’. This would fit with the proposal that enhanced memory for shape during word learning develops after an online attentional shape bias is established (Vlach, 2016). Hence, younger children may preferentially attend to shape in online trials, but encode more general information about the object that would guide them to an overall match when the trial is offline. One possible explanation for the differences between the populations in the current study may be that shape memory for TD children becomes more specialised with experience, whereas autistic children may continue to encode more detail about the object at the expense of encoding the overall shape, despite showing a strong shape bias in online trials. This would be consistent with several theories of attentional differences between autistic and neurotypical individuals, such as weak central coherence (Happé & Frith, 2006), a detail-focused attention bias (Plaisted et al., 1999), or enhanced perceptual function (Mottron et al., 2006), which all suggest that attention in autism defaults to the 'local' component level while TD individuals perceive the 'whole' first.

While standard visibility only affects shape-match choices for autistic participants, labels only appeared to facilitate shape bias in typical development. Generally, shape bias research with similarly-shaped items has found a label effect for TD children. Landau et al. (1988) originally demonstrated this was the case in the seminal shape bias paper, and hypothesised that labels may serve as a cue that count nouns were being referenced. This could then allow children to activate the most relevant attentional strategy. Meanwhile, research that used highly contrasting test items did not find an effect of labels (Field et al., 2016b; Keating et al., Chapter 2). Labels may only serve a facilitatory function when category membership is unclear, an effect that may be redundant in tasks involving objects that are visually distinct and clearly belong to different categories.

While experiment 1 explored the effect of labelling and standard visibility for similar stimuli in a forced choice task, we have previously demonstrated that autistic children who showed a strong shape bias when asked to choose an option from an array nevertheless appeared to have a disrupted shape bias when the same items were presented sequentially in a yes/no task (Keating et al., Chapter 2). When the objects were not in competition, autistic children may be prone to over-extending labels to non-shape-match distractors. While TD children can generalise by shape in both forced-choice and yes/no variants of NNG tasks (Booth et al., 2005; Keating, Chapter 2; Landau et al., 1988), that does not mean that both tasks necessarily tap into the same cognitive mechanisms. It has been argued that a slight shape preference can be enhanced in the forced-choice task as only the ‘best’ example can be chosen, whereas the yes/no task allows children to weigh additional characteristics as important (Samuelson et al., 2009). Additionally, we have argued that performance in forced-choice tasks may be as expected for autistic children because there is no need to make a decision about the distractor items (Keating et al., Chapter 2). The yes/no task, however, requires children to use shape difference as a criterion for category exclusion. In experiment 2 we used a yes/no task to investigate the impact of labels and standard visibility on autistic and neurotypical children’s shape bias when stimuli are perceptually similar.

3.5 Experiment 2: Online and Offline shape bias in autism in a yes/no task

3.5.1 Method

3.5.1.1 Participants

Participants were 16 children with ASD aged between 4 years 7 months and 9 years 7 months (*M* age = 7.18 years; *SD* = 1.46 years), and 14 typically developing children aged between 2 years 9 months and 4 years 8 months (*M* age = 3.69 years; *SD* = 0.51 years). Children were recruited from specialist schools, mainstream primary schools and preschool settings in the North West of England as part of a series of studies and had parental consent to take part. The groups were matched on receptive vocabulary as measured by the British Picture Vocabulary Scale (2nd Edition) (Dunn et al., 1997). The BPVS scores of the final sample for the children with ASD (*M* = 4.25 years; SD = 1.37 years) and TD children (M = 4.14 years; SD = 0.86 years) did not significantly differ (t (25.59) = -0.24, p = .814, two-tailed). Autistic participants had been previously diagnosed by a qualified educational or clinical psychologist, using standardised assessment and professional judgement. As for experiment 1, the Expressive Vocabulary Test (2nd edition) (Williams, 2007) was used to measure expressive vocabulary, and the cognitive battery subscale of Leiter-3 (Roid et al., 2013) was used as a measure of non-verbal intelligence. Available scores are reported in Table 1.

To ensure the current samples had equivalent receptive vocabulary abilities, data for 13 autistic children whose scores exceeded 6.5-years (*n* = 7) or were unable to complete this measure (*n* = 6) were not included in the analysis. Data for 1 TD participant who did not complete all of the conditions was also excluded. Responses for these children remain available in the full dataset.

3.5.1.2 Design

This experiment used a mixed (2x2x2) design, as described for experiment 1. Population was a within-participants variable with two levels (TD/ASD). There were two between participants variables: Standard visibility (online/offline) and Labels (label/no-label). The order of the visibility conditions was counterbalanced so half of the participants experienced the online condition first, and this order was kept consistent across experiments. Therefore, any child who completed the online condition first in experiment 1 would also complete the online condition first in this experiment. Standard visibility and label condition manipulations were as described for experiment 1. Each participant was assigned a random object/label pairing at the beginning of the study which was consistent for both experiment 1 and experiment 2. So, if a participant had already seen set A referred to as a ‘vink’ in experiment 1, set A was still a ‘vink’ in experiment 2. The order of the tasks was counterbalanced across participants, and a minimum gap of 1 day was required between experiment 1 and 2. In practice, most participants had at least 7 days between testing sessions.

3.5.1.3 Materials

Visual stimuli. This experiment used the same materials as experiment 1. Only the first and second triads of stimuli from each set were used due to the individual presentation of each item extending the time children needed to pay attention to the study (see Figure 1). These stimuli were selected to ensure that the shape-match item was seen with both available textures and colours. Only the alternative combinations were omitted. Each image set consisted of one standard, six test items (two colour-match; two texture-match; two shape-match), and one control item that did not match the standard.

Au*dio stimuli.* Experiment 2 used the same audio recordings as experiment 1, with the exception of the test phase question. To be congruent with a yes or no response, the question asked "Is this another…" instead of "Can you find another…". This track was recorded with Audacity® (Version 2.1.3; Audacity Team, 2017) and used the same speaker as the other recordings. The wording for each track is detailed in Table 5.

Table 5: Scripts for pre-recorded audio stimuli for yes/no task

Wording	Phase/condition used in
"Look at that!"	Familiarisation phase: no-label
"Look! A vink!"	Familiarisation phase: label
"Look! A zain!"	Familiarisation phase: label
"Is this another one?"	Test phase: no-label
"Is this a vink?"	Test phase: label
"Is this a zain"	Test phase: label
"Perfect"	Practice phase only
"Uh oh! Try again."	Practice phase only

Presentation software and hardware. This experiment ran on the same Microsoft Surface Pro, using the same bespoke PsychoPy2 programme, as experiment 1. The yes/no task was an identical programme to experiment 2 as described in (Keating et al., Chapter 2). Only different stimuli images and audio files were used. An example of the screen layout during the response phase of this task is illustrated in Figure 9. Only touch-screen responses to the 'tick' or 'cross' buttons were recorded as responses. The programme ignored input to any other location on the screen. The python code required to replicate the game is available on the OSF project: https://osf.io/sgcex

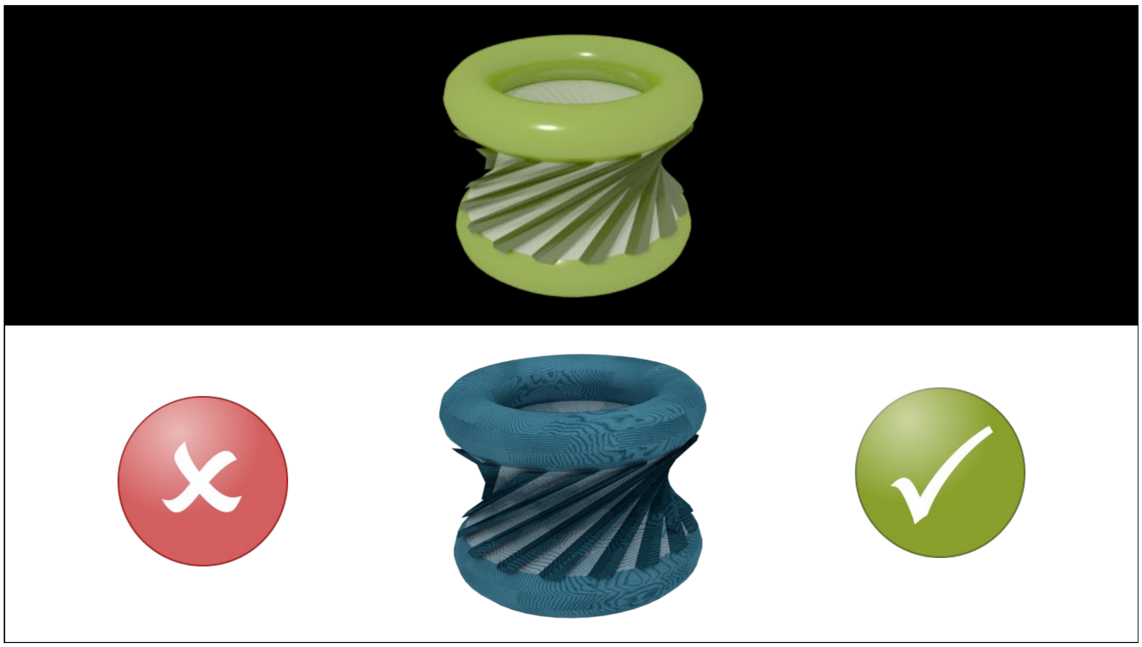


Figure 9: Visual display for the yes/no task while waiting for a response

3.5.1.4 Procedure

Warm-up task. The warm-up task to practice the 'tick' and 'cross' button responses was as described for experiment 1. This task was repeated at each visit to ensure children remembered how to use the buttons to indicate 'yes' or 'no' answers from one session to the next.

3.5.1.4.1 Yes/No NNG task

Practice block. For the yes/no task, the practice block was designed to be similar to that described for experiment 1, with the following adjustments. The familiarisation phase was identical to experiment 1. At the test phase, the green bed image displayed in the test window. In the online condition, the bed in the target window also remained visible, else it was removed in the offline condition. At the same time, the "Is this another one?" audio track played. Once the question finished playing, the tick and cross response buttons appeared on the right and left of the bed. Tapping the tick triggered the 'Perfect' audio, and tapping the cross trigged 'uh oh, try again'. Regardless of accuracy, a green circle appeared around the chosen button. Touching any other place on the screen, including objects, had no effect. Following a correct response on the 'yes' trial, the test phase repeated with the blue butterfly as the test item to practice a 'no' response. After a correct 'no' response, the game proceeded to the experimental trials.

Experimental block. The no-label condition was presented first with the same timings and audio recordings as the practice block, apart from the audio feedback regarding accuracy. A novel stimuli set was used in place of the known items. Following the familiarisation phase, one item was randomly selected for the trial and displayed in the bottom window. Possible images were the six test items for the set, or a copy of the standard. Two copies of the standard were included to give an equal number of trials that could have a 'yes' response regardless of which feature children thought was important. Once a response was recorded, another item from the set was displayed, and the audio track repeated, until a response had been given to all eight items. A final trial displayed an image with no features in common with the standard as an attention check. In the online condition, the standard remained visible in the target window throughout all nine trials. In the offline condition, the standard was displayed for 1 second at the beginning of each trial and removed prior to test item display and the audio track playing.

Once all nine no-label trials had been completed, the familiarisation phase repeated with a new standard for use in the label condition. Label and no-label trials differed only in the wording of the audio track ("Is this another vink?"/ "Is this another one?"). On completion of nine trials in the label condition, a new practice trial for either the online or offline condition, dependent on which remained to be completed.

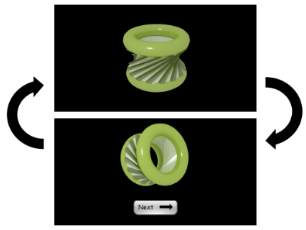
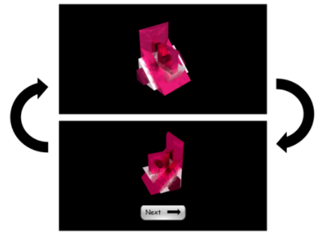
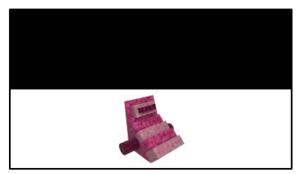
Stages and timings	**Audio track (label/no label conditions)**	**Online block visual display**	**Offline block visual display**
Familarisation phase: Animation of Standard rotating for up to 30 seconds	"Look! A vink!" *or* "Look at that!"		
Test phase Stage 1: 1 second display time	*No audio*		
Stage 2: Audio presentation (approx. 3 seconds)	"Is this another vink?" *or* "Is this another one?"		
Stage 3: wait for participant response	*No audio*		
Stage 4: Response confirmation displayed for 1 second	*No audio*		
Repeat from stage 1 for next triad in the stimuli set	*No audio*		

Figure 10: Game audio and visual display for a trial in the online and offline conditions

3.5.2 Results

3.5.2.1 Data analysis

To test the hypothesis that standard visibility and labels affect shape bias for autistic and TD children in a yes/no NNG task, we created a mixed logit model in RStudio 2022.02.2, Build 485, utilising the “glmer” function of the “lme4” library (Bates et al., 2015). Raw responses to each trial were coded ‘1’ for a tick/yes response and ‘0’ for a cross/no. An item variable ‘Standard Congruent Characteristic’ (SCC) was created to identify which feature the test item had in common with the standard on each trial (colour/texture/shape). To determine if raw responses were consistent with shape bias, SCC was included in the baseline mixed logit model (reference category: shape), along with random intercepts for participant and item set and random slopes for standard visibility. The random effects structure was determined by initially fitting a maximal effects structure and reducing complexity until the model converged (Barr et al., 2013).

To estimate the effect of the independent variables on item acceptance for the different levels of SCC, fixed effects for population (contrast coded: ‘TD’ = -0.5, ‘ASD’ = 0.5), standard visibility (contrast coded: ‘online’ = -0.5, ‘offline’ = 0.5), labelling (contrast coded: ‘label’ = -0.5, ‘no-label’ = 0.5), plus two- and three-way interaction terms, were added to the baseline model and compared. A likelihood ratio test indicated the full model was a significant improvement on the baseline model ($\chi^2(18) = 39.79$, $p = .002$), however, the labelling fixed effect did not reach significance in any interaction or as a main effect. Labelling effects were therefore removed from the model without a significant reduction in goodness of fit ($\chi^2(9) = 9.43$, $p = .399$). Mean-centred receptive vocabulary, and interaction terms with the other predictors, were added t but did not significantly improve it ($\chi^2(12) = 10.41$, $p = .580$). Table 6 shows the model structure and output. Figure 11 shows the observed proportion of trials that participants gave 'accept' or 'reject' decisions on.

Table 6: Mixed logit model estimating probability of accepting a test item in a 'yes/no' task by standard congruent characteristic (shape, texture or colour), population and standard visibility

Effect	Group	Term	β	SE	*z*	*p*	
Fixed		Intercept	0.74	0.38	1.95	.051	
		SCC (Colour)	-1.35	0.25	-5.44	< .001	***
		SCC (Texture)	-1.36	0.26	-5.34	< .001	***
		Population	0.47	0.75	0.63	.528	
		Standard visibility	-0.85	0.41	-2.08	.037	*
		SCC (Colour) x Population	1.40	0.49	2.84	.004	**
		SCC (Texture) x Population	1.69	0.50	3.34	.001	***
		SCC (Colour) x Standard visibility	1.13	0.50	2.28	.023	*
		SCC (Texture) x Standard visibility	1.74	0.51	3.42	.001	***
		Population x Standard visibility	1.01	0.81	1.25	.212	
		SCC (Colour) x Population x Standard visibility	-1.76	0.98	-1.80	.071	
		SCC (Texture) x Population x Standard visibility	-2.05	1.00	-2.04	.042	*
			Variance	SD	Corr		
Random	Participant	Intercept	3.23	1.80			
		Standard visibility	1.33	1.15	-0.46		
	Stimuli Set	Intercept	0.01	0.11			
AIC	789.1				logLik	-378.6	
BIC	862.3				Deviance	757.1	
					df residual	698.0	

Number of obs: 714, participants: 30, stimuli sets: 4
Significance indicators: p>.05*; p>.01**; p>.001***

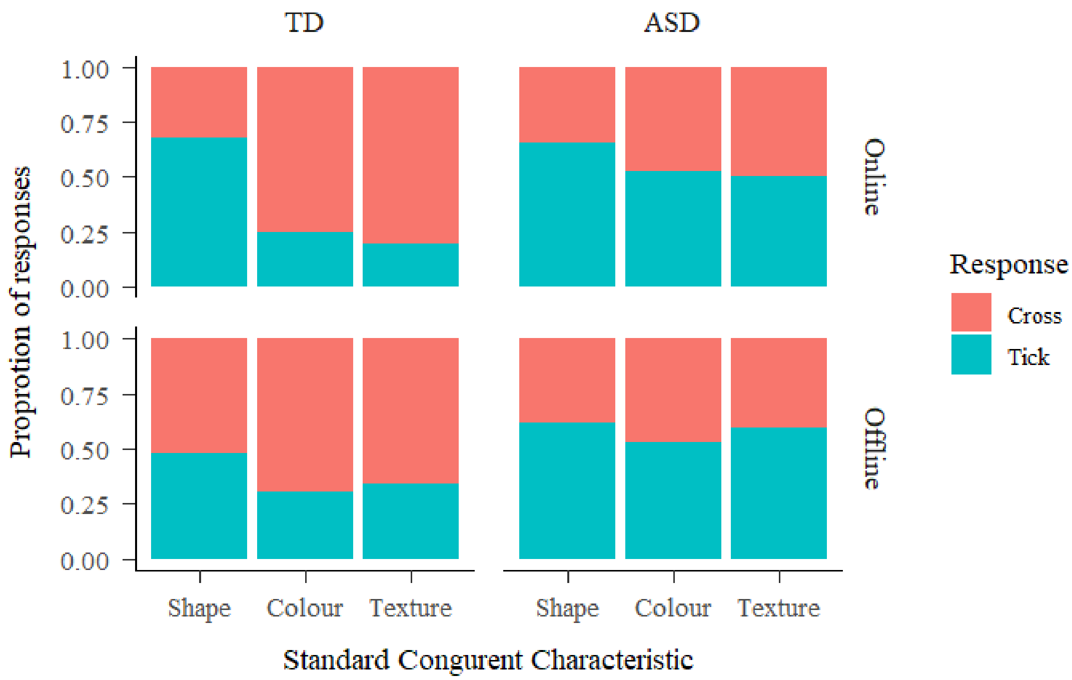


Figure 11: Observed proportional data for 'accept' (tick) and 'reject' (cross) responses to items that were a shape-match, colour-match or texture-match to the standard, by population, in the online and offline condition

3.5.2.2 Responses consistent with shape bias by population

Standard congruent characteristic was a significant predictor of item acceptance. Shape-match SCC had an estimated 0.68 probability of a 'tick' response, whereas the odds were significantly reduced for both texture and colour levels of SCC. This suggests that children's responses were consistent with shape bias overall, however, the two-way interaction between SCC and population was also significant. Post hoc comparisons of SCC by population group were conducted to confirm whether the model predicted shape bias for both groups and are summarised in Table 7. TD participants were significantly more likely to accept shape-match items in comparison to colour- and texture-match items and showed no difference between the non-shape-match distractors. For autistic participants, however, there was no significant differences between the different item characteristics. This showed a pattern consistent with shape bias for the TD participants, but no clear preference for shape for the autistic participants.

Table 7: Post hoc comparisons of probability of accepting items by each level of SCC for TD and autistic participants

	TD		Autism	
Contrast	*z*	*p*	*z*	*p*
Shape/Colour	5.28	<.001	2.13	.085
Shape/Texture	5.48	<.001	1.68	.211
Colour/Texture	0.43	.904	-0.43	.910

As illustrated in Figure 12, population differences in the probability of accepting an item were driven by lower odds of TD participants accepting the colour- and texture-match items . There was no significant difference in acceptance probability for shape-match items for autistic participants (P = 0.72) compared to TD participants (P = 0.63) (z = -0.631, p = .528). In contrast, autistic children were significantly more likely to also accept both colour-match (ASD: P = 0.58, TD: P = 0.18, z = -2.48, p = .013) and texture-match (ASD: P = 0.61, TD: P = 0.15, z = -2.835, p = .005) items. This confirms that there were significant differences between populations for accepting non-shape-match items only. TD children had strong odds of rejecting these items, while autistic children were likely to accept any item regardless of the characteristic it had in common with the standard.

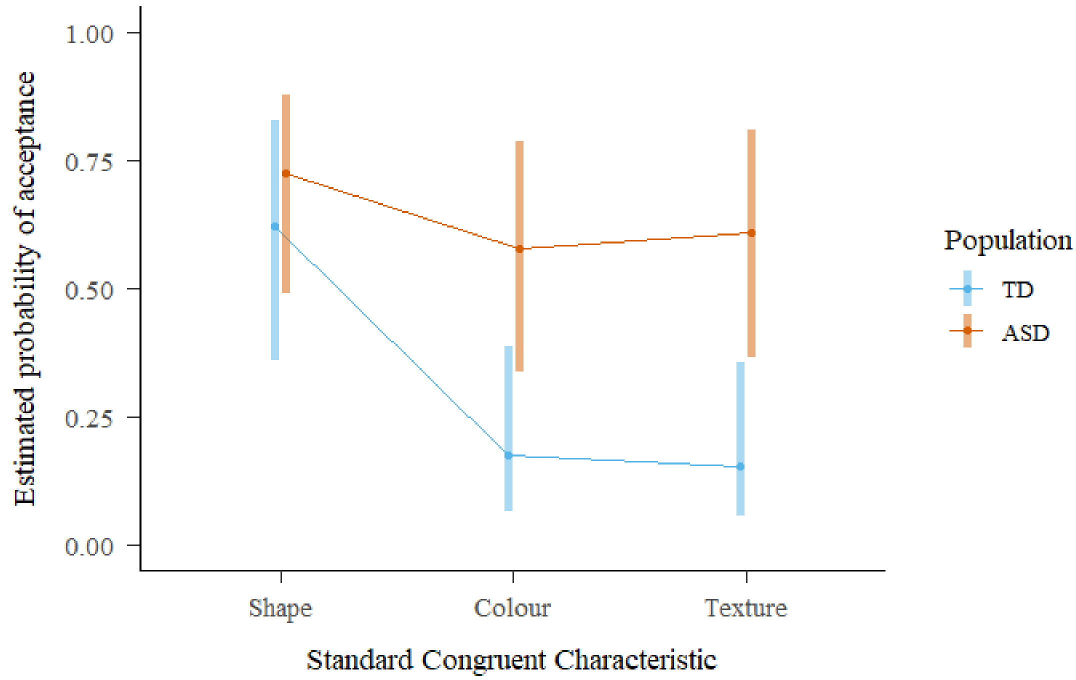


Figure 12: Probability of accepting shape-match, colour-match, and texture-match items by population

3.5.2.3 Effect of standard visibility on item acceptance by population

The relationship between standard visibility and item acceptance significantly differed when the items were colour- or texture-matches. Post hoc pairwise comparisons confirmed that the probability of acceptance was significantly higher for shape-match items in the online condition (P = 0.76) than the offline condition (P = 0.58): z = 2.08, p = .037. Conversely, texture-match items were significantly less likely to be accepted (online: P = 0.25, offline: P = 0.45, z = -2.14, p = .033). However, there was no significant difference for colour-match items (online: P = 0.32, offline: P = 0.38, z = -0.69, p = .492). Overall, this pattern is consistent with increased odds of making inclusion and exclusion decisions that are compatible with shape bias when the standard is visible, compared to when the comparison must be done from memory.

