Most versus Least: A Meta-Analysis of Stimulus Overselectivity

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Applied Psychology in Behaviour Analysis

at

The University of Waikato

by

Simoné Nel

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2026

Abstract

Stimulus overselectivity (OS) occurs when a person's behaviour is controlled by only a limited subset of available cues. This phenomenon is thought to contribute to learning and social challenges in individuals with autism spectrum disorder (ASD). This thesis presents a multilevel meta-analysis quantifying the magnitude of OS in individuals with ASD compared to typically developing (TD) peers. Analyses of 15 studies, producing 41 unique effects, showed a statistically significant population gap (p = .036). The average level of OS was 35.16% for ASD groups and 20.96% for TD groups. Both groups exhibited a general tendency for OS, but the effect was more pronounced in ASD. Notably, this difference was fragile: when the oldest study (Gersten, 1983) was excluded, the gap was no longer statistically significant (p = .129). A major limitation was that only 15 out of 158 relevant studies provided enough data for quantitative analysis, highlighting a widespread reporting gap. These findings were consistent with stimulus control hierarchies, suggesting that the systematic narrowing of control observed in individuals with ASD may be sensitive to non-linear diagnostic shifts within the evolving diagnostic landscape. For practitioners, the findings emphasise the importance of assessing which cues guide a learner's responses and using targeted strategies to broaden cue control in educational settings. For researchers, these results highlight the need for consistent, standardised reporting of element-level data and variability to advance theoretical understanding of overselectivity across populations.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr Tim Edwards, whose insightful feedback, guidance, and encouragement made this thesis possible. Thank you for all your support and for always making time for my emails and meetings.

I would also like to thank Dr Stef Gomes-Ng for her advice and expertise on this topic; your guidance was instrumental in shaping our research. My sincere thanks also to Billie; for all your assistance in getting our research started.

I am also grateful to my colleagues and managers for providing me the opportunity to balance full-time study with my professional responsibilities. I am especially thankful for the time and support provided that made balancing the demands of work and a master's degree manageable.

To my friends, family, and community: thank you for always believing in me, and for your encouragement and prayers. A most special thanks to my husband for your endless patience and grace, for helping me see things clearly, and for always reminding me who I am. Thank you for speaking life over me, praying for me, and reminding me who is ultimately always in control.

Finally, thank you to everyone who listened, encouraged, distracted, or prayed with and for me during the ups and downs of this journey. I truly could not have done it without you. To the one still on the Throne, greatest thanks always be to You, *Psalm 35:5*.

List of Figures

List of Tables

Table of Contents

Most versus Least: A Meta-Analysis of Stimulus Overselectivity

Definition and Significance of Stimulus Overselectivity

Stimulus overselectivity (OS), often described in behavioural literature as restricted stimulus control, is where a behaviour is governed by only a limited subset of available cues—even when multiple cues are present (Bickel et al., 1986; Lovaas et al., 1971; Ploog, 2010; Reed, 2017). A compound stimulus comprises several parts, such as facial features or individual letters in a word (Dube & McIlvane, 1999; Fields & Spear, 2012). Stimulus control refers to the extent to which a particular aspect of the environment reliably governs a response (Bickel et al., 1984). Overselectivity arises when a learner focuses on only one component of a complex stimulus, ignoring others (Dube et al., 2010). For instance, a learner might respond correctly to "SUE" only when the letter "S" is present but fail to discriminate between "SUE" and "SAM" due to attention being captured solely by the first letter (Birnie-Selwyn & Guerin, 1997; Dube & McIlvane, 1999). This narrowed control can limit generalisation, impair the formation of complex concepts, and reduce learning opportunities (Broomfield et al., 2008; Dube & McIlvane, 1999; Reynolds et al., 1974; Varni et al., 1979). In clinical settings, difficulty integrating complex information can negatively impact language development, social interaction, and independence (Birnie-Selwyn & Guerin, 1997; Cook et al., 1982; Schreibman & Lovaas, 1973).