The three-way interaction between SCC, population, and standard visibility was also significant. Figure 13 visualises this interaction. For post hoc comparisons by group, the effect of standard visibility on item acceptance was only significant for the TD group for both shape-match (online: $P = 0.76$, offline: $P = 0.46$, $z = 2.25$, $p = .024$) and texture-match items (online: $P = 0.08$, offline: $P = 0.27$, $z = -2.15$, $p = .032$). For the autistic group, there were no significant differences in acceptance probability between the online and offline conditions for any level of SCC. Being able to see the standard for reference resulted in higher likelihoods of shape bias consistent responses to shape- and texture-match items for the TD group only. There was no such effect for the autistic participants, whose item acceptance probability remained high regardless of standard visibility or which characteristic the test item had in common with the standard.

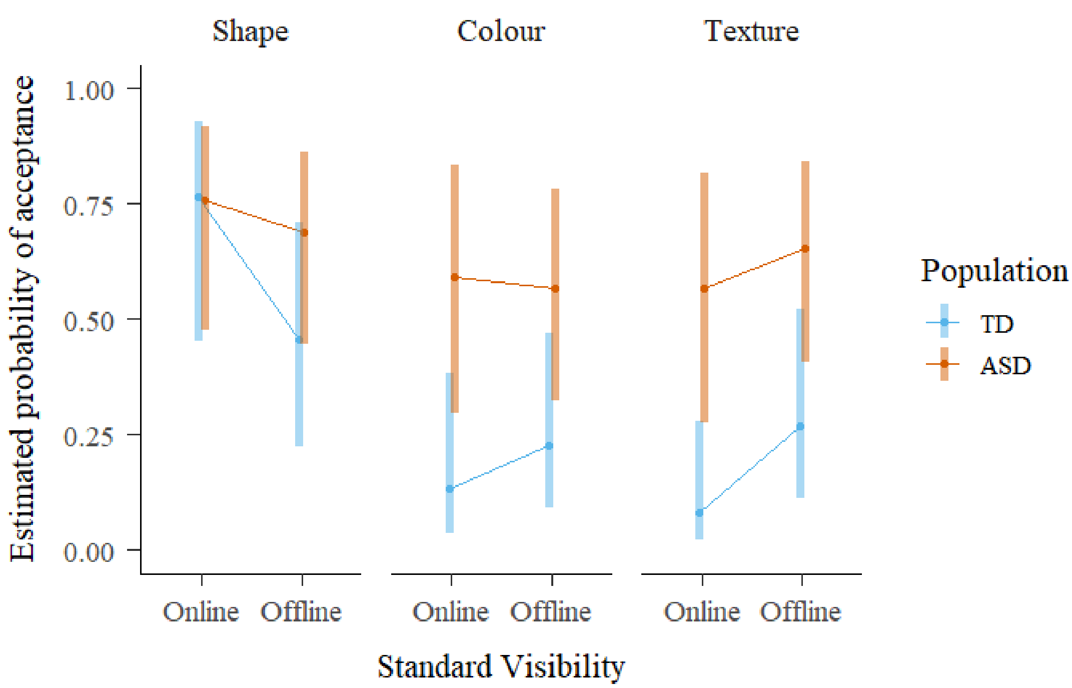


Figure 13: Probability of accepting shape-match, colour-match, and texture-match items in the online and offline condition by population

3.5.3 Discussion

Experiment 2 investigated the effects of population, labelling, and standard visibility on shape bias in autism using a yes/no task. Our results indicated that the standard congruent characteristic (shape, colour, or texture) predicted shape bias consistent responses for TD children, but not for autistic children. While shape-match acceptance was similar for both groups, only the TD group were more likely to reject items that were a different shape to the standard. Furthermore, for TD children, shape bias was stronger when the standard was visible in the online condition, but reduced in the offline condition for items that were a shape-match or texture-match to the standard. There was no such effect of standard visibility on choices made by the autistic participants. Labelling and receptive vocabulary were not significant predictors of children's responses in either group.

We hypothesised that test item SCC would predict children's responses overall, with shape-match items being accepted more often than texture- or colour-match items. However, our findings only supported this prediction for TD participants. Viewed in isolation, the results from experiment 2 indicate that shape did not influence autistic children's decisions on whether to accept an object as a category member. This finding is consistent with label over-extension reported in previous research with autistic participants. For instance, in Hartley and Allen (2014), autistic children were more likely to include both shape- and colour-match items as examples of a novel category. This finding was also replicated by Tovar et al. (2020), who demonstrated that this over-generalisation extended to novel items that had no features in common with the standard. In our previous research using the same methodology as this experiment with highly contrasting stimuli (Keating et al., Chapter 2), we also found over-generalisation of distractors in an almost identical sample of autistic children, however, shape-match items still had the highest likelihood of acceptance. In the current study, low contrast stimuli increased the difficulty of the task, as the distractors were similar enough in shape that they could feasibly belong to the same broad category as the standard. Under these circumstances, the likelihood of accepting texture- and colour-match items was so high that autistic children were just as likely to accept the distractors as the shape-match items. On the other hand, TD participants were still more likely to exclude these candidates, even when the shapes were similar. In Keating et al. (Chapter 2) we proposed that successful shape bias in the yes/no task for TD participants may be the result of using shape differences as an exclusion criteria, and that it is these exclusion decisions that may drive the disruption in autism. Autistic children may be less willing to rule out potential category members on the basis of shape differences alone, perhaps because of uncertainty over the scope of the category. The current pattern of results is consistent with this explanation.

Our hypothesis that standard visibility would affect the likelihood of shape bias consistent responses was supported by our findings, as was our prediction that there would be population differences between the online and offline conditions. In the online condition, TD participants were highly likely to accept shape-match items and highly likely to reject the distractors. Conversely, when the standard had to be held in memory for a period of approximately 3 seconds in the offline condition, shape bias was considerably weaker for the TD group. Here, shape-match acceptance was significantly less likely, whereas texture-match items were more likely to be accepted. For autistic participants, however, all of the test items had a similar chance of being accepted regardless of standard visibility. For shape-match items, standard visibility in the current study had the same effect as in Keating et al. (in prep). That is, for both high- and low-contrast stimuli, overall shape-match item acceptance is higher in the online condition (current study: $P = 0.76$, high contrast study: $P = 0.88$) than the offline (current study: $P = 0.58$, high contrast study: $P = 0.70$). These studies were also consistent in finding no effect of standard visibility for colour-match items (current study - online: $P = 0.32$, offline: P = 0.38; high contrast study - online: $P = 0.30$, offline: $P = 0.37$). However, for texture-match items, the findings were quite different. In Keating et al. (Chapter 2), texture-match items had a similar chance of acceptance in both conditions (online: $P = 0.36$, offline: $P = 0.37$), whereas the current research found significantly higher probability of acceptance in the offline condition (online: $P = 0.26$, offline: $P = 0.46$).

These findings are consistent with our claim that shape difference may be used to inform category exclusion decisions in addition to category inclusion decisions. In the current study, items were similar enough that an accurate memory of shape would be required to accurately reject the distractors. In contrast, items in previous research were so different that even a weak memory of the standard's shape should be sufficient to accurately inform category exclusion decisions. Of the TD group, the need to remember the more similar items in this experiment resulted in a decrease in shape-match acceptance, and an increase in texture-match acceptance. For the autistic participants, however, high levels of acceptance across items did not appear to be reduced by the ability to see the standard. While this finding differs from Keating et al (Chapter 2), it suggests that autistic children were less willing to rule out the similar distractors in this study as they could belong to the same superordinate category.

3.6 General discussion

Experiment 1 found that autistic and TD children both exhibited a shape bias in a force-choice task NNG. However, when asked to give a 'yes' or 'no' response to the same stimuli in Experiment 2, autistic participants were just as likely to accept items that were a texture-match or colour-match to the standard as they were to accept shape-match items. These results demonstrate that autistic children may, or may not, demonstrate a shape bias across different tasks. Furthermore, these results contribute the novel finding that standard visibility had different effects on shape bias in TD and autistic children. In the forced-choice task, the offline condition only disrupted shape bias for the autistic children. However, in the yes/no task, TD children exhibited a weaker shape bias when the standard had to be remembered. Furthermore, we found a stronger shape bias was associated with labelling the standard, but only for TD children in the forced-choice task.

A key finding from this research was the different ways that shape bias emerged in yes/no and forced-choice tasks for autistic and TD participants. The finding that shape bias was disrupted for autistic participants in the yes/no task despite their strong shape bias in the forced choice task replicates our previous study (Keating et al., Chapter 2), providing further evidence that choice of experimental task matters more for autistic participants when studying lexical generalisation. While it has been acknowledged that yes/no and forced-choice NNG tasks ask slightly different questions, and so may involve different processes (Samuelson et al., 2009), TD children tend to consistently generalise labels to objects that are the same shape as the standard regardless of methodology (Landau et al., 1988).

In the yes/no task, each item must be considered individually and there is no need to take the other items into account. The question in this case is 'is this a good example of…'. By contrast, the forced-choice asks children to identify 'the best example of…' and enables decisions based on direct visual comparison. We can think of this as a competitive task where the best example wins. Samuelson (2009) suggested that the importance of shape may be 'amplified' in a forced choice task, as any characteristic with lesser importance would be inhibited in order for shape to be the mostly likely ‘winner’. In the yes/no task, however, characteristics like texture could be important enough to be accepted in addition to shape. The current finding that autistic children only showed a reliable shape bias in the forced-choice task is consistent with the claim that the most important feature is amplified in this task. If the importance of shape to category membership is heavily weighted in typical development, a shape bias is likely in both tasks. However, if it is only weighted as slightly more important than the other features, the forced choice task would be the easier of the two as only the highest weighted feature needs to ‘win’.

Hartley and Allen (2014) previously suggested that additional features of objects (e.g. colour) may be equally weighted in autism, leading to over-generalisation. Since then, additional findings have suggested that shape does still have some advantage over other features (Keating et al., Chapter 2). Perplexingly, over-generalisation also occurs when there are no features in common with the standard at all (Tovar et al., 2020). This suggests both that shape is judged as the most important feature, but that autistic children are, nevertheless, more likely to accept distractor items without taking shape into account when they have a free choice to do so. These findings indicate it may only be category exclusion decisions that are impacted by autism, and that TD children may benefit from increased inhibition of candidates that are a different shape. Autistic children can successfully generalise by shape, but are more reluctant to rule an object out, even when it bears no similarity to the standard. This may be because there is not enough information about the category to rule out distractors through a process of logical deduction, as autistic children are able to do in mutual exclusivity fast mapping tasks (de Marchena et al., 2011). In other category formation tasks, autistic children have been found to generalise a novel category to new examples successfully when they have the opportunity to learn the rule that governs category membership (Klinger & Dawson, 2001). One possibility is that autistic children would make more selective exclusion decisions if given more information about the scope of the category, such as providing them with multiple examples, allowing them to infer the organising features. Investigating manipulations to training that facilitate autistic children's accurate exclusions decisions could be a promising line of future investigation.

We predicted that shape bias would be stronger for labelled items in both experiments, however, this prediction was only supported for TD children in the forced-choice task. In the yes/no task, labels were not a significant predictor for either population irrespective of standard visibility. While our previous study with high contrast stimuli did not identify a labelling effect (Keating et al., Chapter 2), this may have been due to the ease of differentiating high contrast stimuli. With low contrast stimuli, labels supported a shape bias, as has been the case with other research using perceptually similar items (e.g. Landau et al., 1988). Landau et al. (1988) originally proposed that labels may provide a cue that basic level categories are being referenced, whereas unlabelled items may be considered part of a broader, superordinate category. If this were the case, we might have expected to see a higher level of distractor acceptance in the yes/no task in the no-label condition, as this would indicate broader, less exclusive categories being formed when the children were free to do so. However, there was no evidence that children formed larger categories in no label trials.

An alternative explanation to levels of categorisation is that labels serve a function in guiding attention. In typical development, it has been demonstrated that labels can draw attention to commonalities between objects (Althaus & Plunkett, 2015). Furthermore, children can use past experience to direct attention to the most predictive features in non-linguistic learning tasks (Beesley et al., 2015). In this way, the label effect may support a shape bias by directing attention to the most predictive perceptual feature, which is frequently shape for inanimate count nouns. The finding that children accepted a similar number of test items in both label and no-label trials – so did not make broader, more inclusive generalisations when there were no labels - supports the directed attention explanation over the claim of cues to levels of categorisation. However, the labelling effect may only be detectable when the generalisation task is demanding enough for that extra support to be beneficial. With a sufficiently simple task, such as when discriminating between perceptually distinctive stimuli, high success rates may not be further enhanced by the addition of a label. We propose that the yes/no task may tap into category exclusion processes that are not necessary in competitive forced choice scenarios (Keating et al., Chapter 2). With a separate mechanism governing category exclusion, distinct from category inclusion processes, we should not presume that labels serve the same function for both processes. Labels may support attention to any previously useful perceptual features in competitive tasks, providing a 'boost' for that feature, without having the same impact on decisions to rule out distractors.

The current experiment, then, contributes the novel finding that labels may primarily support categorisation in typical development in tasks that involve selecting between competitors, such as the forced-choice task. In a realistic word learning situation, there may be many potential referents for a novel word that are available to a child (Quine, 1960), so having the support of a label to identify the most likely candidate would be a valuable asset. However, the current findings suggest no such label advantage for autistic children, so the mechanism for identifying the referent may differ.

In cases where shape bias was disrupted by one of our experimental manipulations, that disruption presented as an increased preference for texture-matches in both autistic and TD children. While shape is the main organising characteristic for a large proportion of count nouns (Samuelson & Smith, 1999), texture is also predictive of membership for other categories children are familiar with, including animals. TD children demonstrate a shape and texture bias when given cues that objects are animates rather than artifacts, either with perceptual cues like eyes or shoes (Jones et al., 1991), or with additional contextual information (Booth et al., 2005). Furthermore, they generalise by texture when items are deformable (e.g. made from sponge or clay) (Samuelson & Smith, 2000a). While there are examples where colour can be a useful cue, this is usually in addition to another feature. For instance, if categorising mass nouns, colour could help identify custard from other puddings with a similar texture, but colour alone is arguably less useful, and is only classed as an organizing factor for around 4% of nouns in children's early vocabulary structures (Samuelson & Smith, 1999). For this reason, ignoring colour in most cases would be an efficient strategy when generalising labels.

While existing evidence suggests that TD children can tune their attention to include texture when it is relevant and predictive of category membership, that does not mean that children are automatically able to inhibit their attention to texture when it is an irrelevant cue. In the current research we found a slight increase in preference for overall matches in the offline condition for autistic children in the forced-choice task and TD children in the yes/no task. This increase in texture-match acceptance is unlikely to be due to a change in belief about what the organising feature of the category is, but rather what information children have encoded about the standard. When the items were in competition in the forced-choice task, shape-match items still 'won' for TD participants, suggesting that shape was encoded well enough, or texture was inhibited enough, that it did not influence their responses. However, when there was no competition during the yes/no task, shape-match acceptance was reduced. This should not be the case if shape encoding was highly accurate. The finding that autistic children were just as likely as the TD children to ignore texture when they could see the standard suggests that their difference may lie in what perceptual information is automatically processed. While TD children may selectively remember shape over other perceptual features (e.g. Vlach, 2016), automatic attention to alternative features in autism could disrupt this. For instance, there is evidence that autistic individuals may process small, local details before attending to the overall, global properties of items due to weak central coherence (Happé & Frith, 2006), or that they may process additional details due to enhanced perceptual capacity (Mottron et al., 2006). These differences in automatic attention could influence autistic children when generalising based on a mental representation in offline tasks. Additionally, TD children were more likely to reject shape match items in the offline condition, perhaps suggesting that they are less willing to generalise when category membership is uncertain. Conversely, autistic children may be less willing to rule potential category members out without more information.

Our data demonstrate that autistic children acquire and use shape bias to make novel noun generalisation decisions and, under the right conditions, do so indistinguishably from TD children. However, our finding that the offline condition disproportionately affected shape bias for autistic children has implications for how this word learning strategy is investigated with this population. Many studies employ the intermodal preferential looking paradigm (Abdelaziz et al., 2018; Potrzeba et al., 2015; Tek et al., 2008), which requires offline generalisation. The current research shows that no single method of measuring shape bias comprehensively represents the shape bias in autism – combining both forced-choice and yes/no tasks is required to capture both their strengths and weaknesses. Thus, future research in the field would benefit from questioning assumptions that different methods are equally valid when used with neurodiverse populations.

Our findings may also have applications in supporting language development for autistic children. In typical development, shape bias training has been associated with increased rates of language learning (Smith et al., 2002). If autistic children's ability to use the shape bias is improved by presenting visible examples and a choice of competitors, integrating these conditions into educational interventions targeting vocabulary development may be beneficial. Future research could explore the impact of shape bias training on language development for autistic children.

This research was conducted as part of the same project as Keating et al. (Chapter 2), and as such suffered from the same recruitment challenges caused by the covid-19 pandemic. These challenges resulted in a smaller-than-planned for sample size. Another possible limitation is the use of the same stimuli sets across both the forced-choice and yes/no tasks. As many participants took part in both experiments, they would have encountered the same stimuli twice. We took several precautions to limit the risk of confounding the results. Label-items pairings were consistent across the tasks, so children would never hear a set they had seen before associated with a different label. One possible risk was that performance in the offline condition could be impacted by memory of a previous encounter with the stimuli. The same stimuli were assigned to the offline condition in both tasks to limit overall exposure, and a minimum time-period of one-day was left between the two tasks. Furthermore, the order of tasks was counter-balanced, so any advantage from seeing the stimuli previously would have evenly affected the tasks. It was felt that the benefits of observing how the same children performed with the same stimuli on different tasks outweighed the risk, and this has allowed us to build a compelling argument that task demands account for differences in children's responses. However, future research replicating the pattern with different stimuli would be ideal confirmation.

3.6.1 Conclusions

Autistic children show a shape bias when generalising novel object names, however, they do not do so under all of the same conditions as TD children. Autistic children benefit from being able to generalise online from a visible exemplar, and from having a selection of candidates to choose from. However, while labelling items supported attention to shape for the TD participants, labels had no effect on autistic children's responses. Furthermore, in a yes/no task with low contrast stimuli, there was no measurable shape bias for autistic participants. Our research shows that generalisation decisions that are consistent with a shape bias are influenced by multiple task demands, and that these demands to not impact TD and autistic children equally. These findings support our theory that object shape plays an important role in both category inclusion *and* exclusion decisions, but these processes may be subserved by different mechanisms. There may be multiple factors that influence how autistic children respond on a NNG task, however we suggest these differences are not due to a general shape bias 'deficit', but may arise from uncertainty about the status of distractor objects.

Chapter 4: Attention to global shape, not local features, informs offline noun generalisation in autism and typical development

4.1 Introduction to Chapter 4

The previous two chapters have advanced our understanding of the circumstances under which autistic children may, or may not, exhibit a shape bias. In Chapter 2, we proposed that use of shape dissimilarity as a cue to category exclusion may be an important difference between autistic and neurotypical children. In Chapter 3, by increasing the similarity of the stimuli, we also detected population differences for standard visibility and labelling effects. Autistic children benefited from comparing test items to the standard in 'online' trials, but not from labelling as TD children did. However, while this research serves to identify *when* autistic and neurotypical children differ in their use of the shape bias, the underlying cause of those differences lacks an adequate explanation.

In this final paper, we test the hypothesis that autistic children pay attention to small details of stimuli at the expense of the global shape due to 'weak central coherence' (Happé & Frith, 2006). Using the same screen-based game as the other experiments, this study uses stimuli designed to have a small but salient feature that could draw attention away from overall shape. If a local attentional bias explains weaker shape bias in autism, we would expect to see a higher likelihood of generalising labels to objects that are a feature match, particularly in the 'offline' condition when the shape must be remembered.

Author contributions:

Leigh Keating: study design, data collection, statistical analysis, manuscript writing, review.

Katherine Twomey: study design, review. *Calum Hartley*: study design, review.

4.2 Abstract

Typically developing (TD) children's acquisition of novel count nouns may be aided by automatic attentional processes, such as the 'shape bias'. While autistic children can generalise novel nouns by shape, their inconsistent application of shape bias across a range of tasks may be the result of attentional differences, such as 'weak central coherence' – a preference for processing local features before global level characteristics (Happé & Frith, 2006). This experiment investigated how autistic and TD children generalised novel nouns in cases where stimuli had a salient local feature that could distract from global shape. In a yes/no task, children indicated whether each item was another example of a newly introduced 'standard' in both labelled and unlabelled trials. On half of the trials, the standard remained visible for reference (online trials), while the other half required the standard to be briefly held in memory (offline trials). A mixed logit model found that local features were only significant predictors of 'yes' responses for items that were also a global shape-match to the standard for both autistic and TD participants. Furthermore, features only influenced responding in online trials, and no population differences or effect of labelling were identified. These findings indicate that both autistic and TD children prioritised shape for category membership, and only used local features as an additional cue when a standard was visible. Thus, a local attentional bias did not interfere with autistic children's shape bias, challenging the suitability of WCC as an explanation for differences identified in previous research.