Overselectivity is not considered an intrinsic diagnostic trait, but a learned pattern in which training leads to behaviour governed by only a narrow set of features (Reed & Gibson, 2005; Reynolds & Reed, 2011a). In behavioural and learning theory, this selective focus is viewed as a primary reason for disruptions in adaptive responding and new skill acquisition. Importantly, it can result in language delays, social impairments, and reduced

imitative abilities (Cook et al., 1982; Koegel & Wilhelm, 1973; Lovaas et al., 1971). For example, if a person cannot attend to both verbal commands and facial expressions, meaningful social interaction may be compromised (Cook et al., 1982; Schreibman & Lovaas, 1973).

Contemporary perspectives place OS within broader learning theory, highlighting how attentional allocation and stimulus competition shape adaptive responding (Dube & Wilkinson, 2014; Ploog, 2010). In practice, even small changes in how information is presented can alter which cues control behaviour. For example, if a certain element is more salient, appears more often, or is reinforced more strongly (e.g., on a 100% versus a 50% reinforcement schedule), it may dominate responding and reduce the influence of other important features (Ploog et al., 2014; Reed, 2019; Reynolds & Reed, 2018). If this pattern continues, individuals may develop rules or categories that are narrower than intended, lowering the likelihood that skills will generalise to new materials or contexts (Grey et al., 2024; Mueller et al., 2000). This has real-world implications. Practitioners may see learners perform accurately during training but fail to respond when a small aspect of the display changes (Dube & Wilkinson, 2014; Grey et al., 2024). Such patterns are consistent with overselectivity and highlight why the construct matters for both theory and practice (Brown et al., 2022; Dube & Wilkinson, 2014). Understanding these features and consequences helps explain how overselectivity became a key concept in autism research.

Early Research and the Autism–Overselectivity Link

Early research established overselectivity as a prominent feature of autism spectrum disorder (ASD). Researchers frequently used the concept to explain the challenges individuals with ASD face in learning and generalising new skills (Koegel & Wilhelm,

1973; Lovaas et al., 1971). Influential studies by Lovaas and colleagues (1971) showed that, after training with multidimensional cues, children with ASD often responded to only one element of a stimulus. The term "stimulus overselectivity" was introduced to describe this limited range of control (Lovaas et al., 1971; McHugh & Reed, 2007). Subsequent research confirmed strong overselective responding in autistic populations and often compared individuals with ASD to typically developing or intellectually disabled groups (Gersten, 1983; Koegel & Wilhelm, 1973; Lovaas & Schreibman, 1971). As a result, overselectivity became widely regarded as especially common among individuals with ASD and was proposed as a key factor in difficulties acquiring and generalising language, social, and academic skills (Lovaas et al., 1971; Dube et al., 2016; Ploog, 2010).

This early research also suggested that overselectivity contributed to language acquisition difficulties and problems forming lasting social relationships (Lovaas et al., 1971; Schreibman & Lovaas, 1973). Narrow stimulus control was thought to interfere with integrating complex instructional and social cues, thereby explaining broader ASD-related learning challenges (Lovaas et al., 1971; Dube et al., 2016; Ploog, 2010). For instance, a learner might focus exclusively on one feature of a social partner—such as their glasses—instead of integrating verbal and facial cues (Lovaas et al., 1971; Ploog, 2010; Rieth et al., 2015). Early studies documented overselective responding across visual, auditory, and tactile tasks, indicating the problem extended beyond a single sensory modality. These findings pointed to a wider processing impairment in ASD (Gersten, 1983; Ploog & Kim, 2007). This body of evidence influenced both theoretical models of autistic cognition and the development of early behavioural interventions (Reed et al., 2013).

Interpreting this historical literature is complicated by major shifts in the diagnosis and conceptualisation of autism over time. Early research focused on narrower diagnostic groups, often with more severe impairments and a high likelihood of long-term institutionalisation (Lovaas et al., 1971; Schover & Newsom, 1976). The contemporary ASD population is far more diverse, with a wider range of abilities and higher average language skills (Rieth et al., 2015). Earlier clinical estimates suggested that more than 70% of individuals with ASD would never acquire functional speech (Lord et al., 2004; Rieth et al., 2015). Today, fewer than 20% of children with ASD are non-speaking (Lord et al., 2004; Rieth et al., 2015). Earlier diagnosis and broader early interventions may have reduced the development of narrowed stimulus control in current populations (Lord & McGee, 2001; Rieth et al., 2015).