4.3 Introduction

On hearing a novel object named for the first time, typically developing (TD) children over two years old spontaneously generalise that name to other objects that are the same shape (Landau et al., 1988), allowing them to quickly acquire new nouns and support vocabulary growth (Gershkoff-Stowe & Smith, 2004). For autistic children, language acquisition does not always appear so effortless, with some children demonstrating language delays, or no spoken language at all (Anderson et al., 2007). Recent research has investigated whether differences in autistic children's application of the 'shape bias' may contribute to delays in language development (Abdelaziz et al., 2018; Hartley & Allen, 2014; Keating et al., Chapter 2, Chapter 3; Potrzeba et al., 2015; Tek et al., 2008; Tovar et al., 2020). Existing evidence indicates population differences in shape-based generalisation between autistic and TD children, however, these differences are only evident in certain types of task (Keating et al., Chapter 2). As shape bias in TD children is supported by automatic attentional processes (Smith et al., 1996), attentional differences in autism could potentially explain atypical shape bias in certain tasks. Autism is characterised by 'weak central coherence' (WCC) – a tendency to attend to local features first before processing information at a global level (Happé & Frith, 2006). Here, we investigate whether WCC may account for differences in shape-based novel noun generalisation in autistic children.

4.3.1 Shape bias in typical development and autism

The shape bias was first documented by Landau et al. (1988), who explained it as the prioritisation of object shape over other perceptual features (e.g. size, colour, or texture) when generalising object names to new examples. As a perceptual feature, shape is a reliable cue for basic category membership (Rosch et al., 1976), whereas features like texture and colour are rarely unique to a single category, and so are less informative in isolation. A typical experimental task demonstrating the shape bias begins by introducing children to a novel object (commonly referred to as the 'standard'), often accompanied by a label (e.g. Landau et al., 1988). Then, children are introduced to new objects that match the standard on one or more perceptual features and asked to indicate if they are also examples of the standard. These 'novel noun generalisation' (NNG) tasks can be forced-choice, in which children choose one of several candidate objects as belonging to the same category as the standard, or 'yes/no', in which children are asked to make a category judgement on each item in turn. The forced-choice task has been the most frequently replicated, and English-learning TD children usually prioritise shape over other characteristics once they have around 50-150 count nouns in their receptive vocabulary (Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999). For rigid objects, young children show the same shape preference in both yes/no and forced-choice tasks (Samuelson et al., 2009).

It has been argued that young TD children's attention to shape is an automatic process, which is enhanced when objects are named. In a yes/no NNG task, Smith et al. (1996) used stimuli constructed from a large 'base', which formed the overall shape of the object, and a small yet salient 'part'. They found that 3-year-olds generalised labels to items with the same base/shape as the standard, but not differently shaped items that shared the part. However, when children were asked to make a similarity judgement without items being named, only exact matches to the standard received a high proportion of 'yes' responses indicating category membership. This finding suggests that hearing a novel count noun cues TD children's attention towards global shape. Automatic attentional biases, such as the shape bias, and early language acquisition are thought to develop in tandem: as the number of shape-based nouns in child's vocabulary increases, attention to shape also increases, resulting in more efficient word learning and faster vocabulary growth (Smith et al., 2002).

However, in autism, delays in language development are common, and up to one third of autistic children may have no spoken language by the age of nine years (Anderson et al., 2007). Reduced attention to shape and less efficient language acquisition in comparison to TD children could contribute to their language learning difficulties. Recent evidence suggests that autistic children may have difficulty using a shape bias under certain circumstances. In a sorting task, Hartley and Allen (2014) introduced children to a novel item (e.g. a *blicket*) and asked them to sort objects and pictures of objects into either a box for '*blickets*' or a box for 'other things'. While TD participants tended to only include items that were the same shape as the standard in the 'blicket' box, autistic children were prone to over-generalise and include items that matched on shape and/or colour in the new category. These findings have been replicated in another sorting task (Tovar et al., 2020), and also support the findings of our recent studies using screen-based 'yes/no' tasks (Keating et al., Chapter 2, Chapter 3). In Keating et al.'s yes/no task (in prep, see Chapter 2), both TD and autistic children were likely to include shape-match items in the same category as the standard. However, the groups differed significantly in their decisions on items that differed in shape. While the TD participants were also likely to reject colour- and texture-match items, the autistic participants were more likely to accept the distractors too. We proposed that autistic children associate shapes with count nouns, however, they do not do so exclusively and may not 'rule out' items that are a different shape to the standard as TD children appear to do.

Furthermore, naming items may not support automatic attention to shape for autistic children. Tek and colleagues (2008) used an intermodal preferential looking (IPL) paradigm to investigate shape bias with a method that did not require an explicit response, specifically looking times to two images of test items. Shape-match and colour-match competitors were displayed on a screen, and children's looking time to each item was recorded for trials with labels ("Where's the dax?") and without labels ("Which one looks the same?"). While TD children looked more towards the shape-match object in the label condition compared to the no-label condition, looking times for autistic children were not influenced by naming. Population differences associated with labelling have also been reported in a forced-choice NNG task (Keating et al., Chapter 3). For TD children, labels resulted in a stronger shape bias, with shape-match items more likely to be chosen over colour-match or texture-match competitors when the standard was named. Autistic children also showed a shape bias in the same task, but they were unaffected by having the standard named.

If automatic attention to shape supports TD children's word learning (Landau et al., 1988; Smith et al., 2002), it is essential to consider the influence of attentional differences on language acquisition in autism. The proponents of WCC originally proposed that autism is characterised by an advantage in local processing at the expense of global processing. However, modern versions of this theory (Happé & Booth, 2008) now argue that there is no deficit in global processing in autism, but that there may be a bias towards local processing first. WCC has been presented as an explanation for differences in autistic individuals' responding across various visuo-spatial and auditory tasks (see Happé & Frith, 2006 for review).

While TD individuals prioritise processing global characteristics before processing local elements (Navon, 1977), research with autistic participants suggests the opposite may be true for them. Local level processing may occur first, with global processing only occurring once that initial attentional pull has been overcome. In a meta-analysis, Van der Hallen et al. (2015)

concluded that global processing is slower in autism while local level processing may be more automatic, indicating a difference in default prioritisation of attention rather than a deficit in global processing. If this preference applies in a novel noun learning situation, not only could TD children attend to the global shape first, but it would be unnecessary for them to process local details at all in order to demonstrate a shape bias. Conversely, in order for autistic children to attend to the global shape, they need to first process the local details, posing more risk of interference on an individual item level.

Most studies that report differences in autistic children's shape bias have implicated WCC as a potential cause. Specifically, local attention to small parts or irrelevant details of objects may distract attention away from the global shape (Hartley & Allen, 2014; Potrzeba et al., 2015; Tek et al., 2008), whereas TD children generalise by shape due to their tendency to prioritise processing of global (shape) features. Furthermore, Potrzeba et al. (2015) suggested that a local attention bias could disrupt or delay the process of learning to make higher order generalisations on the basis of shape. Consequently, it is plausible that WCC could affect autistic children's attention in-the-moment during a NNG task. If autistic children are more likely to prioritise attention to local details in a NNG task, that would provide evidence that WCC may have a significant impact on their language acquisition.

4.3.2 Task demands on shape bias in autism

While there is strong evidence of shape bias differences in autism, there are also cases where autistic children exhibit a shape bias that is indistinguishable from TD children matched on receptive vocabulary. With a screen-based forced-choice task (Keating et al., Chapter 2, Chapter 3), autistic children prioritised object shape over other perceptual features as expected for their vocabulary level with a variety of different stimuli. Field et al. (2016b) also found a shape bias in autism on a similar task using physical objects. These findings contrast with evidence of weaker shape bias autism, however, differing task demands may offer an explanation. By considering how methodological variations impact how shape bias presents in a task, we may be better able to identify which underlying processes are involved (Kucker et al., 2019). Keating and colleagues investigated the impact of task differences using screen-based versions of both the forced-choice and yes/no NNG tasks with high contrast (Keating et al., Chapter 2) and low contrast (Keating et al., Chapter 3) stimuli. In both studies, autistic children exhibited a similar shape bias to TD children on the forced-choice task, however, their shape bias was disrupted when the same stimuli were presented in a yes/no task. Furthermore, when the test stimuli shared shape commonalities with the standard, we found that shape bias was influenced by the ability to see the standard for comparison during generalisation. In 'online' trials, where children were able to refer to the standard at all times, preference for shape was similarly strong for both groups. However, in 'offline' trials, where the standard had to be briefly held in memory, autistic children exhibited a weaker shape bias than TD children in the forced-choice task. TD children were also affected by the standard visibility, but only in the yes/no version of the task. In the forced-choice task, the memory requirement did not impact TD children's shape bias. A similar pattern was observed in the IPL task compared to a force-choice task in Tek et al. (2008); despite exhibiting weaker shape bias in the offline IPL task, the same autistic children preferred shape-match items in a pointing task where the standard remained in view on the table.

The finding that shape bias differs for autistic and TD children in online and offline tasks is relevant to the role of automatic attention to shape in NNG tasks. A stronger attentional pull towards global shape for TD children could support shape bias in tasks where the standard is not visible. Conversely, a default attention to detail in autism may cause a disruption in offline tasks that are not observed when children are allowed to make direct comparisons to a visible example.

4.3.3 Aims of the current research

The objective of this study was to investigate whether a local attentional bias in autism accounts for disrupted generalisation in yes/no NNG tasks. We tested this possibility by asking children to generalise labels across novel objects with salient local features using a yes/no task. This allowed children freedom over their strategy, so they could generalise by global shape, by feature, or be more conservative and only generalise if both shape and feature matched the standard. Furthermore, this task has been effective for revealing population differences in previous work (Keating et al., Chapter 2, Chapter 3). To preserve TD children's global shape preference, local features were small and positioned so that they did not overlap with the outer contour of the item. Additionally, the features had no obvious function. While we expected that TD children would attend to shape regardless, there is evidence that autistic children may prefer to generalise by function (Field et al., 2016a), so features were intended to appear decorative only.

We expected that, overall, objects that were the same global shape as the standard would be more likely to receive 'yes' responses, and this shape preference would be greater for labelled items than for unlabelled items. In line with the WCC hypothesis, we predicted that autistic children would accept more feature-match objects than TD children. We also suspected that labels would influence feature-match acceptance. While the most likely outcome would be that labels would enhance automatic attention to global shape (Smith et al., 1996), and hence reduce feature acceptance, it is possible that labels could enhance attention to any commonality (e.g.

Althaus & Plunkett, 2015) resulting in a preference for both global shape and local feature to match the standard. For this reason, our hypothesis for local feature was non-directional.

Furthermore, we were interested in whether visibility of the standard would interact with attention to the shape or feature. There were two contrasting hypotheses on the interaction between standard visibility and population. Hypothesis 1 predicted shape acceptance would be weaker in the offline condition for autistic participants, as WCC suggests the local feature should be attended to first and so have a stronger working memory representation. Alternatively, hypothesis 2 predicted a higher probability of shape-match acceptance for both groups in the offline condition. This outcome would be expected if greater initial attention was paid to shape rather than the salient local feature (i.e. neither group would display a local processing bias).

4.4 Method

4.4.1 Participants

Participants were recruited from preschools, mainstream schools, and specialist schools in the Northwest of England. The sample for this study included 17 autistic (ASD) children (*M* age = 6.86 years; *SD* = 1.38 years) and 17 typically developing children (M age = 3.70 years, *SD* = 0.49 years) matched on receptive vocabulary as measured by the British Picture Vocabulary Scale (BPVS), 2nd edition (Dunn et al., 1997). There was no significant difference in age equivalent vocabulary scores between the groups (ASD: *M* BPVS age equivalent = 4.27 years; *SD* = 1.48 years, TD: *M* BPVS age equivalent = 4.00 years; SD = 0.88 years), t (26.03) = -0.65, p = .524, two-tailed). Expressive vocabulary was measured using the Expressive Vocabulary Test (EVT), 2nd edition (Williams, 2007), and non-verbal cognition was measured with the cognitive battery subscale of Leiter-3 (Roid et al., 2013). Due to testing being interrupted by school closures during the Covid-19 pandemic, many children were unable to complete these measures (EVT: 7 autistic and 11 TD, Leiter-3: 11 autistic and 11 TD). The available data are summarised in Table 1 .

A further 11 autistic children involved in the overall project completed this task, but were not included in the analysis as their receptive vocabulary age equivalent was higher than 6.5 years (n = 8) or they were unable to complete the BPVS (n = 3). This decision was taken to ensure the receptive vocabulary of the autism group was similar to that of the TD group, however, the excluded data are available on OSF for future analysis: https://osf.io/sgcex/. Additionally, data for one autistic participant were excluded due to failure to complete the experimental task.

Of the 34 children who completed this task, 29 of them (15 autistic and 14 TD) also took part in other experiments that were part of the project on different days (see Keating et al., in prep; Chapters 2 and 3). The order of task completion was counterbalanced for participants who took part in every study. The remaining five participants took part in the current study only.

This research was approved by the Lancaster University Ethics committee (REF: FST18017). Informed consent was obtained from caregivers of each child who took part. All participants received stickers throughout testing as rewards. At the end of the final session, they were also given a choice of book to take home with them, along with a debrief sheet for their caregiver. Participating schools were also given an Amazon gift card as a thank you gift for supporting the research.

Table 1: Sample description (standard deviation and range in parentheses)

Population	*N*	Chron. Age (years)	BPVS age equiv. (years)	EVT age equiv. (years)	Leiter-3
TD	17	3.70 (0.48; 2.75-4.67)	4.00 (0.85; 3.00-5.42)	4.79 (0.68; 3.75-5.50)	46.83 (8.98; 34-62)
ASD	17	6.85 (1.34; 4.58-9.08)	4.27 (1.44; 2.5-6.42)	3.64 (1.03; 2.17-4.92)	37.5 (13.94; 24-56)

4.4.2 Design

The study employed a mixed (2 x 2 x 2) design with one between-participants factor (Population: typical development, autism) and two within-participants factors (Standard visibility: online, offline; Label: label, no-label). Participants completed the online and offline conditions in a counterbalanced order. No-label trials were always completed first to avoid the label trials priming children to generalise by shape.

4.4.3 Materials

4.4.3.1 Visual stimuli

Novel objects were computer-generated 3D models created using Blender (Version 2.78; Blender Development Team, 2017), an open source software suite suitable for modelling and animating 3D assets. Each object was designed to look like a realistic, solid artifact of uniform colour made from plastic, with a feature of a contrasting colour and different material. Contrasting colours and prominent positioning ensured that features were clearly visible, however, they did not alter the outer contour of the global shape. Four sets of eight objects were created, each with one object designated as the standard. For instance, the Set 1 standard was a slotted turquoise block with three cylinders as the base, and a gold, metal sphere on the front as the feature. Sets also included a contrasting item with no characteristic congruent with the standard (e.g. a green, spiked block base with a red velvet lattice feature). This ensured each set contained an equal number of stimuli that matched or differed on each characteristic (e.g. four the same shape and four of the alternate shape). Of the remaining items in the set, six were constructed from each combination of global shape and feature, with varying colour. The final item of each set was an unrelated control item that was not perceptually similar to either the standard or any of the test items. All four sets are presented in Figure 1 .

Visual stimuli used in the experiment were a static image of each object, rendered from the same angle, and images of a red circle containing a white cross and a green circle with a white tick. We also used Blender to create a two-second video of the standard rotating on its y-axis, followed by an x-axis and y-axis spin, beginning and ending at the same viewpoint as the static image so that the clip could be looped continuously.

Static images and a video clip were also created for use in known object practice trials. These were created from pre-existing 3D models available in Microsoft 3D Paint. The known object images were a green bed and a blue butterfly. A two-second video of the green bed rotating with the same motion described for the novel objects was recorded using MS Powerpoint.

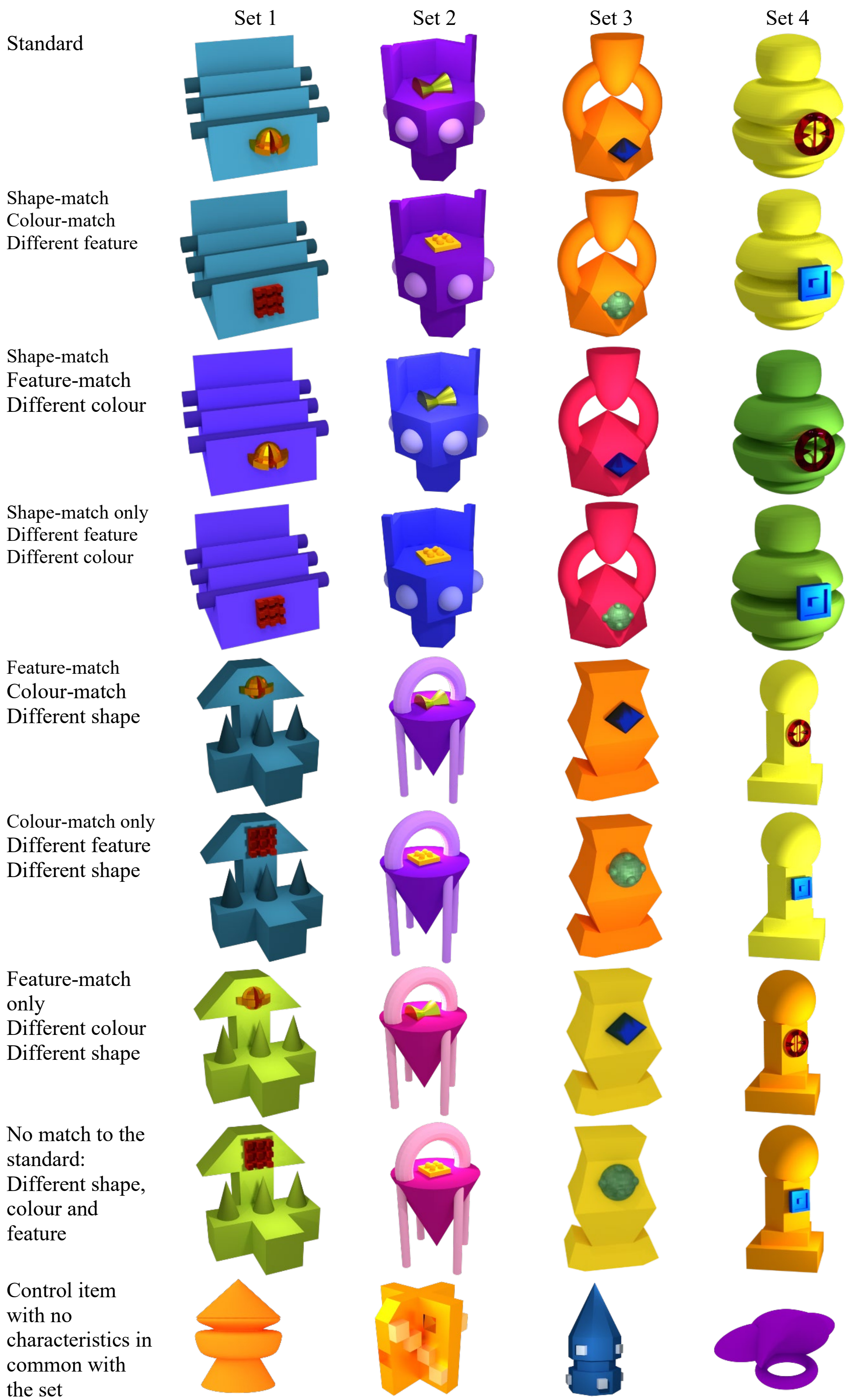


Figure 1: Complete stimuli sets 1-4

4.4.3.2 Audio stimuli

Audio stimuli were created using Audacity® recording and editing software (Version 2.1.3; Audacity Team, 2017), and featured a female voice with a British English accent. The wording and timings for each track are detailed in Table 2. Two plausible English non-words, *numble* and *tookie*, were selected from the English Lexicon Project (Balota et al., 2007).

Table 2: scripts for audio stimuli

Wording	Clip duration (seconds)
"Look at that!"	0.9
"Look! A numble!"	1.6
"Look! A tookie!"	1.4
"Is this another one?"	1.1
"Is this a numble?"	1
"Is this a tookie?"	1
"Perfect"	0.8
"Uh oh! Try again."	1.7

4.4.3.3 Presentation software

A Microsoft Surface 4 tablet laptop running Windows 10 was used to administer the study. The visual and audio stimuli were presented using Psychopy 2 (Peirce, 2007). The programme ran five mini-games which formed the complete project, in a fixed order that was counterbalanced for each participants. The four other mini-games are described in previous papers (Keating et al., Chapter 2, Chapter 3). All timings, randomisation of positions and allocation of stimuli to conditions was managed by the programme, as was the recording of participants responses. Figure 2 shows a typical layout of the screen while waiting for a child's response. Touching the 'tick' or the 'cross' button would be logged as a response, whereas tapping anywhere else on the screen had no effect.

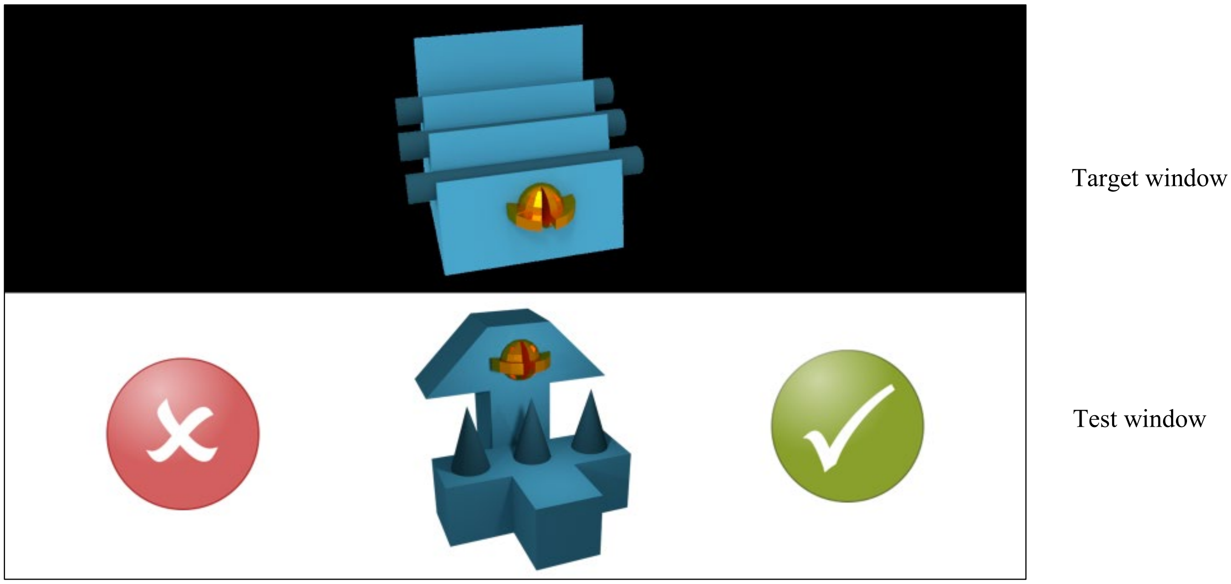


Figure 2: Example of the visual display on screen during an online trial

We also created an interactive PowerPoint presentation designed to familiarise children with the yes/no task. The presentation consisted of six slides, each with a static image of a known item (a horse, a ball, or a banana, each seen twice) and the same green tick and red cross circles used in the main experiment. A word was printed above the item for the experimenter to read out. On three slides the word was the same as the image, while three had the wrong word (*shoe*; *train*; *elephant*). The slideshow was programmed to progress only when the correct button (tick or cross) was pressed.

The code to recreate the experiment is freely available on the OSF repository associated with the project: https://osf.io/sgcex/.

4.4.4 Procedure

Participants took part in this study as part of a three-day project which included two other studies. The order of study completion was counterbalanced, so the current experiment could take place on the first, second, or third visit on different days. The gap between testing was between 1 to 7 days. Sessions took place in a quiet room at the child's own school or nursery, and a familiar adult was invited to observe. A typical session began by completing a standardised (either BPVS, expressive vocabulary test, or Leiter 3) followed by a warm-up exercise and then the main experimental task.

4.4.4.1 Warm-up task

The experimenter informed the child that they were about to play a game on a computer that was going to ask them some questions, so first they were going to practice. The Surface Pro tablet computer was either stood on the desk in front of the child or given to them to hold on their lap if they preferred, with the button training PowerPoint presentation loaded. The first slide showed an image of a banana along with the tick and cross buttons, and the text '*shoe?'* at the top of the screen. The experimenter asked if it was a picture of a shoe. When the child indicated it was not, either verbally or by shaking their head, the experimenter agreed and asked them which button meant 'no'. If the child correctly identified the red button with the cross, they were then encouraged to try tapping it. If they did not know, the experimenter told them which button to use and invited them to try tapping it. When the red button was tapped, the presentation moved on to the next slide. The next image depicted a horse with the text '*horse?'*, so the process was repeated to model a 'yes' response. Participants were then encouraged to try the remaining four slides alone, only receiving reminders from the experimenter if they were unsure of what was expected. If, after completing all six slides, children had not managed to use the buttons correctly three times in a row, the presentation was restarted and they could try again. All participants gave three consecutive correct responses within 12 practice slides.

4.4.4.2 The NNG task

Following the warm-up task, the experimenter started the experiment software on the computer. The task then followed the pre-programmed procedure and required no further intervention from the experimenter. Participant responses were logged by the software. On start-up, the computer programme randomly distributed the four stimuli sets to each of the four condition combinations (Online/Offline x Label/No-Label), and labels were randomly assigned to one of the two labelling condition sets. The experiment consisted of two main blocks, one for the online condition and one for the offline condition. Each block began with a practice trial, followed by all trials for the first unlabelled set, then all trials for the first labelled set. At the end of the block, the second block began automatically with practice trials for the second online/offline condition. Once both blocks were completed, children had responded to all 36 objects across the four sets and the experiment ended.

4.4.4.2.1 Practice trials

Each block began with a practice trial consisting of a familiarisation phase with known items, which was identical in both conditions, and a generalisation phase. This phase was designed to familiarise children with test trials in which the standard was not visible (Offline condition).

4.4.4.2.2 Familiarisation phase

Children saw the animation of a rotating green bed and heard '*look at that!'*. After one complete two-second loop of the video, a '*next'* button appeared. Children could then choose to end the video or continue to watch for up to 30 seconds, after which the familiarisation phase would end automatically. As autistic children can benefit from having more time to process perceptual information for generalisation tasks (e.g. Hartley et al., 2020), child-led timing on the familiarisation phase allowed them to look for longer if they wished to.

4.4.4.2.3 Generalisation phase

The screen was divided into a black top half and a white bottom half. The top half always displayed the standard, while the bottom half always presented test items and the response buttons. The clear division of the display area was intended to mimic similar tasks with real objects where candidate items are presented on a tray and the standard is either separate or not visible. The screen layout is shown in Figure 2.

For the online condition, a static image of the bed appeared at the top of the screen for one second. Then an exact copy appeared in the bottom half of the screen while the audio track asked '*Is this another one?'*. After the audio track finished, the green and red response buttons appeared on either side of the lower bed. If children pressed the green button, the '*Perfect'* audio track played and the trial progressed. If children pressed the red button, the '*Uh oh! Try again'* track played, and the programme waited until the correct response was given. The second practice trial followed the same procedure, but presented a blue butterfly in the lower half of the screen, and required a red 'no' response before the experiment would proceed.

In the offline condition, following familiarisation, the static bed image appeared for one second in the top window and then disappeared. At the same time, the bed image appeared in the bottom half and the same audio track played. The procedure with the response buttons was identical to the online condition. Once a correct response was given, the green bed was displayed again, alone on the top half of the screen, for one second, and disappeared when the blue butterfly was presented.

After two correct responses were gained for the two practice trials, the programme progressed to the main experimental block for the Online/Offline condition that had just been practiced.

4.4.4.3 Experimental trials

Children completed either the online or offline block first (counterbalanced order) and were presented with one stimuli set without labels followed by a second set with labels.

4.4.4.3.1 Exemplar familiarisation phase

Familiarisation proceeded as described in the practice block, but this time the rotation animation was of the standard for one of the four stimuli sets. The first familiarisation phase was always for a 'no label' set and used the same audio track as in the practice block ('*Look at that!*'). Children could manually move on after two seconds or watch the animation loop for a maximum of 30 seconds.

For familiarisation of a labelled exemplar, the audio track included the label ('*Look! A tookie/numble*') but was otherwise identical.

4.4.4.3.2 Generalisation phase

This phase consisted of nine trials for each stimuli set, one for each of the eight item variations plus one for the control item (see Figure 2). The timings and position of the stimuli for the generalisation phase of the experimental trials were the same as in the practice block. A static image of the standard for the set appeared on the top half of the screen. In the online condition, the standard remained visible for the duration of all nine trials. In the offline condition, it appeared for one second at the beginning of each trial as a reminder, then disappeared before the object to be considered was displayed. After the standard had been displayed for one second, one of the eight candidate items appeared on the bottom half of the screen. The order was randomly determined by the computer, with the exception that the control item was always presented on the final trial. The audio track played ('*Is this another one?*') before the tick and cross response buttons appeared, to prevent children answering before hearing the audio. Once the audio finished and the response buttons appeared, the programme waited until the child gave a response before proceeding to the next trial. The selected response button was highlighted with a green circle for 1 second so children could clearly see what they had chosen. During this time, children were allowed to change their answer, and only the final answer was logged. Unlike the practice trial, no feedback was given on any experimental trials. After one second, the bottom half of the screen was cleared and the next trial began.

Once children had responded to all nine items in the no-label condition, the familiarisation of the exemplar for the label condition began. The procedure was the same, but the labelled versions of the audio tracks were used (‘*Is this a numble/tookie?’*). When both labelled and unlabelled trials for the first block were completed, a new known item practice phase began for the second block. Figure 3 shows the screen display for each stage of the procedure.

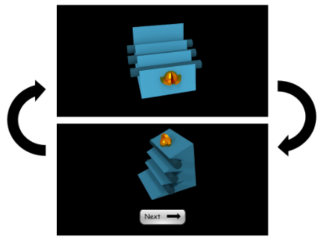
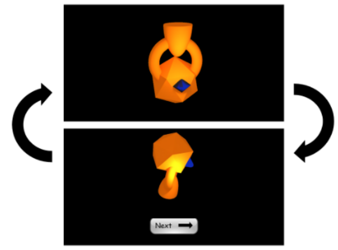
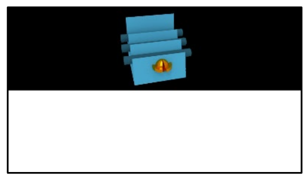
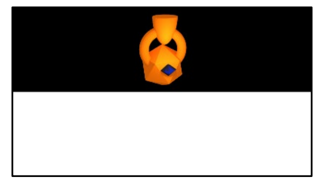
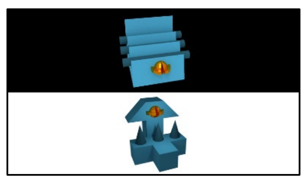
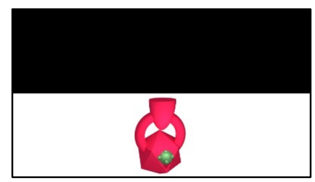
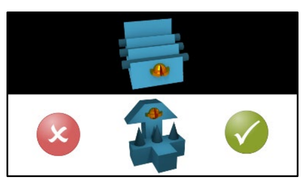
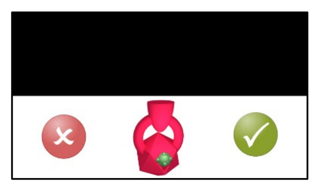
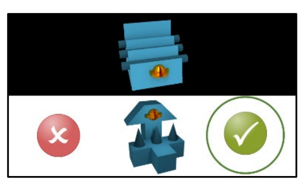
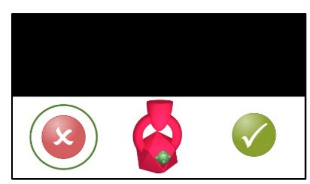
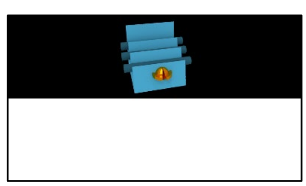
Stages and timings	**Audio track (label/no label conditions)**	**Online block visual display**	**Offline block visual display**
Familiarisation phase: Animation of Standard rotating for up to 30 seconds	"Look! A tookie!" *or* "Look at that!"		
Test phase Stage 1: 1 second display time	*No audio*		
Stage 2: Audio presentation (approx. 3 seconds)	"Is this another tookie?" *or* "Is this another one?"		
Stage 3: wait for participant response	*No audio*		
Stage 4: Response confirmation displayed for 1 second	*No audio*		
Repeat from stage 1 for next triad in the stimuli set	*No audio*		

Figure 3: Visual display, timing and audio for an example trial

4.5 Results

4.5.1 Data analysis

All statistical modelling and post-hoc comparisons described were conducted with the R libraries “lme4” (Bates et al., 2015) and “emmeans” (Lenth, 2022) in RStudio, Version 1.3.1073. The dependent variable was the binary response on each individual trial: ‘tick’ (coded 1) or ‘cross’ (coded 0). To determine if these raw responses were consistent with a perceptual bias, binary item variables were coded to indicate which dimensions the test item matched the standard on. The item variables were ‘global shape’, ‘local feature’, and ‘colour’. Each variable could be ‘congruent’ (1) with the standard or ‘incongruent’ (0). The within-participants independent variables were standard visibility (contrast coded: online: -0.5, offline: 0.5) and labels (contrast coded: no-label: -0.5, label: 0.5). Population was included as a between-participants variable, with typical development coded as -0.5 and autism coded 0.5. There were 32 experimental trials for each participant.

4.5.2 Item effects and fixed effects modelling

The analysis was conducted using generalised linear mixed-effects models for binomially distributed outcomes (also known as a ‘mixed logit model’), with the raw response to each trial modelled as the outcome. First, the random effects structure was determined by reducing the complexity of the maximal structure until the model converged (Barr et al., 2013). This established the baseline model for comparison, and contained a random intercept for participants, and a random slope for labels by participant.

Item variables were added to the model first in a single step. If ‘tick’ responses were more likely when global shape was congruent, and not affected by variations in colour or feature, this would be consistent with a shape bias. In the pre-registered plan, we stated that we would test whether global shape, local feature, and item colour predicted the likelihood of items being accepted, with the intention of ruling out colour as a source of variance. This was tested by comparing the change in deviance from the baseline model using a likelihood ratio test (Bates et al., 2015). The full item effects model was a significant improvement on the baseline ($\chi^2(7)$ = 140.08, p < .001). As predicted, colour was not a significant predictor of children's responses, so was removed without reducing model fit ($\chi^2(4)$ = 1.39, p = .846). Table 3 shows the summary of the best-fitting item effects model. As planned, only the effects of shape and feature were considered as interactions with the experimental condition variables.

Table 3: Summary of fixed effects and random effects for best fitting model of item characteristics

Effect	Group	Term	Estimate	Std Error	z value	p value	
fixed		(Intercept)	-0.71	0.27	-2.61	.009	**
		Global Shape (congruent)	0.99	0.20	4.89	<.001	***
		Local Feature (congruent)	-0.02	0.21	-0.10	.919	
		Global Shape x Local Feature	1.12	0.29	3.81	<.001	***
			Variance	Std. Deviation	Corr		
random	participant	(Intercept)	1.54	1.24			
		labels	0.53	0.73	-0.04		
AIC	1226.5				logLik	-606.3	
BIC	1261.4				deviance	1212.5	
					df residual	1078	

Number of obs: 1085, participants: 34
Significance indicators: p>.05*; p>.01**; p>.001***

The fixed effects for the experimental conditions (task type, labels, and population) were added to the item model, along with interaction terms with the item variables. This model was significantly better fitting than the item-only model ($\chi^2(12)$ = 30.91, p < .001). To investigate the effect of receptive vocabulary on item acceptance, we added the mean-centred BPVS age equivalent scores as a covariate. The addition of receptive vocabulary further improved model fit ($\chi^2(4) = 12.11$, $p = .017$). The final fixed effects model is reported in Table 4.

Table 4: Summary of best fitting model

Effect	Group	Term	β	Std Error	*z* value	*p* value	
fixed		Intercept	-0.76	0.27	-2.84	.005	**
		Global shape (congruent)	1.06	0.21	4.98	<.001	***
		Local feature (congruent)	0.00	0.21	-0.01	.990	
		Population	0.87	0.54	1.62	.105	
		Labels	-0.29	0.33	-0.88	.381	
		Standard visibility	0.12	0.30	0.42	.676	
		Receptive vocabulary	-0.03	0.02	-1.60	.109	
		Shape (congruent) x Feature (congruent)	1.14	0.30	3.75	<.001	***
		Shape (congruent) x Population	-0.43	0.42	-1.02	.309	
		Feature (congruent) x Population	-0.24	0.42	-0.57	.571	
		Shape (congruent) x Labels	1.09	0.42	2.58	.010	**
		Feature (congruent) x Labels	0.46	0.42	1.09	.274	
		Shape (congruent) x Standard visibility	0.73	0.41	1.77	.077	.
		Feature (congruent) x Standard visibility	0.21	0.42	0.50	.614	
		Shape (congruent) x Receptive vocabulary	0.03	0.02	2.23	.026	*
		Feature (congruent) x Receptive vocabulary	0.01	0.02	0.85	.398	
		Shape (congruent) x Feature (congruent) x Population	-0.35	0.61	-0.57	.569	
		Shape (congruent) x Feature (congruent) x Labels	-0.51	0.60	-0.84	.402	
		Shape (congruent) x Feature (congruent) x Standard visibility	-1.72	0.60	-2.85	.004	**
		Shape (congruent) x Feature (congruent) x Receptive vocabulary	0.01	0.02	0.23	.816	
			Variance	Std. Deviation	Corr		
random	participant	(Intercept)	1.52	1.23			
		labels	0.52	0.72	0.17		
AIC	1215.5				logLik	-584.7	
BIC	1330.3				deviance	1169.5	
					df residual	1062	

Number of obs: 1085, participants: 34

Significance indicators: *p*>.05*; *p*>.01**; *p*>.001***

4.5.3 Global shape and local feature

The model intercept shows the log odds for the estimated probability of acceptance for items that had neither global shape nor local feature in common with the standard. Items with these characteristics were significantly less likely to receive a 'tick' response than the 0.5 chance level. In comparison, objects that matched on global shape were significantly more likely to be accepted, as would be expected in a standard shape bias task. The local feature alone was not a significant predictor in the model, nor were the two-way interactions between feature and population, labels, standard visibility, or receptive vocabulary. However, local feature did interact significantly with global shape. Post-hoc comparisons showed that, for items that were the same global shape as the standard, there was no difference in probability (P) of acceptance on the basis of local feature: (feature congruent: $P = 0.32$; feature incongruent: $P = 0.32$; $z = 0.01$, $p = .990$). When global shape was a match, however, items with a standard-congruent local feature had a significantly higher probability of acceptance ($P = 0.81$) compared with incongruent features ($P = 0.57$): $z = -5.197$, $p < .001$ (see Figure 4 for model estimates and Figure 5 for observed proportional data). This suggests that congruent local features enhanced the probability of accepting a congruent global shape, but did not increase the chances of accepting a differently-shaped distractor.

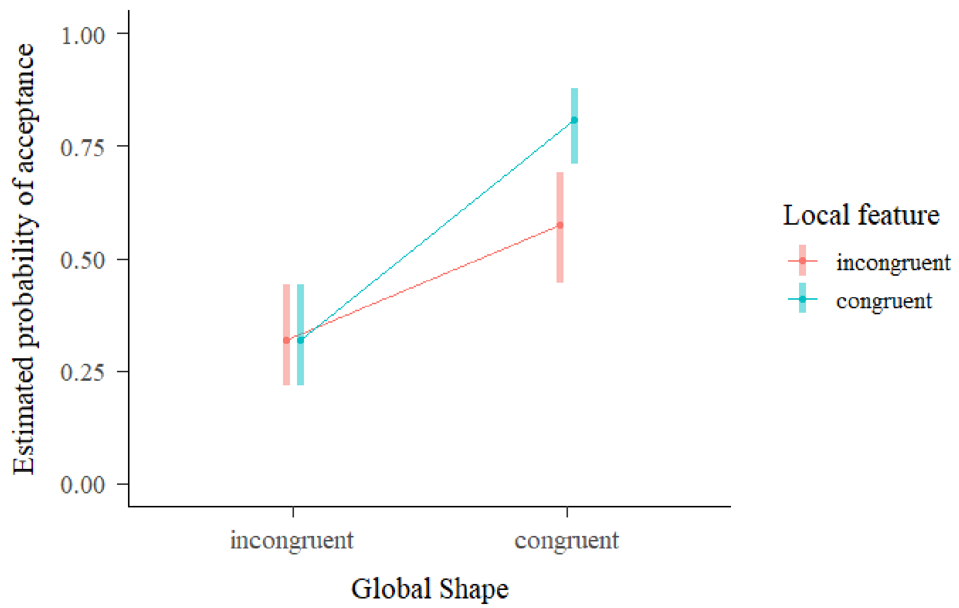


Figure 4: Model estimates of local feature match acceptance for levels of global shape

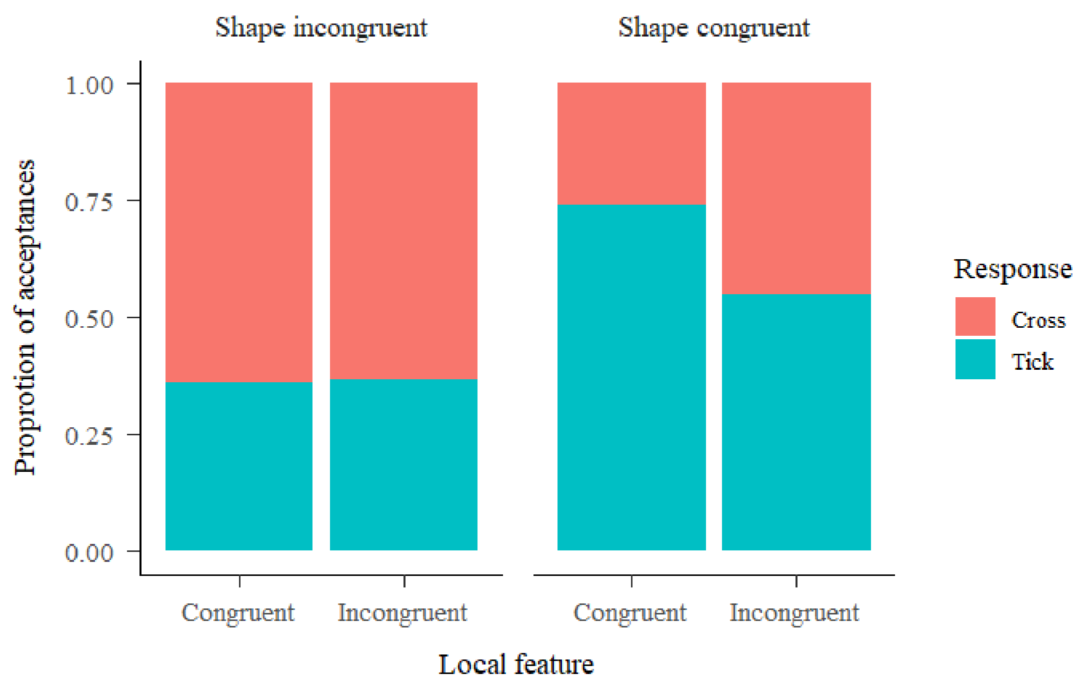


Figure 5: observed proportional data for local feature acceptance for global shape congruent or incongruent items

4.5.4 Population

Population was not a significant predictor in the final model, nor were any interactions including Population. Thus, we found no evidence that autistic children performed significantly differently to TD children when responding in this task (Figure 6).

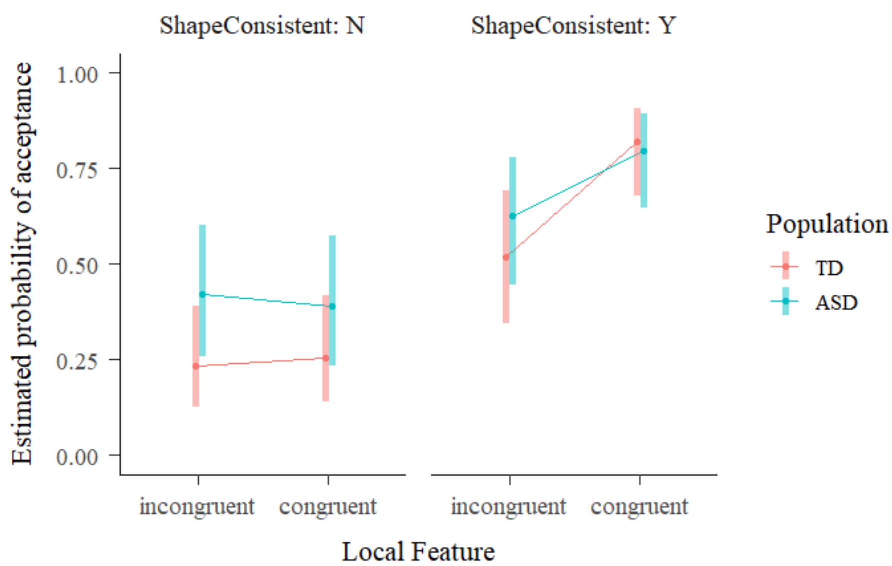


Figure 6: Estimates of local feature acceptance for levels of global shape by population

4.5.5 Labels

There was a significant Shape-congruent x Labels interaction: there was a greater effect of labels on item acceptance when the test item was a global shape-match for the standard. Post-hoc tests confirmed that global shape-congruent items had a significantly higher probability of being accepted in label trials ($P = 0.78$) than no-label trials ($P = 0.62$), and this difference was significant regardless of the local feature (Figure 7). Labelling made no significant difference for items with an incongruent global shape (label: $P = 0.31$, no-label: $P = 0.33$; $z = 0.23$, $p = 0.821$). In other words, labels only enhanced same-shape acceptance, and did not make a significant difference to the odds of rejecting distractor items of a different shape.