As a result, overselectivity was described as a core stimulus control difficulty in ASD and was closely linked to overfocused attention (Ploog, 2010; Rincover et al., 1986; Dube et al., 2016). Early interventions often targeted "attention shaping," aiming to reduce narrow control so individuals with ASD could benefit from teaching methods that require responding to multiple cues (Dube & McIlvane, 1999; Koegel & Schreibman, 1977; Ploog, 2010). These demonstrations established narrowed stimulus control as a central explanation for the unique learning profiles observed in ASD (Lovaas et al., 1971; Ploog, 2010; Schreibman, 1997).

Therefore, it can be suggested that the reported strength of the ASD–overselectivity association is shaped by historical context and sampling methods, rather than reflecting a universal or stable feature of the current, broader population. Interpretation of the ASD literature must account for the transition from the *Diagnostic and Statistical*

Manual of Mental Disorders, Third Edition (DSM-III; American Psychiatric Association, 1980) to the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5; American Psychiatric Association, 2013). This shifted the diagnosis from a narrow category of global impairment to a single spectrum encompassing a wide range of intellectual, linguistic, and adaptive profiles (American Psychiatric Association, 2013; Rieth et al., 2015). For example, individuals who may have been excluded from an autism diagnosis under DSM-III criteria, due to average or above-average cognitive abilities, are now included within the broader ASD spectrum in DSM-5 (American Psychiatric Association, 2013; Rieth et al., 2015). The change has resulted in more heterogeneous research samples. Strong associations found in earlier, more severe samples are often diluted, especially when studies control for developmental level or compare individuals with ASD to clinical groups matched for cognitive functioning (Kelly & Reed, 2021; Rieth et al., 2015). These points suggest that the strong ASD–overselectivity link identified in early studies may partly reflect historical sampling and conceptual context, rather than a stable, disorder-specific phenomenon (Kelly & Reed, 2021; Ploog, 2010). While initial OS research focused on autism, later studies expanded our understanding of overselectivity by examining its occurrence across a wider range of populations.

Prevalence of Overselectivity Across Populations

Subsequent research has shown that overselectivity is not unique to ASD but occurs across a range of typical and atypical groups (Dube & McIlvane, 1999; McHugh & Reed, 2007; Reed & Gibson, 2005). OS has been documented in individuals with profound intellectual disabilities, Down syndrome, Rett syndrome, and traumatic brain injuries (Dube et al., 1999, 2016; Fabio et al., 2021; McHugh & Wood, 2013). In typical development

(TD), overselective responding appears more often in early childhood, especially in preschoolers under 36 months (McHugh & Reed, 2007). These findings suggest that difficulty with multiple simultaneous cues may reflect a general developmental delay or a specific limitation in cognitive capacity, rather than a trait exclusive to ASD (McHugh & Reed, 2007; McHugh & Wood, 2013).

Beyond TD, the presence of overselectivity in other clinical populations supports the idea that restricted control is a dynamic process shaped by task and environmental factors (Fabio et al., 2021; Reed et al., 2013). For instance, individuals with intellectual and learning disabilities frequently show overselective control during multi-element discrimination tasks, which can significantly affect educational progress (Bailey, 1981; Dube & McIlvane, 1999). Recent research using interactive virtual avatars with individuals with Rett syndrome has demonstrated improvements in attention and retrieval, indicating that the degree of overselectivity can be influenced by cue salience and delivery (Fabio et al., 2021). Studies also report increased overselective responding in older adults compared to younger adults, particularly when tasks include distractors—a pattern that matches age-related declines in attention (Kelly et al., 2016; McHugh & Reed, 2007). Together, these findings support the interpretation that overselectivity reflects general limits in cognitive capacity, attentional allocation, or flexibility, rather than an ASD-specific limitation.