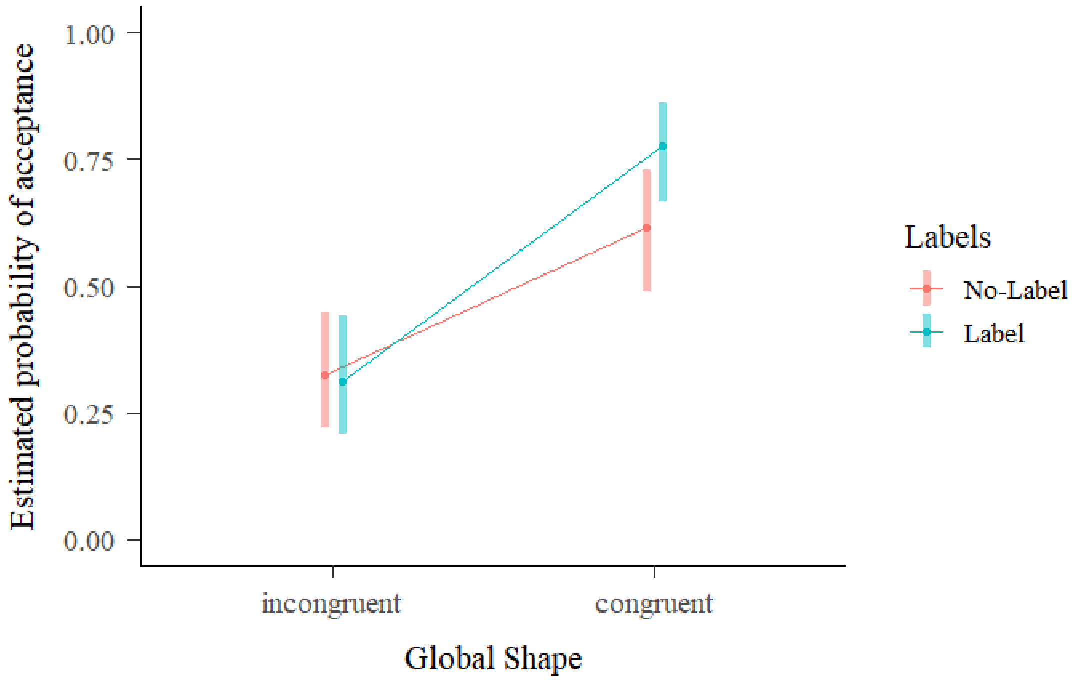


Figure 7: Estimated probability of accepting items with a global shape that was congruent or incongruent with the standard for label and no-label trials

4.5.6 Standard visibility

For standard visibility, there was a significant two-way interaction with global shape, and a three-way interaction including local feature congruence. Post-hoc comparisons of the effect of standard visibility for different item variable combinations confirmed that differences were only significant for global shape-match items. For items where both shape and feature were congruent with the standard, the probability of acceptance was significantly higher in the online condition ($P = 0.85$) compared to the offline condition ($P = 0.75$): $z = 2.03$, p =.042. For global shape-congruent items with the different local feature, however, acceptance was more likely in the offline condition (online: $P = 0.47$, offline: $P = 0.67$; $z = -2.98$, $p = .003$). When global shape was not a match to the standard, there was no significant difference between online and offline trials (see Figure 8). Furthermore, item acceptance was only significantly affected by local feature for shape-match items in the online condition (online: $P = 0.85$, offline: P =0.47; z = -

5.96, $p < .001$). Feature-congruent and feature-incongruent items did not significantly differ in acceptance probability for items that were a different global shape, or same shape items in the offline condition. This suggests that local feature only affected acceptance when the standard was visible. Overall, the odds of rejecting an item that was not a shape match did not change as a result of labels or standard visibility. Only shape-match acceptance decisions were predicted by the experimental conditions.

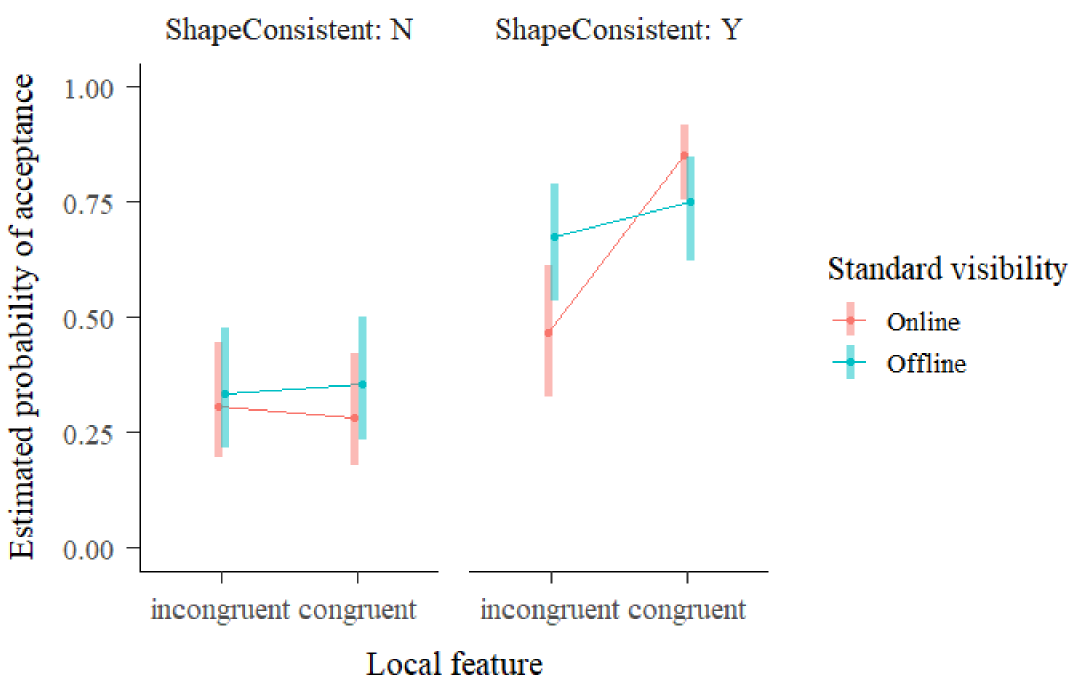


Figure 8:Probability of accepting items with congruent and incongruent global and local features in online and offline trials

4.5.7 Vocabulary

Receptive vocabulary was not a significant predictor for the baseline item characteristics, indicating the model did not detect an effect for items with incongruent global shapes and local features. The interaction between vocabulary and global shape was significant, with higher receptive vocabulary scores associated with an increasing probability of accepting global shape-match items, and a decreased probability of accepting differently-shaped items. However, post hoc comparisons indicated that acceptance probability did not significantly differ between the upper and lower measures of receptive vocabulary for either shape congruent or incongruent items.

4.6 Discussion

This study investigated whether the presence of a salient local feature interfered with shape bias for autistic children. Our findings indicate that global shape was good predictor of object acceptance for both autistic and TD participants overall – in line with Chapters 2 and 3, both groups demonstrated a shape bias. Higher probabilities of accepting local feature matches were only found for shape-match items. This suggests that both populations took the local feature into account in the NNG task, but only after similarity of global shape had been considered. If the item was a different global shape to the standard, the feature had no effect on acceptance likelihood. Furthermore, features had a greater impact when children could see the standard for direct comparison in the online condition. We found no evidence of local features overriding attention to shape for autistic children. Overall, these findings suggest that global shape was prioritised as the primary cue for generalisation by both groups; the local feature was only used to provide additional confirmatory information and did not affect generalisation if the item could be ruled out on the basis of global shape. This means our prediction that autistic children would accept more items with congruent local features was not supported. On the contrary, the autistic participants did not differ significantly from the TD participants in any condition. Both groups preferred test objects to match the standard on both global shape and local feature, however neither group generalised by feature *instead* of global shape. Furthermore, while we did observe an effect of standard visibility as predicted, the direction did not support the WCC hypothesis. Acceptance by local feature did not increase when the standard had to be held in working memory. Instead we found the opposite: both groups were more likely to use both the shape and feature when they could see the standard, but generalised by shape alone in the offline condition.

The finding that TD children prioritised global over local details in this study was consistent with Smith et al. (1996), who found that 3-year-olds generalised labels by global shape and not a local part. TD children in the current study also prioritised shape, and did so more strongly for labelled items. Smith et al. (1996) differed in that there was no generalisation trials where both the shape and part matched the standard while another characteristic differed – only a perfect match had both. Our study included both a perfect match and a differently coloured item that was both a shape and feature match, so added the opportunity to generalise by both. Our results replicated the finding that feature-matches on a differently shaped base were unlikely to be accepted as category members, however, children generalised more systematically when both shape and feature matched the standard compared to a shape only match. This does not suggest that the local feature was unimportant *per se*, but rather that the feature was secondary to global shape in terms of informing generalisation. If the object could be ruled out because it was a different shape, the feature was no longer considered. However, if it was a global shape match, then children would decide if the feature was also important to category membership.

Surprisingly, autistic children also prioritised global shape over the local feature when making categorisation decisions. While there is growing evidence of autistic children performing as well as TD children in forced-choice generalisation tasks (Field et al., 2016b; Hartley et al., 2019, 2020; Keating et al., Chapter 2, Chapter 3; Tek et al., 2008), this has not been the case in non-competitive shape bias tasks like the current yes/no task (Keating et al., Chapter 2, Chapter

3). In previous yes/no tasks, autistic children were less likely to reject colour- and texture-match distractor items, resulting in a weaker shape preference being observed compared to TD children. The current finding of a shape preference supports our previous study with highly contrasting shapes, in that the shape-match items were accepted with great consistency, so the shape bias was clearly observed above the distractors that were accepted at chance levels. However, the studies differ in observed over-generalisation to the items that were not a shape-match. In Keating et al. (Chapter 2), autistic children were more likely than TD children to generalise labels to items that differed in shape, and that finding was not replicated here. That appears to be accounted for by slightly higher probability of TD children accepting different global shapes in the current study (around 29% in the current study compared to 22% in Keating et al. (Chapter 2)), and a reduction for autistic participants (41% currently down from 53%), effectively closing the gap between the groups. So, while autistic participants in this study did not over-generalise significantly, the findings do not contradict our previous results.

Notably, autistic children did not generalise on the basis of local features. If the item was not a shape-match, autistic children had almost identical odds of accepting both feature-congruent and feature-incongruent test objects. This suggests that they did not treat the local feature as important to category membership. While it is possible that the local feature still caused a distraction for autistic participants, drawing their attention away from shape and increasing the overall difficulty of the task, the results of the online and offline conditions suggest this is unlikely. In offline trials where the standard had to be held in memory, children were less likely to generalise by the local feature, and there was no difference by population. In other words, overall, participants only generalised by both feature and shape if they could see the standard, suggesting that the feature was more weakly-encoded than shape. In the online condition, acceptance rates for global shape matches with incongruent local features were around 50%, so shape bias was disrupted in favour of conserving both characteristics. A possible explanation for this finding is that the complexity of the stimuli and the short presentation time prevented children encoding enough detail about the feature into working memory. For the global shape, however, automatic attention to the shape (Smith et al., 1996), and better memory for that characteristic (Vlach, 2016) could support shape bias in the offline condition despite distraction from the local feature.

These results indicate a hierarchy of importance for the characteristics in this study. Shape was clearly the primary cue, given priority attention which allowed it to be available in the offline condition. Feature was a secondary cue, only considered once a positive shape match had been made. Having identified an object as a shape-match, children could then reject the item as a candidate for having a different feature. However, in the offline condition, the reduced time to process the stimuli may have restricted attention to only the characteristics with the highest priority, so the local feature did not interfere. Of particular importance to the current research is the finding that this order was the same for the autistic participants, suggesting that preference for processing local details does not explain weaker shape bias in autism.

Contemporary interpretations of WCC argue for a 'local first' attentional style that can be overcome to focus on global information (Happé & Frith, 2006). However, our results did not support this pattern as both groups used shape as the primary cue, and both were less influenced by the feature in the offline condition. An alternative model of autistic attention, enhanced perceptual functioning (Mottron et al., 2006) may provide a better explanation. According to this theory there is no local bias, but a general superiority of perception that is flexible to task demands. Children can process local details, but not to the detriment of global meaning.

Presuming that autistic children use past word learning experience to extract the weighting of shape for generalisation, which is likely based on the current and previous findings that shape bias improves with receptive vocabulary (Keating et al., Chapter 2), it would follow that attention to shape would be intact for a task such as this. Children would have no difficulty learning that shape is a useful cue and making higher order generalisation (see Smith et al., 2002). This is supported by our growing body of evidence that autistic children's responses to same-shape items is remarkably similar to those of TD children (Keating et al., Chapter 2, Chapter 3). Additionally, this account predicts that autistic children would have an advantage for processing the local feature, but only if there was some benefit to doing so. In the current study the meaning and importance of the feature are never confirmed to children, so based purely on past learning experiences, attention to shape would be the most accurate label generalisation strategy.

In line with our predictions, we did find an effect of labelling in this research. In the current study, shape-match items were more likely to be accepted in the label condition, which is also the case in previous research that included label and no-label trials (e.g. Landau et al., 1988; Smith et al., 2002; Tek et al., 2008). While the label effect is not uncommon, and may support attention to shape by providing a facilitatory 'binding' effect (Samuelson et al., 2013), there are also examples of labels having no impact (Diesendruck & Bloom, 2003; Field et al., 2016b; Keating et al., Chapter 2). In studies with a label effect (e.g. Landau et al., 1988), the test objects typically change one characteristic, so a shape change trial would involve a small alteration to the shape and otherwise use the same colour, texture, and size. Conversely, in studies where labels made no difference, stimuli were perceptually very different and only matched on one characteristic, leading us to suggest that labels only support the shape bias if the task is sufficiently challenging. With obviously different objects, and with wording in the no-label condition that still implies it is a generalisation task, the demands may be simple enough that there is nothing further to gain from a label. The stimuli in the current task were also very different shapes and so should be easily identified as such, however, children may have additional uncertainty around the status of the feature. The fact that the texture is held constant across the objects, maintaining similarity on another characteristic may increase the challenge enough for labels be useful. Our data supports this interpretation, as no-label global shape-match objects only had a 47% chance of being accepted when the local feature did not match the standard. However, this increased to 67% when the same objects were named, suggesting that labels enhanced attention to shape and inhibited the influence of the feature.

We also found that shape bias was stronger for children with higher receptive vocabulary, as has been the case in previous research (Gershkoff-Stowe & Smith, 2004; Keating et al., Chapter 2, Chapter 3; Samuelson & Smith, 1999). While children were more likely to accept global shape matches as vocabulary increased, vocabulary had no effect on local feature acceptance. This indicates that labels and children's own past word learning experience influence automatic attention to shape, while the status of the feature is considered secondary. It may be that the decision on the feature is a more top-down process, as children make a conscious decision on whether the feature is relevant to category membership.

These findings have implications for word learning interventions for autistic children. It is becoming increasingly clear that autistic children can demonstrate a strong shape bias in a variety of situations, just as TD children do. A growing body of evidence suggests that shape-based generalisation of labels to new exemplars is a shortcut that is available to autistic children as a word learning strategy. This raises the possibility that shape bias could form the basis of an intervention. TD infants have been shown to benefit from shape bias training, resulting in learning words at an increased rate (Smith et al., 2002). Given the strengths of shape bias we have observed, it is possible this could also be effective for autistic children.

4.6.1 Limitations

One thing to note is that there is a great deal of variability between the autistic children. Some children performed extremely well at this task, while others were prone to accept everything. It may be that further investigation of individual differences would reveal a factor that predicts which children were more prone to errors and which made successful exclusion decisions. What this means for the current study is that we are not proposing that autistic children 'guess' trials with different shapes, but that there may be an undiscovered source of uncertainty. Some children may reason that 'shape' is the organising feature of the category and so give answers that are identical to typical development. Others may include everything because they are reluctant to rule them out without more information about the category. Uncovering the source of individual differences should be a topic of future research, as this has the potential to inform language teaching strategies for these children.

4.7 Conclusion

In conclusion, these findings suggest weak central coherence are an unlikely explanation for shape bias differences in autism, as autistic children prioritised global shape over a local feature when generalising novel nouns. Features were only considered as an *additional* cue for category membership for shape-match items, and did not affect children's decisions on objects that were differently-shaped. Furthermore, the feature was not taken into account when generalisation had to be completed from memory, suggesting this was not encoded as robustly in working memory as global shape. We therefore propose that attention to shape supports category inclusion decisions for both autistic and TD children, as both populations generalise to shape-match items to a similar degree. The decision to exclude non-shape-match items from a category, on the other hand, may be supported by a separate process from the learned association between shape and categories that autistic children find more challenging.

Chapter 5: General Discussion

This research began with a simple question: do autistic children exhibit a shape bias? Given the evidence that this can be a powerful word learning tool for TD children (e.g. Gershkoff-Stowe & Smith, 2004; Smith et al., 2002), it was reasonable to ask whether a difference, or deficit, in this heuristic could account for language difficulties in autism (Anderson et al., 2007). However, in reviewing the fledgling branch of research into the shape bias in autism, it quickly became apparent that this question was incorrect. We should not be asking *if* autistic children exhibit a shape bias, but rather *when*. And once we understand the circumstances under which autistic children do, or do not, utilise this bias when generalising novel nouns, we can begin the process of explaining why they find certain circumstances are more challenging than others.

In the five experiments described in this thesis we replicated some key variations in the methods used in previous research to determine whether the task demands accounted for shape bias differences in autism. We used forced-choice and yes/no variations of the novel noun generalisation (NNG) task to explore the effect of competition between stimuli (e.g. Field et al., 2016b) compared to non-competitive sequential decisions (e.g. Hartley & Allen, 2014). We also repeated the tasks with stimuli sets that varied in perceptual similarity to investigate the impact of stimuli used in the research. Within each experiment we also manipulated the visibility of the 'standard', to test the effect of adding a working memory demand, and the use of labels to see if children benefited more strongly from lexical cues.

Our findings demonstrate that the choice of task used to observe shape bias is of critical importance for investigating this word learning strategy in autism. Different tasks that elicit very similar responses when used with TD children are not equivalent when used with autistic children. In the current research, autistic and TD children showed a comparable shape bias when given forced-choice tasks, however, the same stimuli presented in the yes/no variation resulted in considerable disruption for autistic participants in two out of three experiments. Furthermore, when visual distinctions between the stimuli were smaller, such that they could belong to the same superordinate category, autistic children benefited more from being able to see the standard whereas TD children benefited from the presence of labels (Chapter 3). In short, methodological choices that researchers made with little concern when investigating typical development significantly impacted the strength of the shape bias observed in autistic children.

The results of this research indicate that autistic children can exhibit a shape bias appropriate to their receptive vocabulary under a variety of experimental circumstances. We can broadly draw the following conclusions: (1) shape is prioritised over colour and texture in forced-choice tasks (experiments 1 and 3); (2) overall global shape is prioritised over a local feature for generalisation (experiment 5); (3) autistic and TD children are likely to accept candidates that are a shape-match as belonging to the same category as the standard in a yes/no task (experiments 2, 4 and 5). In short, there is nothing to suggest that autistic children do not treat shape as the most important cue to category membership when generalising count nouns. While we did find some group differences for visibility of standard and label manipulations, these differences speak more to how the information available in the task is processed rather than children's underlying categorisation strategy. For instance, a child can believe that shape is the most important feature, but if they cannot remember an object's shape, they cannot accurately act upon that belief. Their response will be mitigated by their processing limitations and biases. For both autistic and TD children, the shape-category link appears to be robust. The key difference is only observed when a task requires children to actively *exclude* candidates from the category. In two out of three yes/no tasks, autistic children accepted colour and texture-match distractors at around chance level, while TD children were more consistent at rejecting them (experiments 2 and 4). In experiment 5, the difference between the groups was not significant, however, the pattern was the same as the other yes/no tasks with autistic children being more likely to accept distractors than TD children were. So, despite evidence that shape is prioritised as the most important feature when generalising new category members, category exclusion decisions to not appear to be similarly organized around shape similarity for autistic children.

This finding concerning category exclusion has interesting implications, not just for autism, but also for category and word learning more generally. There is an abundance of literature and discussion on how children identify what things *are*, but discussion surrounding how children identify that something does not belong to a category is more limited, and usually relative to category inclusion. For instance, Quinn et al. (1993) describe evidence from a novelty preference task that infants form a category of ‘cats’ that includes novel cats and excludes novel dogs. However, novelty preference tasks involve competition between stimuli, so the claim that a novel dog is excluded is only made in comparison to the inclusion of the more familiar cat. In such cases, whether candidates are truly ‘excluded’ from the category as opposed to just being ‘more novel’ than the competitor is questionable. In non-competitive tasks such as the yes/no task, the exclusion decision is made overtly, however, it is tempting to interpret ‘exclusion’ and the ‘absence of inclusion’ as one in the same. For instance, in a dynamic word learning model, Samuelson et al. (2009, 2013) operationalised ‘no’ decisions in the model as a lack of ‘yes’ response, which proved to be a reasonable emulation of children’s responses to the task. However, our finding that such ‘no’ responses for different shapes do not mirror the ‘yes’ response for same shape in autism suggest that this is not the full picture. We need to consider the role of shape as a predictive cue for category inclusion and exclusion decisions separately. Before reflecting onto this new perspective, we will address the existing claims in the research regarding shape bias acquisition in autism.

5.1 Do autistic children acquire a shape bias?

This was an important starting question when we embarked on this series of studies, because a shape bias deficit had the potential to explain particular characteristics of autistic children's language. As it has been demonstrated that many autistic children can acquire sizable vocabularies and use language well (Anderson et al., 2007; Charman et al., 2003; Norrelgen et al., 2015), the suggestion that language delays may be explained by a reduced ability to exploit this attentional shortcut was certainly worth exploring. Tek et al. (2008) proposed that autistic children may have difficulty acquiring the shape bias due to attentional differences (e.g. if their attention is pulled to local details rather than global shape) after finding that children do not exhibit the bias despite having a sufficiently developed vocabulary. This supposition that autistic children do not acquire a shape bias is reiterated in follow-up research reporting similar results. (Potrzeba et al., 2015; Tek & Naigles, 2017). The researchers argue, reasonably, that the ability to acquire such a sizable vocabulary without exploiting the valuable shape bias shortcut suggests that autistic children acquire their language through some alternative mechanism. However, there are two major points to take into consideration with this claim: (1) the method of measuring shape bias; (2) the definition of shape bias.

Firstly, the conclusion that autistic children do not use shape bias when generalising novel nouns was drawn primarily from intermodal preferential looking (IPL) tasks, and a key conclusion of this thesis is that the choice of task can greatly impact how shape bias presents in autistic children. The absence of a shape bias in one task cannot necessarily be generalised to other tasks, even if similar responding is observed in typical development. It is possible that the shape bias can be acquired, but not used consistently across different conditions. Secondly, the definition of 'shape bias' when measured using the IPL task differs from the classic definition. Originally, shape bias was described as the higher weighting of shape over other perceptual features, and while it was described as being stronger in a naming context, a preference for shape without naming was still considered to be a 'shape bias' in Landau et al. (1988). However, Tek et al. (2008) and the subsequent IPL research devised a measure of 'shape bias performance', calculated by subtracting the percentage of time looking at the shape-match in the no name trials from the percentage of time looking in the named trials. Under this definition, a high preference for shape in both conditions would not be classified as a 'shape bias' due to the lack of a naming effect. Using this definition, the claim that autistic children do not acquire a shape bias could still be supported if it is reframed in terms of the label effect (e.g. autistic children may acquire a shape-category association rather than a shape-label-category association, and so have limited benefit from the label condition). However, using Landau et al.'s (1988) original conception of the shape bias, the current research provides convincing evidence that the autistic participants had successfully acquired a shape bias.

While autistic children did not consistently exhibit a shape bias across the five experiments described in this thesis, a significant preference for shape was measured in both forced-choice tasks. This finding is difficult to explain without accepting that a shape bias has been 'acquired' insofar as autistic children gave responses that were consistent with the belief that shape was the most important characteristic for category membership. Acquisition difficulty cannot explain weaker shape bias performance in the yes/no tasks if children are able to use it in forced-choice tasks, therefore we deduce that acquisition is unlikely to be the cause of the observed population differences. Furthermore, our findings suggest that autistic children do have a shape bias that is appropriate to their receptive vocabulary level, even if it is not exhibited in every task. This does not support the suggestion that their vocabulary is acquired by an alternative route that does not involve the shape bias. In cases of language delays, it may be that both shape bias and language acquisition are delayed together, and so language ability could affect the application of a shape bias. However, we have not observed a larger vocabulary with an absent shape bias when the differences between the tasks are taken into consideration.

Based on our findings, I suggest that shape bias is successfully acquired by autistic children. Measurements prioritise the effect of labels on shape bias may risk overlooking the roots of shape bias as an attentional learning strategy. In typical development, it is proposed that shape bias is acquired implicitly, learned from repeated experience of objects that are similar shapes being paired with the same label (Smith et al., 2002, 2010). Generally speaking, implicit learning appears to be intact in autism (Foti et al., 2015). Specific to categorisation, Klinger and Dawson's (2001) well-known research demonstrating that autistic children had difficulties forming 'prototypes' for a new category of novel creature, nevertheless showed they could form categories as well as TD children by both implicit and explicit rule learning. For instance, the rule for a creature being a 'mip' could be 'has big feet', and autistic children were able to learn and apply the rule to include new exemplars in the category, even when they were not told what the rule was.

Ultimately, shape bias acquisition as described in Smith et al. (2002) can be construed as implicit rule learning on a much larger scale. In their 4-step proposal for shape bias development, children initially learn word and object pairs individually (step 1), then notice that those objects that share a name are also the same shape (step 2). This step is arguably similar to Klinger and

Dawson's implicit rule learning condition, with 'shape' being the rule for category membership. Step 3, however, involves taking this rule and making higher order generalisations, applying the 'same shape' rule to completely new categories. This enables step 4, in which children can now successfully generalise newly encountered labels to other same-shaped things from a single example. If this associative learning process can be accomplished purely by implicit learning of the statistical regularities of their existing vocabularies, perhaps we can expect autistic children to be able to do this successfully once they have reached the 50-150 count noun threshold, as TD children do. This is also supported by our findings from experiment 5, that autistic children generalised by global shape and not local feature, as this would be most consistent with their past learning experience. Even though we proposed that autistic children may have more capacity to pay attention to the feature, this process of implicit learning from their existing vocabulary would not lead children to prioritise the feature over global shape. Children should only implicitly learn associations that actually exist between their vocabulary and the real world, and words that are associated with features are comparatively rare. For example, while the category 'lamp' could include anything that has a lightbulb regardless of the global shape, such categories are far less common than those which are organised by shape, so children are unlikely to infer that the feature alone will be 'the rule'.

This leads on to the prediction that other attentional biases that rely on implicit learning of associations between perceptual characteristics, such as texture-bias for deformable things (Samuelson & Smith, 2000a), may also be observed in autism in competitive forced-choice tasks. This opens up an obvious possibility that autistic children may have more difficulty reaching that 50 count noun threshold in the first place, which would delay them having the information required in order to learn the 'rule'. However, this very early word learning process is beyond the scope of this thesis, and occurs prior to any kind of shape bias being demonstrated. For the purposes of this discussion, we are now interested in why autistic children who have apparently successfully developed a shape bias do not consistently use it in NNG tasks as TD children appear to.

5.2 Shape bias supports category inclusion, not exclusion, in autism

If the question is 'when do autistic children exhibit a shape bias', in the current research we can broadly say it is when the task can be completed exclusively by selecting shape-match items. Generalisation appears to become more challenging when exclusion decisions are needed. As stated earlier, we now need to consider the role of shape as a predictive cue for category inclusion *and* exclusion decisions separately. In general, discussion about the shape bias for categorisation usually means shape bias for category inclusion, and does not explicitly consider category exclusion. As we discussed in the chapter on high contrast stimuli, in the commonly used forced-choice task, children are asked to pick one candidate to include in the same category as the standard, without any requirement to exclude the alternatives. Our findings from the two forced-choice tasks in experiment 1 and 3 suggest that, on a behavioural level, shape bias for category inclusion decisions is similar for autistic and TD children when given this competitive forced-choice task. It has been argued that the forced-choice task allows characteristics that have only slight importance to become amplified (Samuelson et al., 2009). So, for the shape-match object to be chosen in a forced-choice task, it need only be considered slightly more important than the other competing features. This also allows for the possibility that children with different strengths of weighting could still come out with similar results. It would not matter if shape was 1% more important or 100% more important, as any increased degree of importance would be enough for the shape-match items to be categorised over the other options. Our findings that autistic children show a typical shape bias in this forced-choice scenario fits with the findings of previous research. Notably, Field et al. (2016b) and Hartley et al. (2019, 2020) also found that autistic children generalised by shape in forced-choice tasks with highly different stimuli. Additionally, Tek et al.'s (2008) found that their autistic participants chose shape-match options at greater than chance levels in a forced-choice pointing task. While the possibility remains that similar responses are reached by different mechanisms, or that labels are not as useful as they are in typical development, the evidence suggests that autistic children are able to generalise successfully by shape in forced-choice tasks.

In a yes/no task, on the other hand, alternative characteristics such as texture may also be accepted – there does not have to be just one 'winner'. We could think of this as a non-competitive NNG task. In our yes/no tasks we found that autistic children include shape-match objects in the category at very similar rates to TD children, as they did in the competitive forced-choice task. The key difference in our non-competitive tasks was that autistic children were more likely to also include distractors in the category (experiment 2 and 4). They were also slightly more likely to do so in experiment 5, however, this was not significantly different to TD children in this case. The findings of experiments 2 and 4 agree with the findings of other non-competitive tasks. In Hartley and Allen's (2014) sorting task, autistic children were also very successful at sorting same-shape items into a box for the new category. However, the children were also more likely to include different-shaped colour-match items. This pattern was replicated by Tovar et al. (2020), who also found autistic children had difficulty excluding novel items with nothing in common with the standard. Across these non-competitive tasks that required children to actively rule out items, autistic children did not exhibit a shape bias.

The majority of recent research into shape bias in autism uses the IPL task (Potrzeba et al., 2015; Tek et al., 2008; Tek & Naigles, 2017), which is not easily defined as a competitive or non-competitive shape bias task. While the task can be thought of as competitive in that there are two visible objects that can be directly compared to each other, rather than each item being independently compared with the standard, there is also no requirement for one of the candidates to 'win'. Equal looking time – a draw – is a valid outcome. We could argue then that responding to this task does not merely require category inclusion, as the forced-choice task does. A reluctance to exclude the alternatives, even if the shape-match is judged to be the best, could appear as equal looking at both. This does fit with the findings that autistic children had roughly equal looking times at both test items when given this task (Potrzeba et al., 2015; Tek et al., 2008). There are other possible explanations for difficulty in the IPL task. For instance, enhanced perception may mean autistic children do not need to ignore the distractor as they have enough capacity to attend to both (Mottron et al., 2006). However, the role of category exclusivity in the IPL task would be worthy of consideration in future research to gain a more complete picture of the conditions under which autistic children exhibit a shape bias.

The finding that autistic children exhibit a shape bias in category inclusion, but not exclusion, raises a fundamental question about the nature of the shape bias. Current explanations appear to conceptualise category 'belonging' and 'not belonging' as binary states of the same thing (Quinn et al., 1993; Samuelson et al., 2009). The object shape is a good predictor of category membership, so the closer the shape is to the standard, the higher the probability that the label can be generalised. If the shape is the same, the probability of belonging is high and the candidate is accepted, whereas different shapes are less likely to belong and so can be excluded.

Figure 1 illustrates how this single inclusion decision could be applied in a yes/no task to achieve shape bias consistent responses.

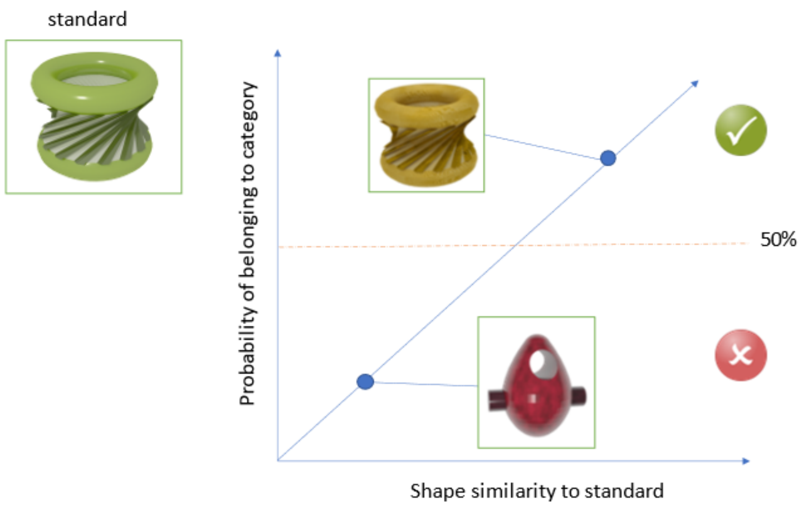


Figure 1: Illustration of how shape similarity could predict acceptance or rejection decisions

This idea of how shape can inform 'yes' or 'no' responses has been more formally represented using dynamic field models (e.g. Samuelson et al., 2009, 2013). Such models consist of 'fields' that represent clusters of neurons that become more active (or more inhibited) in response to an input (see Schöner et al., 2015). These fields have an 'activation threshold' which allows them to have a binary output – if there is enough activity to reach the threshold it can be 'active', or with activity levels that are below the threshold it remains 'inactive'. In the existing dynamic field models of shape bias, these yes/no decisions are also represented by a combined inclusion/exclusion decision. Shape similarity generates activity in the field, and if it is enough to reach the threshold the model gives a 'yes' response. Critically, 'no' responses are represented by a lack of activation. If it is not an active 'yes' then the response is a 'no'. This model is a good fit for behavioural responses of TD children, although our findings with autistic children cannot be accounted for in this way as they did not reliably give 'no' responses to differently shaped items. However, separate inclusion and exclusion decisions may be able to explain both sets of results.

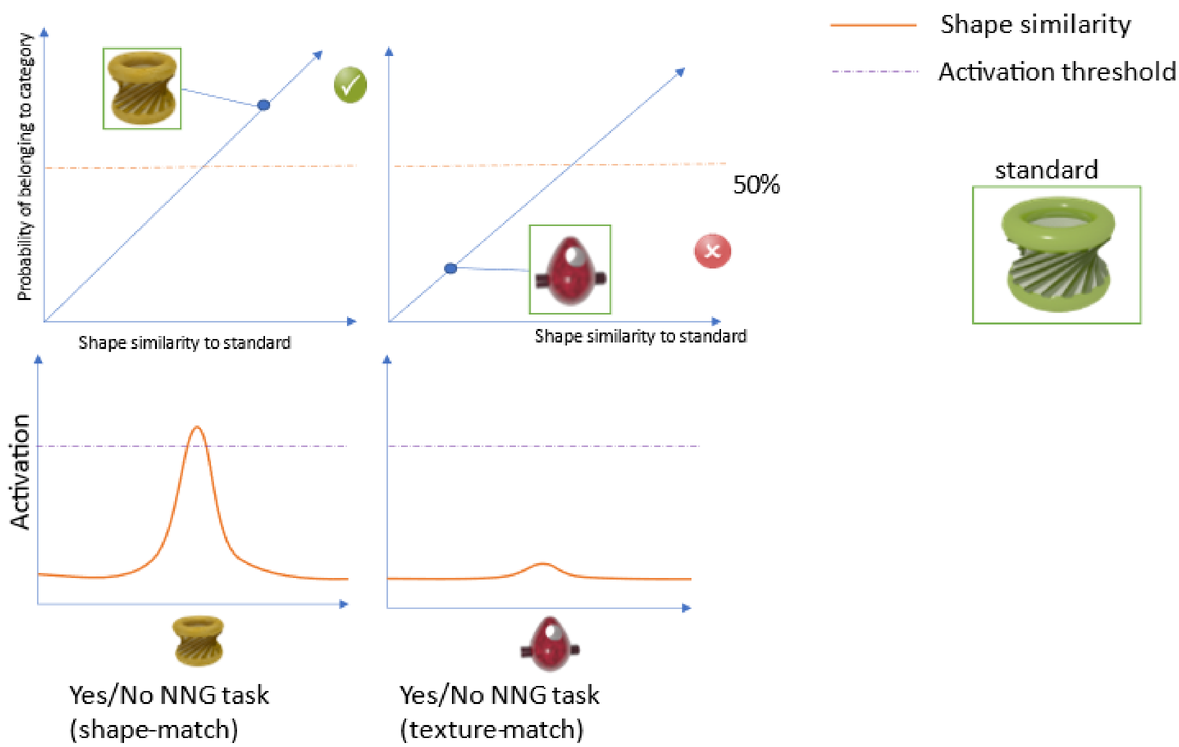


Figure 2: Illustration of how combined inclusion/exclusion decisions can be represented with dynamic fields

With a combined inclusion/exclusion judgement, illustrated in Figure 2, the item that is a shape match for the standard generates a high amount of activity in the field, resulting in an 'accept' decision. The item with a different shape presented to the same field would generate far less activity, the threshold would not be reached, and so the decision would be 'reject'. The same answers could also be reached with a separate exclusion process informed by shape differences (Figure 3).

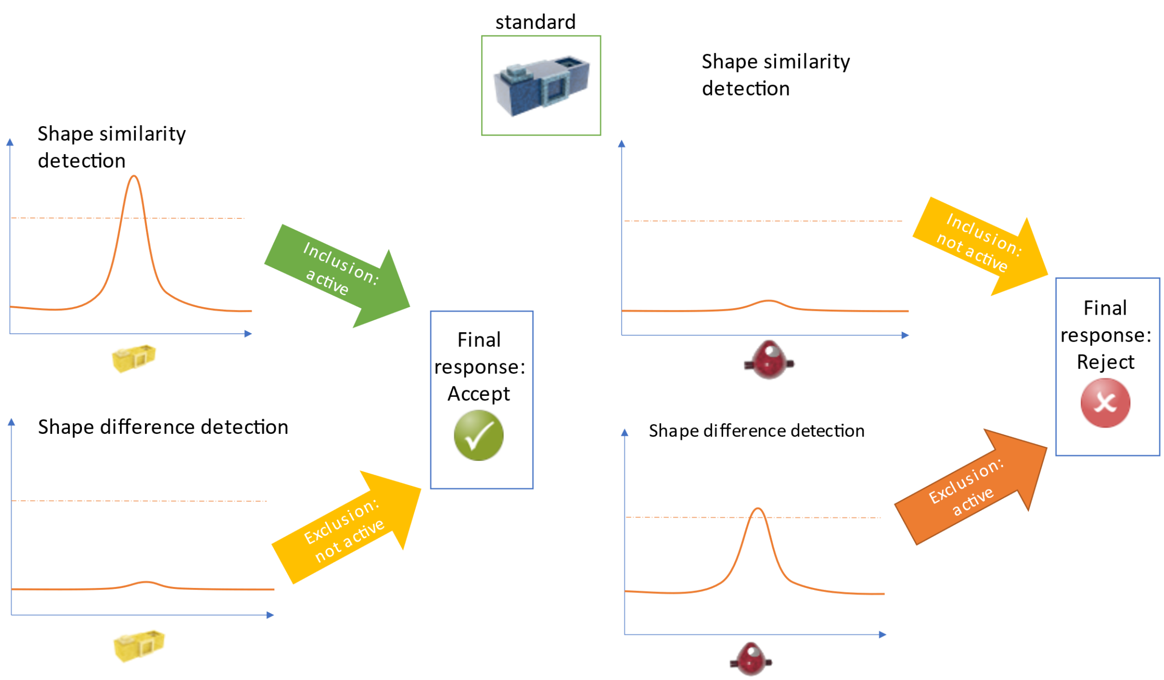


Figure 3: illustration of responses to a yes/no task based on separate inclusion/exclusion decisions

With a separate exclusion process, the 'reject' decision is not merely a failure to 'accept' the item. Exclusion is not defined by the absence of inclusion, but is an active decision based on a how different the shape is from the standard. The appeal of the dual process explanation over the combined one is that this a) it allows for a third 'unsure' state if neither the inclusion nor exclusion decisions become active, which we predict would behaviourally result in a 'best guess' response when encouraged to give one, and b) the results for autistic participants in the current series of experiments could be accounted for by a difference in the exclusion process only. Their inclusion decisions based on shape similarity remain similar to TD children. This would also account for findings that autistic children are prone to over-generalising to any unknown object, even if it has nothing in common with the standard (Tovar et al., 2020). While accepting objects that are the same colour or texture as the standard could be explained by increased attention activating an alternative 'inclusion' decision based on factors other than shape, such explanations struggle to account for this completely novel generalisation where there is nothing to match. As there is evidence that autistic children can exclude items that belong to a known category (Hartley & Allen, 2014; Tovar et al., 2020), the difference may not lie in making exclusion decisions in general, but in ruling out other unknown items when there is no other information about the category. The tendency for autistic children's exclusion decisions to be around 50% in the current task would be consistent with the theory that they are uncertain about the status of those objects. In principle, this could have implications for early word learning while children's vocabularies are sparser and the world is made up of more 'unknown' objects, but may be less of a drawback once a more sizable vocabulary has been obtained.

Instead, it appears any novel object has the capacity to be treated as 'uncertain' and a potential category member. Only the certain knowledge that the thing is something else appears to offer consistent protection against this. So, if implicit attentional learning supports a shape bias for category inclusion, and this appears to be intact for autistic children, we should question how children come to use shape bias for exclusion decisions. This was an unexpected finding for the current research, so our explanations are only suggestions and require further investigation.

5.3 Shape bias as a lexical tool: the 'label effect'

If shape bias is the product of learning associated properties, is it truly a 'word learning' strategy, or does its role in supporting word learning arise as a result of learning categories? In this series of studies, all task variations that elicited a shape bias did so for both labelled and unlabelled items. The only studies in which labels were associated with a stronger shape bias were the forced-choice task with low contrast stimuli (study 3), in which TD children had a slightly increased chance to choose the texture-match distractor on unlabelled trials, and the global vs. local task (study 5) for objects that were a shape-match to the standard only. With these exceptions, labels made no significant difference to generalisation for TD or autistic participants. Overall, labels appeared to enhance attention to shape when the task was more challenging, however there was no evidence that they were necessary for shape-based generalisation. Children prioritised shape regardless, just with greater consistency when items were labelled.

This question about the link between language and shape bias is not a new one. In their original paper, Landau et al. ask if "any such bias would be more likely to originate in general perceptual processes or in the language-learning process itself" (1988, p. 316). To answer this, they suggested shape bias was initially a lexical effect, based on the finding that younger children were sensitive to the presence of a label and demonstrated a stronger shape bias when generalisation took place in a lexical context. Older children and adults, however, did not appear to be influenced by labels, and generalised by shape in variations of the NNG task even when there was no name. They proposed that young children may be more concerned about what words can be used to refer to, and that shape bias becomes more general to non-lexical categorisation tasks with development. Smith et al. (2010) suggest that labels may serve as a sort of 'contextual cueing' (e.g. Chun & Jiang, 1999), with the syntactic position of the novel word in the sentence providing an indication that it was a noun, which in turn would prime attention to shape as the most relevant information. In this way, we could see labels as supporting a more general process of category learning by cuing attention appropriately rather than being the primary subject of the learning. Consequently, language acquisition may be facilitated following shape bias development because children learn categories along with their labels.