Several factors may explain why overselectivity persists in TD adults under certain conditions (Reed & Gibson, 2005; Reynolds & Reed, 2011b). Overselective responding can be readily induced when adults face increased task demands, such as concurrent cognitive or memory loads (Broomfield et al., 2008; Reed & Gibson, 2005; Reynolds & Reed, 2011b). The length of the retention interval—the delay between a sample and the

presentation of choices—also significantly influences the degree of restricted control (Broomfield et al., 2008; Reed & Gibson, 2005). As this interval increases, TD adults respond with greater overselectivity, resembling the deficits seen in individuals with ASD (Reed, 2006; Stromer et al., 1993). Considering OS as a reflection of general limits in cognitive flexibility, rather than an ASD-specific pathology, formed a primary rationale for the population-level comparisons in this thesis (Reed, 2006, 2012). Moreover, the widespread occurrence of overselectivity across groups has driven researchers to investigate its underlying mechanisms.

Explanations for Overselectivity: Attention at Learning Versus Performance at Test

Theoretical accounts currently diverge on whether overselectivity reflects failures of attention during learning or failures of performance at retrieval (Broomfield et al., 2008; Leader et al., 2009; Reed et al., 2009). Attention-deficit accounts argue that overselectivity arises from not observing or encoding all relevant stimulus elements during training (Dube et al., 1999; Lovaas et al., 1971; Reed, 2006). In essence, if a cue is not attended to initially, it cannot guide behaviour later (Dube et al., 1999; Lovaas et al., 1971; Reed, 2006). Eyetracking analyses strongly support this perspective, frequently revealing failures to maintain consistent sensory contact with all relevant stimuli in individuals with neurodevelopmental disabilities (Dube et al., 1999, 2010; Perez et al., 2014). Notably, frame-by-frame analyses show that narrowed stimulus control is often linked to extremely short observing durations, sometimes just 0.19 to 0.26 seconds (Dube et al., 1999, 2010). For example, research involving participants with moderate intellectual disabilities revealed that these brief fixations and observing failures occurred on nearly 90% of error trials

(Dube et al., 1999, 2010). Adequate sensory contact is therefore a prerequisite for effective performance (Dinsmoor, 1985).

In contrast, performance-based accounts argue that multiple elements are encoded during training, but their behavioural expression is suppressed by cue competition at retrieval (Broomfield et al., 2008; Miller & Matzel, 1988). In this view, overselectivity is primarily a retrieval issue, similar to overshadowing in animal conditioning, where a salient stimulus overshadows a less salient one, even when both are paired with reinforcement (Broomfield et al., 2008; Reed, 2011; Reynolds & Reed, 2018). Retrospective revaluation serves as a key test for these theories. After training, researchers may weaken the dominant (overselected) cue through extinction, without directly training the previously underselected element (Broomfield et al., 2008; Reed et al., 2009). If the underselected cue begins to control behaviour after the extinction procedure, this confirms that it was encoded during training but suppressed during retrieval (Broomfield et al., 2008; Reed et al., 2009). Several studies have observed this recovery pattern in higher-functioning individuals, supporting the idea of an over-sensitive comparator mechanism at the retrieval stage (Gomes-Ng et al., 2023; Leader et al., 2009; Reed et al., 2009).

Synthesising these debates, contemporary research suggests both processes can contribute to overselectivity (Kelly et al., 2015; Reed et al., 2009). OS is multifaceted, with mechanisms that vary depending on age, ability, and task demands (Kelly et al., 2015; Reed, 2011). Performance-based accounts are supported by recovery patterns in higher-functioning samples (Leader et al., 2009; Reed et al., 2009). Whereas the persistent failure of underselected cues to re-emerge in lower-functioning or neurologically compromised

samples is more consistent with encoding limitations (McHugh & Reed, 2007; Reynolds et al., 2012). To reconcile these findings, contemporary research has proposed the prioritisation deficit hypothesis (Etzel et al., 1981; Ploog et al., 2014). According to this account, individuals with ASD may not consistently attend to the most relevant features in their environment, reflecting an atypical comparator process rather than a complete failure of perception (Etzel et al., 1981; Ploog et al., 2014). This recognition of multiple mechanisms sets the stage for more nuanced theoretical models, such as the stimulus control hierarchy, discussed below.