The children in our research were slightly older than in Landau et al., with the TD participants in our study aged between 2.5 and 4 years compared to their 2 and 3-year-old participants. Our stimuli were also visually more distinctive from each other. Thus, the lack of a label effect in our research does not necessarily conflict with their findings. If labels play a supportive role, our easier tasks may not have been aided by labels. Our finding that only the most similar stimuli had any labelling effect at all supports this position. This implies that shape bias is not necessarily a word learning strategy, but a categorisation strategy that is made easier when that category is named. Children expect things that look the same to be the same, but that grouping can happen without a name. Categories and their labels are tightly entwined (Markman, 1989). Labels have been described as "invitations to form categories" (Booth & Waxman, 2002a), and may support categorisation by drawing attention to commonalities (Althaus & Plunkett, 2015). Furthermore, labels can override perceptual similarities to form different categories (Plunkett et al., 2008; Westermann & Mareschal, 2014). However, none of these accounts suggest that labels are *required* for the formation of perceptual categories. In fact, there are ample examples of perceptual categories being formed without labels being present (e.g. Plunkett et al., 2008; Quinn et al., 1993). Hence, shape bias may arise from children's early word learning experience, but become a category learning tool that can be utilised without the need for a label to learn. This explanation is more cohesive if we suppose that labels act as features to categories, as proposed in unsupervised feature-based accounts of categorisation (Gliozzi et al., 2009; Plunkett et al., 2008).

There are still questions about the effect of labels in autism. In the current research, we only found a population difference for labelling in experiment 3 (Chapter 3), in the forced-choice task with low contrast stimuli. In the other experiments we found either no label effect (experiments 1, 2 , and 4) or no detectable populations differences (experiment 5). This meant we could only make limited conclusions about whether labels have a different effect for autistic children. Further investigation would require investigation with a task that does induce the label effect in typical development, such as using stimuli with only one dimension of change as in (Landau et al., 1988), and should cover the developmental phase where it is believed to make the most impact (i.e. around 2 years of age). However, our findings do support the claim that shape bias is not exclusively a lexical strategy, and explanations of shape bias in autism should not be limited to label effect. While the label effect is useful for understanding the development of shape bias, and may reveal something about differences in how the shape-category link is learned, the relevance of this effect being absent in older children may be overstated.

Of further interest is that we found no label effect for category exclusion decisions. Experiments 2 and 4 had no effect of label, and in experiment 5 labels only made a difference for items that were already the same shape. They did not increase the chances of rejecting a differently shaped object. This is consistent with explanations that see labels drawing attention to perceptual similarities (e.g. Althaus & Plunkett, 2015), so it is unsurprising that we do not observe a label effect when the perceptual similarities are obvious enough that labels do not add anything. If we are suggesting that the category exclusion process may be different, there was no evidence that labels play a different role in reducing the risk of accepting a distractor.

5.4 Implications for word learning in autism

The findings of this work indicate there is a challenge ahead to disentangle differences in autism that arise from the demands of the experimental task and how that might interact with children's real-world daily word learning situations. We have claimed that autistic children generalise accurately based on shape in a forced-choice task, but are more prone to over-generalisations in a yes/no task. However, neither NNG task represents a natural word learning situation. In a typical interaction between a caregiver and a child, the caregiver might point towards a field and say “Look, can you see a cow?”. It would arguably be uncommon for a parent to point at a cow and ask an infant to confirm or deny if it was one. So, in a common word learning scenario, the competitive attentional process of the forced-choice task would appear to be the most applicable, and this is the task which resulted in shape bias consistent responses for autistic children. The ability to identify the ‘best’ referent for a word from options that are present in the visual field is not challenged by the current findings, and while the realities of processing a busy scene in real life may present additional challenges for autistic children with enhanced perceptual capacity that our neatly presented individual objects on a screen do not, that is a different question beyond the scope of the current work. The implication of this research is that autistic children have more difficulty ruling things out as potential category members, but is the ability essential for natural word learning? If language acquisition can be achieved by attentional learning alone, including learning attentional biases, a sizable vocabulary could be acquired without the need to rule candidates out at all.

One of our motivations for investigating shape bias in autism was to determine if a shape bias training intervention could support word learning. Our findings show that autistic children do prioritise shape, so such an intervention may be of limited use to children of this age who already exhibit a shape bias. However, for TD infants, shape bias training has been shown to support faster language acquisition in 17-month-olds (Smith et al., 2002) and so we might predict such training could have a similar impact for autistic infants. Whether such an intervention would improve language outcomes for autistic children, or support them in reaching their existing potential earlier is difficult to predict, however, it seems that even a reduction of a language delay would be a desirable outcome worth further investigation.

5.5 Implications for demands of the research task

One of the key aims of this research was to address the assumption that tasks that appear equivalent in TD were also equivalent in autism. As expressed by Kucker et al. (2019, p. 162), "…the shape bias is not just about nouns or objects. It is about how the whole system works – memory, attention, object recognition, statistical learning…". A difference in any part of the system has the potential to prevent two tasks from being equivalent. In addition to the differences between the forced-choice and yes/no tasks discussed here, we found that the visibility of the standard, which affects the working memory and attentional demands of the task, affected shape bias for autistic children when the stimuli were similar (study 3). When autistic children had to complete the task from memory, they were more likely to choose the texture-match item, which was not the case for the TD children. While the visibility of the standard did not matter for objects that had more obvious differences, this serves as a reminder that shape bias is not a single entity, but the product of multiple processes and cognitive functions. A shape bias difference may be observed in one task but not in another, so it is essential that future research into shape bias with any neurodiverse population considers whether the choice of task introduces demands that could disproportionately affect one group of participants over another.

5.6 Compatibility with other shape bias accounts

Our proposal, that the ability to use shape-difference to 'rule out' objects as category members is separate from the ability to generalise based on perceptual similarity, differs from other accounts of shape bias. ALA, and the subsequent dynamic systems account, aim to explain how children learn to generalise new words to novel instances. Explaining a lack of generalisation, and why children may exclude candidates from a category, is not the problem these accounts set out to solve. Hence, we see our suggestion of a separate exclusion process as an expansion to these approaches, rather than an alternative, that may account for children's behaviour in a wider range of situations.

Our explanation for our findings is built on the foundations laid by Linda Smith and colleagues in their attentional learning account (Smith et al., 1992, 1996). While we have used the definition of shape bias as the *behaviour* of generalising novel words on the basis of shape (Kucker et al., 2019), ALA was interested in exploring the role that automatic attention played in underpinning this behaviour. By design, their aim was not to fully account for shape bias as a behaviour, but to show how children can exploit regularities, such as the link between count nouns and object shape, to support word learning. Smith and colleagues (2003) proposed that automatic attention is a 'background process' running behind conscious reasoning, thereby acknowledging that other contextual information could influence responses on NNG tasks. Our approach agrees with ALA on these points, and sees ALA as providing a useful account of how children learn where to direct attention for efficient word learning based on their early vocabularies (Smith et al., 2002). While we propose a separation of inclusion and exclusion decisions, which does not feature in their account, we see it as plausible for automatic attention to play a role in both.

If attention is automatically drawn to the most informative cue, this should happen regardless of whether the object is a shape-match for the standard or not. It seems likely that at least some processing of shape would be needed to judge similarity and familiarity. If this check finds that shape does match the standard, it *could* then be included in the category without needing to process additional information about the object. That is not to say that additional processing does not happen. The results of our 5th experiment (Keating et al., Chapter 4) suggested children used additional features when they had time to fully consider them. However, automatic attention provides a way that quick generalisations *can* be made without further processing being necessary if shape is a match. Hence ALA could play a role in more efficient inclusion decisions on the basis of shape.

If shape is not a match, automatic attention to shape would still be useful for helping children to determine that quickly. In our separate process model, we argue that a lack of inclusion, as described above, is not equivalent to an exclusion decision. We propose that excluding an object as a possible category member requires conscious reasoning, with the information on the object's shape forming one source available to guide the final choice. However, that choice is only necessary when children are engaged in a task that requires them to make one. On this point we share the dynamic system’s position that task demands interact with underlying cognitive processes to produce the shape bias behaviour.

The dynamic systems approach to shape bias (Colunga & Smith, 2008; Samuelson & Horst, 2008) is also rooted in ALA, and is already effective for explaining how children's word generalisations are affected by their existing knowledge and the context of the task. As discussed previously (Section 5.2), we see the current implementation of dynamic systems models as best describing same-shape inclusion decisions. However, our position is that existing knowledge and current task demands also interact to inform exclusion decisions, but these may include different or additional knowledge and processes from those used to make quick generalisations. So, we suggest an extension to current dynamic approaches rather than an alternative.

We should also mention compatibility with the 'shape-as-cue' (SAC) account (Bloom, 2000). This sees shape bias as arising from children's belief that object shapes serve as a cue to the *kind* of object, and that children's generalisations are based on deeper properties than perceptual similarities. We have argued that this position is not as incompatible with ALA as previously claimed, and that SAC may be useful for describing more mature categorisation in older children and adults with greater conceptual knowledge. The body of evidence used to support the SAC account comes from only forced-choice tasks, which we have argued can be completed without making exclusion decisions. However, Cimpian and Markman's (2005) forced-choice NNG did include an option to choose 'none of the above'. Notably, this study used known objects and asked children to extend labels in a new 'froggy language', and so children did have existing conceptual knowledge of the items. They found that children choose 'none of the above' more often than expected by chance when there was not another basic-level example of the category to select, despite choosing a shape-match candidate when they were not given the option to reject them all. This finding fits well with our proposal that conceptual knowledge will be utilised to support exclusion decisions, and could be used to override shape-based generalisation when needed. We suggest this additional information would come into play after an automatic exploit had been attempted, allowing children to use both their existing knowledge and the automatically processed information from the task at hand, to decide on what response to give.

5.7 Limitations

As with any research, these studies had a number of limitations. Firstly, the sample size was smaller than originally intended as data collection was interrupted due to the impact of Covid-19 on school closures. Pre-registrations for each of the studies are available on the Open Science Framework (experiments 1 & 3: https://osf.io/sv97u; experiments 2 & 4: https://osf.io/69rxf; experiment 5: https://osf.io/a3z9s). Each of the experiments had 12 to 16 participants in each group. While these are not unusually small sample sizes for developmental research with atypical populations, ideally future research extending this work would build in replication. Additionally, the testing disruption and school closures prevented us from taking all of the planned demographic measures. The intention was to use the Childhood Autism Rating Scale (CARS-2) to confirm children were included in the correct group. We were not able to collect this for the majority of children, and so group membership in the study has been based on the schools' informing us of the children's diagnosis. Expressive vocabulary and non-verbal intelligence measures were collected for about half of the children whose data was analysed. As a result, we were not able to conduct the additional exploratory analysis on these measures that had been planned.

A further limitation of the sample was the verbal ability of the participants included in the data analysis was restricted to a specified range. Children had to meet the minimum score on the BPVS 2nd edition, which meant children who could not engage with this measure were excluded. While some of the autistic participants did have delayed language compared to the TD control group, all of the participants were verbal to some degree. We have suggested that shape bias acquisition requires at least 50 count nouns (Gershkoff-Stowe & Smith, 2004), and so it remains a possibility that non-verbal autistic children, or those with greater language challenges, may not acquire the bias as the children in the current study did. We also set an upper limit on receptive vocabulary, the equivalent of the expected score for a 6.5-year-old, in order for the autistic and TD groups to be matched on this measure. While we did collect data from older autistic children with excellent receptive vocabulary scores, the lack of a TD sample for comparison meant they were not included in the analysis for the studies reported here. As a result, these findings are a limited snapshot of a developmental window, and do not show how the shape bias may be acquired in autism, or how it changes over time. It is possible that the differences found here would not be present in the older children with age-expected receptive vocabulary scores. However, the additional data we have collected could be used for future analysis.

The methods in this research were limited to variations of a novel noun generalisation task, and explicit conscious decisions about category membership. However, as discussed previously, a current prominent line of work in this field uses the IPL method, which we have not touched on here. Future investigation of task demands could also include the IPL method. We have attempted to extract some features of the task for testing, specially the offline element, but the implicit nature of the measure is not considered in the current research at all. It remains a possibility that autistic children's shape bias consistent responses are the result of conscious adjustment from top-down processes rather than automatic attention.

5.8 Future directions

Throughout this thesis, within each chapter and in this discussion we have touched on potential avenues for future research. A great many questions about the role of shape bias for language acquisition in autism remain, and we have only just begun to scratch the surface. The investigation of shape bias training as an intervention for autistic children is one such important avenue. To further investigate the key finding however, a next step would be to explore the possibility that category exclusion is different for autistic children, and uncover why that might be the case. Identifying underlying mechanisms remains a priority, as this is how we can inform the most effective interventions and find ways to present information to children in a way that works with their strengths, and not against them. Having concluded in this thesis that perceptual differences are less likely to be the source of difference, the next line of research could consider other explanations for why the use of shape bias for exclusion decisions appears to be different in autism.

One possibility is that autistic children are less willing to rule out unknown items when they have only a single example. Categories in real life are variable in scope; for instance ‘cats’ are perceptually similar while ‘dogs’ are highly variable, so autistic children may not reject differently shaped items due to uncertainty about the category. How do they know if the example belongs to a narrow or a broad category? From a single example there is no way to be certain – children can only make a best guess based on their previous category learning experience. TD children respond as if they default to a same-shape rule in the absence of any other information, but will form broader categories when they are presented with perceptually different exemplars with the same label (e.g. Plunkett et al., 2008). If uncertainty about the exclusion rule underpins autistic children’s inconsistent answers, manipulating variables that can increase that certainty could support exclusion decisions. For instance, presenting multiple exemplars is one possible way of doing this (Twomey et al., 2014). Given multiple examples that are the same shape, children may be more confident in rejecting differently shaped items, compared to learning a set that allows some shape variability or having a single example.

Another potential line of research could test whether autistic children prefer to rely on mutual exclusivity principles over perceptual cues for limiting generalisation. While autistic children are able to rule out known objects (de Marchena et al., 2011; Hartley & Allen, 2014; Tovar et al., 2020), there is more recent evidence that typically developing children do not need to consciously exclude known items, and may instead associate labels with the most novel object when given a choice (Twomey et al., 2016). This presents two interesting possibilities. 1) Autistic children also generalise to the comparative most novel object on offer, or 2) autistic and TD children use different strategies to make mutual exclusivity-like responses to tasks. If autistic children do use principles of ME to inform their exclusion decisions in an explicit top-down strategy, we could expect their rejection decisions about an item to be more strongly influenced by how similar it is to a known object than by how different it is from the standard.

In addition to extending the study of category exclusion decisions, the huge body of TD shape bias research that already exists offers a rich source of further investigation. Our findings that autistic children exhibit a shape bias differently depending on the task demands means it would be useful to explore previous shape bias manipulations with this population to determine what underlies these differences. For instance, do autistic children learn other associations between categories and informative perceptual features, such as material for deformable objects (Samuelson & Smith, 2000a), or increased texture bias when there are cues to animacy (Jones et al., 1991)? If children learn these associations implicitly, as we have suggested they could, other perceptual word learning biases could be observed. Additionally, replications of research showing TD children's shape bias is influenced by the syntactic context the word is presented in, specifically if the word is presented as a noun or a verb (Landau et al., 1992), and the influence of providing additional conceptual information about the object (Booth et al., 2005) could provide valuable insights. Understanding what sources of information autistic children tend to draw on to make word learning and categorisation inferences has the potential to be an important foundation for designing teaching strategies that present information using children's preferred channels.

5.9 Conclusions

This research has reached some clear conclusions about shape bias in autism, specifically that autistic children do acquire this attentional exploit, and are able to use it to make forced-choice generalisations of novel stimuli to a new category. This challenges earlier claims that autistic children do not use shape as a primary categorisation cue (Potrzeba et al., 2015; Tek et al., 2008). Instead, we have argued that they judge same-shape items to belong to the same category and make generalisations to that effect (Keating et al., Chapter 2, Chapter 3). Furthermore, we found no evidence of weak central coherence interfering with attention to global shape over a salient local feature (Keating et al., Chapter 4). Autistic children were not distracted by uninformative features any more than TD children were, though enhanced perceptual functioning may enable them to process more information about the object. However, in tasks that also require decisions to be made about distractor items, the ability to use shape-difference to exclude non-referent objects is not as consistent in autism as it is in typical development. For some autistic children, the imagined boundary of a new category may be unclear, allowing accurate positive identifications to be made while remaining uncertain about what can be ruled out. While learning associations between shapes and nouns appears to be intact, it is possible that shape as a shared reference point for social communication also requires excluding the unknown. If so, this the next challenge we should look to.

Consolidated bibliography

Abdelaziz, A., Kover, S. T., Wagner, M., & Naigles, L. R. (2018). The Shape Bias in Children With Autism Spectrum Disorder: Potential Sources of Individual Differences. *Journal of Speech, Language, and Hearing Research*, *61*(11), 2685–2702. https://doi.org/10.1044/2018_JSLHR-L-RSAUT-18-0027

Adams, N. C., & Jarrold, C. (2012). Inhibition in autism: Children with autism have difficulty inhibiting irrelevant distractors but not prepotent responses. *Journal of Autism and Developmental Disorders*, *42*(6), 1052–1063. https://doi.org/10.1007/s10803-011-1345-3

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The Role of Discourse Novelty in Early Word Learning. *Child Development*, *67*(2), 635–645. https://doi.org/10.2307/1131837

Allen, M. L., Hartley, C., & Cain, K. (2016). iPads and the use of 'apps' by children with autism spectrum disorder: Do they promote learning? *Frontiers in Psychology*, *7*(AUG), 1–7. https://doi.org/10.3389/fpsyg.2016.01305

Althaus, N., & Plunkett, K. (2015, October). Categorization in infancy: Labeling induces a persisting focus on commonalities. *Developmental Science*, 1–11. https://doi.org/10.1111/desc.12358

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., Welch, K., & Pickles, A. (2007). Patterns of Growth in Verbal Abilities Among Children With Autism Spectrum Disorder. *Journal of Consulting and Clinical Psychology*, *75*(4), 594–604. https://doi.org/10.1037/0022-006X.75.4.594

Audacity Team. (2017). *Audacity(R)* (2.1.3) [Computer software].

Baldwin, D. A. (1991). Infants' Contribution to the Achievement of Joint Reference. *Child Development*, *62*(5), 875–890. https://doi.org/10.2307/1131140

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395–418. https://doi.org/10.1017/S0305000900008345

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' Reliance on a Social Criterion for Establishing Word-Object Relations. *Child Development*, *67*(6), 3135–3153. https://doi.org/10.2307/1131771

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *30*(2), 285–297. https://doi.org/10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S. (2000). Theory of Mind and Autism: A Review. *International Review of Research in Mental Retardation*, *23*(JANUARY 2000), 169–184. http://dx.doi.org/10.1016/S0074-7750(00)80010-5

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beesley, T., Nguyen, K. P., Pearson, D., & Le Pelley, M. E. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *Q J Exp Psychol (Hove)*, *68*(11), 2175–2199. https://doi.org/10.1080/17470218.2015.1009919

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*(2), 105–141. https://doi.org/10.1016/0010-0277(96)00706-8

Blender Development Team. (2017). *Blender—A 3D modelling and rendering package* (2.78) [Computer software]. Blender Foundation. http://www.blender.org

Bloom, P. (2000). *How children learn the meanings of words*. MIT Press.

Bloom, P. (2003). Mindreading, Communication and the Learning of Names for Things. *Mind & Language*, *17*(1-2), 37–54. https://doi.org/10.1111/1468-0017.00188

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science*, *9*(3), 200–204. https://doi.org/10.1111/1467-9280.00038

Booth, A. E., & Waxman, S. (2002a). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948–957. https://doi.org/10.1037/0012-1649.38.6.948

Booth, A. E., & Waxman, S. R. (2002b). Word learning is 'smart': Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, *84*(1), 11–22. https://doi.org/10.1016/S0010-0277(02)00015-X

Booth, A. E., & Waxman, S. R. (2008). Taking stock as theories of word learning take shape. *Dev Sci*, *11*(2), 185–194. https://doi.org/10.1111/j.1467-7687.2007.00664.x

Booth, A. E., Waxman, S. R., & Huang, Y. T. (2005). Conceptual information permeates word learning in infancy. *Dev Psychol*, *41*(3), 491–505. https://doi.org/10.1037/0012-1649.41.3.491

Booth, R., & Happé, F. (2018). Evidence of Reduced Global Processing in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(4). https://doi.org/10.1007/s10803-016-2724-6

Brian, J. A., Tipper, S. P., Weaver, B., & Bryson, S. E. (2003). Inhibitory mechanisms in autism spectrum disorders: Typical selective inhibition of location versus facilitated perceptual processing. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*(4), 552–560. https://doi.org/10.1111/1469-7610.00144

Carey, S. (2010). Beyond Fast Mapping. *Language Learning and Development*, *6*(3), 184–205. https://doi.org/10.1080/15475441.2010.484379

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur communicative development inventory (Infant Form). *Journal of Child Language*, *30*(1), 213–236. https://doi.org/10.1017/S0305000902005482

Chun, M. M., & Jiang, Y. (1999). Top-Down Attentional Guidance Based on Implicit Learning of Visual Covariation. *Psychological Science*, *10*(4), 360–365. https://doi.org/10.1111/1467-9280.00168

Cimpian, A., & Markman, E. M. (2005). The absence of a shape bias in children's word learning. *Developmental Psychology*, *41*(6), 1003–1019. https://doi.org/10.1037/0012-1649.41.6.1003

Colle, L., Baron-Cohen, S., Wheelwright, S., & van der Lely, H. K. J. (2008). Narrative Discourse in Adults with High-Functioning Autism or Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *38*(1), 28–40. https://doi.org/10.1007/s10803-007-0357-5

Colunga, E., & Smith, L. B. (2008). Flexibility and variability: Essential to human cognition and the study of human cognition. *New Ideas in Psychology*, *26*(2), 174–192. https://doi.org/10.1016/j.newideapsych.2007.07.012

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1308–1318. https://doi.org/10.1080/23273798.2018.1471512

Davis-Kean, P. E., & Ellis, A. (2019). An overview of issues in infant and developmental research for the creation of robust and replicable science. *Infant Behavior and Development*, *57*, 101339. https://doi.org/10.1016/j.infbeh.2019.101339

de Marchena, A., Eigsti, I. M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, *119*(1), 96–113. https://doi.org/10.1016/j.cognition.2010.12.011

Diesendruck, G., & Bloom, P. (2003). How Specific Is the Shape Bias? *Child Development*, *74*(1), 168–178. https://doi.org/10.1111/1467-8624.00528

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale (2nd Ed.)*. NFER-Nelson.

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691. https://doi.org/10.1016/j.rasd.2010.09.001

Ellis Weismer, S., Gernsbacher, M. A., Stronach, S., Karasinski, C., Eernisse, E. R., Venker, C. E., & Sindberg, H. (2011). Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *Journal of Autism and Developmental Disorders*, *41*(8), 1065–1075. https://doi.org/10.1007/s10803-010-1134-4

Field, C., Allen, M. L., & Lewis, C. (2016a). Are Children with Autism Spectrum Disorder Initially Attuned to Object Function Rather Than Shape for Word Learning? *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-015-2657-5

Field, C., Allen, M. L., & Lewis, C. (2016b). Attentional Learning Helps Language Acquisition Take Shape for Atypically Developing Children, Not Just Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *46*(10), 3195–3206. https://doi.org/10.1007/s10803-015-2401-1

Foti, F., De Crescenzo, F., Vivanti, G., Menghini, D., & Vicari, S. (2015). Implicit learning in individuals with autism spectrum disorders: A meta-analysis. *Psychological Medicine*, *45*(5), 897–910. https://doi.org/10.1017/S0033291714001950

Fritzley, V. H., & Lee, K. (2003). Do Young Children Always Say Yes to Yes–No Questions? A Metadevelopmental Study of the Affirmation Bias. *Child Development*, *74*(5), 1297–1313. https://doi.org/10.1111/1467-8624.00608

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not Tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, *105*(1), 218–228. https://doi.org/10.1016/j.cognition.2006.09.005

Funabiki, Y., & Shiwa, T. (2018). Weakness of visual working memory in autism. *Autism Research*, *11*(9), 1245–1252. https://doi.org/10.1002/aur.1981

Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, *75*(4), 1098–1114. https://doi.org/10.1111/j.1467-8624.2004.00728.x

Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, *33*(4), 709–738. https://doi.org/10.1111/j.1551-6709.2009.01026.x

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*(1), 171–183.

Graham, S. A., & Diesendruck, G. (2010). Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. *Cognitive Development*, *25*(2), 111–123. https://doi.org/10.1016/j.cogdev.2009.06.002

Hall, D. G., & Graham, S. A. (1999). Lexical Form Class Information Guides Word-to-Object Mapping in Preschoolers. *Child Development*, *70*(1), 78–91.

Happé, F., & Booth, R. (2008). The Power of the Positive: Revisiting Weak Coherence in Autism Spectrum Disorders. *Quarterly Journal of Experimental Psychology*, *61*(1), 50–63. https://doi.org/10.1080/17470210701508731

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Hartley, C., & Allen, M. L. (2014). Brief report: Generalisation of word-picture relations in children with autism and typically developing children. *Journal of Autism and Developmental Disorders*, *44*(8), 2064–2071. https://doi.org/10.1007/s10803-014-2074-1

Hartley, C., Bird, L. A., & Monaghan, P. (2019). Investigating the relationship between fast mapping, retention, and generalisation of words in children with autism spectrum disorder and typical development. *Cognition*, *187*(March), 126–138. https://doi.org/10.1016/j.cognition.2019.03.001

Hartley, C., Bird, L. A., & Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition*, *200*(July 2019), 104265. https://doi.org/10.1016/j.cognition.2020.104265

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Hupp, J. M. (2008). Demonstration of the shape bias without label extension. *Infant Behavior and Development*, *31*(3), 511–517. https://doi.org/10.1016/j.infbeh.2008.04.002

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development*, *9*(1), 45–75. https://doi.org/10.1016/0885-2014(94)90019-1

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*(3), 499–516. https://doi.org/10.2307/1131126

Keates, J., & Graham, S. A. (2008). Category markers or attributes: Why do labels guide infants' inductive inferences? *Psychological Science*, *19*(12), 1287–1293. https://doi.org/10.1111/j.1467-9280.2008.02237.x

Keating, L., Twomey, K. E., & Hartley, C. (Chapter 3). *Shape bias for perceptually similar stimuli during 'online' and 'offline' novel noun generalisation for autistic and typically developing children*.

Keating, L., Twomey, K. E., & Hartley, C. (Chapter 2).

Kim, S. H., Paul, R., Tager-Flusberg, H., & Lord, C. (2014). Language and communication in autism. In *Handbook of autism and pervasive developmental disorders: Diagnosis, development, and brain mechanisms, Volume 1, 4th ed.* (pp. 230–262). John Wiley & Sons, Inc.

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, *13*(2001), 111–124. https://doi.org/10.1017/S0954579401001080

Kucker, S. C., Samuelson, L. K., Perry, L. K., Yoshida, H., Colunga, E., Lorenz, M. G., & Smith, L. B. (2019). Reproducibility and a unifying explanation: Lessons from the shape bias. *Infant Behavior and Development*, *54*(October 2018), 156–165. https://doi.org/10.1016/j.infbeh.2018.09.011

Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, *31*(6), 807–825. https://doi.org/10.1016/0749-596X(92)90040-5

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Landau, B., Smith, L. B., & Jones, S. S. (1998). Object shape, object function, and object name. *Journal of Memory and Language*, *38*(38), 1–27. https://doi.org/10.1006/jmla.1997.2533

Lazenby, D. W. C., Sideridis, G. D., Huntington, N., Prante, M., Dale, P. S., Curtin, S., Henkel, L., Iverson, J. M., Carver, L., Dobkins, K., Akshoomoff, N., Tagavi, D., Nelson, C. A., & Tager-Flusberg, H. (2016). Language Differences at 12 Months in Infants Who Develop Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *46*(3). https://doi.org/10.1007/s10803-015-2632-1

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Lotter, V. (1978). Follow-Up Studies. In M. Rutter & E. Schopler (Eds.), *Autism: A Reappraisal of Concepts and Treatment* (pp. 475–495). Springer US. https://doi.org/10.1007/978-1-4684-0787-7_32

Mandler, J. M. (2000). Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development*, *1*(1), 3–36. https://doi.org/10.1207/S15327647JCD0101N

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. mit Press.

Markson, L., Diesendruck, G., & Bloom, P. (2008). The shape of thought. *Dev Sci*, *11*(2), 204–208. https://doi.org/10.1111/j.1467-7687.2007.00666.x

Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and Developmental Receptive Language Disorder—a Comparative Follow-up in Early Adult Life. I: Cognitive and Language Outcomes. *Journal of Child Psychology and Psychiatry*, *41*(5), 547–559. https://doi.org/10.1111/1469-7610.00642

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research*, *4*(6), 422–437. https://doi.org/10.1002/aur.223

Naigles, L. R., & Tek, S. (2017). 'Form is easy, meaning is hard' revisited: (Re) characterizing the strengths and weaknesses of language in children with autism spectrum disorder. *Wiley Interdiscip Rev Cogn Sci*. https://doi.org/10.1002/wcs.1438

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., Gillberg, C., & Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, *19*(8), 934–943. https://doi.org/10.1177/1362361314556782

Oakes, L. M., & Kovack-Lesh, K. A. (2007). Memory processes and categorization in infancy. *Special Issue: The Development of Categorization*, *11*(4), 661–677.

O'Hanlon, C. G., & Roberson, D. (2007). What constrains children's learning of novel shape terms? *Journal of Experimental Child Psychology*, *97*(2), 138–148. https://doi.org/10.1016/j.jecp.2006.12.002

Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders*, *31*(3), 257–263. https://doi.org/10.1023/A:1010794902139

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perry, L. K., Axelsson, E. L., & Horst, J. S. (2016). Learning What to Remember: Vocabulary Knowledge and Children's Memory for Object Names and Features. *Infant and Child Development*, *25*(4), 247–258. https://doi.org/10.1002/icd.1933

Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, *2*(NOV), 1–12. https://doi.org/10.3389/fpsyg.2011.00345

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of*

Child Psychology and Psychiatry, and Allied Disciplines, *39*(5), 765–775. https://doi.org/10.1111/1469-7610.00375

Plaisted, K., Swettenham, J., & Rees, L. (1999). Children with autism show local precedence in a divided attention task and global preference in a selective attention task. *Journal of Child Psychology and Psychiatry*, *40*(5), 733–742.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681. https://doi.org/10.1016/j.cognition.2007.04.003

Potrzeba, E. R., Fein, D., & Naigles, L. (2015). Investigating the shape bias in typically developing children and children with autism spectrum disorders. *Front Psychol*, *6*, 446. https://doi.org/10.3389/fpsyg.2015.00446

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), 13–23. https://doi.org/10.1016/j.cognition.2005.01.008

Quine, W. V. O. (1960). *Word and object.* Technology Press of the Massachusetts Institute of Technology.

Quinn, P. C., & Eimas, P. D. (1997). A Reexamination of the Perceptual-to-Conceptual Shift in Mental Representations. *Review of General Psychology*, *1*(3), 271–287. https://doi.org/10.1037/1089-2680.1.3.271

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptual similar natural categories by 3 and 4 month old infants. *Perception*, *22*(4), 463–475. https://doi.org/10.1068/p220463

Remington, A., Swettenham, J., Campbell, R., & Coleman, M. (2009). Selective attention and perceptual load in autism spectrum disorder. *Psychological Science*, *20*(11), 1388–1393. https://doi.org/10.1111/j.1467-9280.2009.02454.x

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter International Performance Scale- 3rd Edition*. Stoelting.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalization: Moment-to-moment interactions shape children's naming biases. *Infancy*, *11*(1), 97–110. https://doi.org/10.1207/s15327078in1101_5

Samuelson, L. K., & Horst, J. S. (2008). Confronting complexity: Insights from the details of behavior over multiple timescales. *Dev Sci*, *11*(2), 209–215. https://doi.org/10.1111/j.1467-7687.2007.00667.x

Samuelson, L. K., Horst, J. S., Schutte, A. R., & Dobbertin, B. N. (2008). Rigid thinking about deformables: Do children sometimes overgeneralize the shape bias? *Journal of Child Language*, *35*(3), 559–589. https://doi.org/10.1017/S0305000908008672

Samuelson, L. K., Schutte, A. R., & Horst, J. S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalization. *Cognition*, *110*(3), 322–345. https://doi.org/10.1016/j.cognition.2008.10.017

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*(1), 1–33. https://doi.org/10.1016/S0010-0277(99)00034-7

Samuelson, L. K., & Smith, L. B. (2000a). Children's Attention to Rigid and Deformable Shape in Naming and Non-Naming Tasks. *Child Development*, *71*(6), 1555–1570.

Samuelson, L. K., & Smith, L. B. (2000b). Children's Attention to Rigid and Deformable Shape in Naming and Non-Naming Tasks. *Child Development*, *71*(6), 1555–1570. https://doi.org/10.1111/1467-8624.00248

Samuelson, L. K., Spencer, J. P., & Jenkins, G. W. (2013). A Dynamic Neural Field Model of Word Learning. *Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence*, 1–27. https://doi.org/10.4018/978-1-4666-2973-8.ch001

Schöner, G., Spencer, J., & Group, D. F. T. R. (2015). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press. https://books.google.co.uk/books?id=WiexCgAAQBAJ

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as Process: Contextually Cued Attention and Early Word Learning. *Cognitive Science*, *34*(7), 1287–1314. https://doi.org/10.1111/j.1551-6709.2010.01130.x

Smith, L. B., Jones, S. S., & Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology*, *28*(2), 273–286. https://doi.org/10.1037/0012-1649.28.2.273

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*(2), 143–171. https://doi.org/10.1016/0010-0277(96)00709-3

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name Learning Provides On-the-Job Training for Attention. *Psychological Science*, *13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

Smith, L. B., Jones, S. S., Yoshida, H., & Colunga, E. (2003). Whose DAM account? Attentional learning explains Booth and Waxman. *Cognition*, *87*(3), 209–213. https://doi.org/10.1016/s0010-0277(02)00236-6

Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology*, *42*(6), 1339–1343. https://doi.org/10.1037/0012-1649.42.6.1339

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, *38*(2), 179–211. https://doi.org/10.1016/0010-0277(91)90051-5

Suh, J., Eigsti, I.-M., Canfield, A., Irvine, C., Kelley, E., Naigles, L. R., & Fein, D. (2017). *Language representation and language use in children with optimal outcomes from ASD* (pp. 225–243). Berlin: Walter de Gruyter GmbH. https://doi.org/10.1037/15964-012

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). *Processes of Language Acquisition in Children With Autism: Evidence from Preferential Looking*. *78*(2), 542–557.

Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., & Chadwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and Down syndrome children. *J Autism Dev Disord*, *20*(1), 1–21.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tecoulesco, L., Fein, D., & Naigles, L. R. (2021). What categorical induction variability reveals about typical and atypical development. *Journal of Child Language*, *48*(3), 515–540. https://doi.org/10.1017/S0305000920000392

Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? *Autism Research*, *1*(4), 208–222. https://doi.org/10.1002/aur.38

Tek, S., Jaffery, G., Swensen, L., Fein, D., & Naigles, L. R. (2012). The shape bias is affected by differing similarity among objects. *Cognitive Development*, *27*(1), 28–38. https://doi.org/10.1016/j.cogdev.2011.09.009

Tek, S., & Naigles, L. (2017). The shape bias as a word-learning principle: Lessons from and for autism spectrum disorder. *Translational Issues in Psychological Science*, *3*(1), 94–103. https://doi.org/10.1037/tps0000104

Tovar, Á. E., Rodríguez-Granados, A., & Arias-Trejo, N. (2020). Atypical shape bias and categorization in autism: Evidence from children and computational simulations. *Developmental Science*, *23*(2), 1–14. https://doi.org/10.1111/desc.12885

Twomey, K. E., Morse, A. F., Cangelosi, A., & Horst, J. S. (2016). Children's referent selection and word learning. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *17*(1), 101–127. https://doi.org/10.1075/is.17.1.05two

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's More Like It: Multiple Exemplars Facilitate Word Learning. *Infant and Child Development*, *23*(2), 105–122. https://doi.org/10.1002/icd.1824

Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., & Wagemans, J. (2015). Global processing takes time: A meta-analysis on local–global visual processing in ASD. *Psychological Bulletin*, *141*(3), 549–573. https://doi.org/10.1037/bul0000004

Vlach, H. A. (2016). How we categorize objects is related to how we remember them: The shape bias as a memory bias. *J Exp Child Psychol*, *152*, 12–30. https://doi.org/10.1016/j.jecp.2016.06.013

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *369*(1634), 20120391. https://doi.org/10.1098/rstb.2012.0391

Williams, K. T. (2007). *Expressive Vocabulary Test, second edition*. Pearson.

World Health Organization. (2022). *ICD-11: International classification of diseases (11th revision)*. https://icd.who.int/en

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250. https://doi.org/10.1016/S0010-0277(02)00109-9

Zhang, M., Jiao, J., Hu, X., Yang, P., Huang, Y., Situ, M., Guo, K., Cai, J., & Huang, Y. (2020). Exploring the spatial working memory and visual perception in children with autism spectrum disorder and general population with high autism-like traits. *PLOS ONE*, *15*(7), e0235552. https://doi.org/10.1371/journal.pone.0235552

Appendix A: Random effects model reduction

The binomial mixed effects models built for chapter 2, 3 and 4 all had a random effects structure determined by the method recommended by Barr et al. (2013). This is to start with the maximal random effects structure and to simplify in steps until the random effects model, and subsequent desired fixed effects models, successfully converge. The following tables detail the reduction steps taken for each experiment in order to determine the random effects for each model.

Chapter 2

Table 1: Comparison of random effects models for experiment 1, chapter 2: forced-choice task with high contrast stimuli. Each model is compared with the indicated nested model.

		Model fit				LRT test against nested					Model
Model	By-subject	AIC	BIC	logLik	deviance	nested	df.resid	chisq	*p*		convergence
1	(1 + online.condition * label.condition | participant) + (1 |stim.set)	466.4	516.4	-221.2	442.4		464				Singular fit
2	(1 + online.condition + label.condition | participant) + (1 | stim.set)	464.3	497.6	-224.1	448.3	1	468	5.82	.213		Singular fit
3.1	(1 + online.condition | participant) + (1 | stim.set)	459.8	480.7	-224.9	449.8	2	471	1.56	.668		Successful
3.2	(1 + label.condition | participant) + (1 | stim.set)	466.7	487.6	-228.4	456.7	2	471	8.46	.037	*	Successful

NB. Models 3.1 and 3.2 were both potential random effects candidates following the convergence issues with model 2. As both variables were of theoretical interest the decision was driven by the data. Both models were compared with model 2. The model including labels as a random effect (3.2) showed a significant reduction in goodness of fit compared to the more complex model. The model containing task type (3.1) was no different to model 2 and converged successfully, so was preferred.

Table 2: Comparison of random effects models for experiment 2, chapter 2: yes/no task with high contrast stimuli. Each model is compared with the indicated nested model.

		Model fit				LRT test against nested				Model
Model	By-subject	AIC	BIC	logLik	deviance	nested	df.resid	chisq	*p*	convergence
1	(1 + online.condition * label.condition | participant) + (1 |stim.set)	844.65	898.78	-410.33	820.65		660			Singular fit
2	(1 + online.condition + label.condition | participant) + (1 | stim.set)	837.62	873.7	-410.81	821.62	1	664	0.97	.915	Singular fit
3.1	(1 + online.condition | participant) + (1 | stim.set)	833.07	855.63	-411.54	823.07	2	667	1.45	.693	Successful
3.2	(1 + label.condition | participant) + (1 | stim.set)	833.7	856.25	-411.85	823.7	2	667	2.08	.556	Singular fit
4	(1 | participant) + (1 | stim.set)	829.72	843.26	-411.86	823.72	3.1	669	0.65	.722	Successful

NB. Models 3.1 and 3.2 were both potential random effects candidates following the convergence issues with model 2. As both variables were of theoretical interest the decision was driven by the data. Both models were compared with model 2. The model including labels as a random effect (3.2) was a singular fit so the model containing task type (3.1) was preferred.

Chapter 3

Table 3: Comparison of random effects models for experiment 1, chapter 3: forced-choice task with low contrast stimuli. Each model is compared with the indicated nested model.

		Model fit				LRT test against nested					
Model	By-subject	AIC	BIC	logLik	deviance	nested	df.resid	chisq	*p*		Model convergence
1	(1 + online.condition * label.condition | participant) + (1 |stim.set)	517.0	565.8	-246.5	493.0		416				Singular fit
2	(1 + online.condition + label.condition | participant) + (1 | stim.set)	511.2	543.7	-247.6	495.2	1	420	2.15	.708		Successful at baseline but singular for more complex models
3.1	(1 + online.condition | participant) + (1 | stim.set)	515.8	536.1	-252.9	505.8	2	431	10.59	.014	*	Successful
3.2	(1 + label.condition | participant) + (1 | stim.set)	506.3	526.6	-248.2	496.3	2	423	1.12	.773		Successful at baseline but singular for more complex models
4.1	(1 | participant) + (1 | stim.set)	512.3	524.4	-253.1	506.3	3.2	425	9.96	.007	**	Successful
4.2	(1 + label.condition | participant)	504.4	520.7	-248.2	496.4	3.2	424	0.12	.733		Successful

NB. Models 3.1 and 3.2 were both potential random effects candidates following the convergence issues with model 2. As both variables were of theoretical interest the decision was driven by the data. Both models were compared with model 2. The model including task type as a random effect (3.1) showed a significant reduction in goodness of fit compared to the more complex model. Modelling the fixed effects was attempted with model 3.2, however convergence failed once interaction terms were added. Model 4.2 was then choosen as the random effects structure for the final analysis.

Table 4: Comparison of random effects models for experiment 2, chapter 3: yes/no task with low contrast stimuli. Each model is compared with the indicated nested model.

		Model fit				LRT test against nested					Model
Model	By-subject	AIC	BIC	logLik	deviance	nested	df.resid	chisq	*p*		convergence
1	(1 + online.condition * label.condition | participant) + (1 |stim.set)	845.3	900.2	-410.7	821.3		702				Singular fit
2	(1 + online.condition + label.condition | participant) + (1 | stim.set)	838.8	875.4	-411.4	822.8	1	706	1.4888	.829		Singular fit
3.1	(1 + online.condition | participant) + (1 | stim.set)	833.5	856.3	-411.7	823.5	2	709	0.6562	.884		Successful
3.2	(1 + label.condition | participant) + (1 | stim.set)	839.4	862.2	-414.7	829.4	2	709	6.5529	.088	.	Singular fit
4.1	(1 | participant) + (1 | stim.set)	835.9	849.6	-415	829.9	3.1	711	6.4338	.040	*	Successful
4.2	(1 + online.condition | participant)	831.5	849.8	-411.8	823.5	3.1	710	0.065	.065		Successful

NB. Models 3.1 and 3.2 were both potential random effects candidates following the convergence issues with model 2. As both variables were of theoretical interest the decision was driven by the data. Both models were compared with model 2. The model including labels as a random effect (3.2) was a singular fit so the model containing task type (3.1) was preferred. Model 4.1 and 4.2 were both potential simplifications. Both were compared to model 3.1. 3.1 was a significantly better fit than 4.1, whereas 4.2 was not significantly different. Therefore 4.2 was the preferred simplified model.

Chapter 4

Table 5: Comparison of random effects models for chapter 4: yes/no task with global shape/local feature stimuli. Each model is compared with the indicated nested model.

		Model fit				LRT test against nested				Model
Model	By-subject	AIC	BIC	logLik	deviance	nested	df.resid	chisq	*p*	convergence
1	(1 + online.condition * label.condition | participant) + (1 |stim.set)	1365.5	1425.4	-670.8	1341.5		1073			Singular fit
2	(1 + online.condition + label.condition | participant) + (1 | stim.set)	1360.8	1400.7	-672.4	1344.8	1	1077	3.31	.508	Singular fit
3.1	(1 + online.condition | participant) + (1 | stim.set)	1359.7	1384.7	-674.9	1349.7	2	1080	4.92	.178	Singular fit
3.2	(1 + label.condition | participant) + (1 | stim.set)	1361.2	1386.1	-675.6	1351.2	2	1080	6.38	.095 .	Singular fit
4.1	(1 + online.condition | participant)	1357.7	1377.7	-674.9	1349.7	3.1	1081	0	1	Successful
4.2	(1 + label.condition | participant)	1359.2	1379.2	-675.6	1351.2	3.2	1081	0	1	Successful

NB. Models 3.1 and 3.2 were both potential random effects candidates following the convergence issues with model 2. As both variables were of theoretical interest the decision was driven by the data. Both models were compared with model 2. Both 3.1 and 3.2 were a singular fit. The variance for the stimuli set was near 0 on both, so the stim.set intercept was removed. Models 4.1 and 4.2 were equally good fits. The final fixed-effects structure was tried with both, and again both were an equal fit. Model 4.2 was chosen to capture individual variation in the effect of labels.