Stimulus Control Hierarchies (the Bickel Nuance)

A more nuanced theoretical perspective, developed by Bickel (1986) and Bickel et al. (1984), frames overselectivity as a stimulus control hierarchy rather than a simple failure of learning. In this framework, S+ refers to stimulus elements linked with reinforcement, while S− indicates those associated with non-reinforcement (Bickel, 1986; Perez et al., 2014). Microanalyses revealed that overselective responding often emerges from behaviour governed by various stimulus–response relations, including context, stimulus position, and even S− elements (Bickel, 1986; Bickel et al., 1984). This finding is significant because it suggests that overselectivity does not reflect a total lack of control but rather shows the specific hierarchy an individual has acquired through environmental shaping (Bickel et al., 1984).

Because responses are often conditional on which elements are presented together, traditional averaging across participants or sessions can hide the actual pattern of stimulus control (Bickel, 1986; Sidman, 1980). For example, a learner might respond at chance (50%) to an element in one context, but at 100% in another, revealing a consistent hierarchy

that disappears when scores are aggregated (Bickel et al., 1984, 1986; Sidman, 1980). By focusing on the difference between the most-selected element and the least-selected element in this thesis, I sought to preserve the individual signal of restricted control and account for variation in which cues dominate behaviour (Bickel et al., 1984; Mason et al., 2021). Poor acquisition, in these contexts, likely reflects a failure to discriminate low-preference dimensions in specific situations, rather than a general deficit in learning (Schneider & Salzberg, 1982). As these theoretical perspectives evolve, it is crucial to consider how overselectivity is measured and the implications of different methodological approaches.

Measurement Inconsistency and Problems of Inference

A major challenge for the field lies in how overselectivity is measured and reported. Although substantial literature exists, conclusions about the nature and prevalence of OS are limited by inconsistent measurement. Studies have used a range of metrics, including the difference between the most-selected and least-selected elements, overall accuracy, and categorical indices (Dube & McIlvane, 1999; Reed & Gibson, 2005; Reynolds & Reed, 2011a). However, these do not always measure the same underlying phenomenon. For instance, low accuracy might reflect general task difficulty or poor motivation, not necessarily restricted control (Mason et al., 2021). Similarly, categorical thresholds for "overselective" or "non-overselective" can hide important within-participant variability (Mason et al., 2021; Reynolds & Reed, 2011a). In comparison, high compound accuracy may mask reliance on a single cue if alternative elements are rarely tested in isolation

(Dickson et al., 2006; Dube & McIlvane, 1999). For example, a learner might always select "A" from a compound "AB" when tested against a neutral "C," but fail to select "B" when it is presented alone, revealing restricted control that compound accuracy would obscure (Perez et al., 2014). In this context, accuracy refers to the proportion of trials where the S+ (the correct cue) is selected over the neutral comparison (Stromer et al., 1993). Thus, someone may appear highly accurate on compound trials because one cue dominates, but their restricted control is revealed only when individual elements are tested separately (Stromer et al., 1993; Perez et al., 2014). Some indices may even confuse overselectivity with random errors or broader cognitive limitations, especially if they do not check for whether errors consistently favour particular cues (Dube & Wilkinson, 2014; Dube et al., 2016). Recent recommendations stress the use of element-level, statistically robust indices and continuous measurement approaches that directly assess the distribution of control across stimulus components, such as reporting the proportion of correct responses for each element (Dube et al., 2016; Mason et al., 2021; Reynolds & Reed, 2011b). For these reasons, this thesis prioritised studies with element-level performance and variance, allowing more precise estimation of restricted control and reducing confounds from task difficulty or random error (Dube & Wilkinson, 2014; Reynolds & Reed, 2011b). These measurement issues underscore the need for a thorough, quantitative synthesis to clarify the actual nature and magnitude of overselectivity across populations.

Rationale for a Population-Focused Quantitative Synthesis

Given these conceptual and methodological issues, a population-focused meta-analysis is needed to determine whether overselectivity is indeed more pronounced in the

ASD population. Narrative and scoping reviews have mapped the breadth of the literature and identified overselectivity across ASD, TD, and other clinical populations. However, these reviews cannot precisely estimate group differences or formally test whether autism is associated with a distinct magnitude of restricted control after accounting for developmental level and task parameters (Kelly et al., 2015; Ploog, 2010). Narrative syntheses are especially limited when primary studies use different metrics and small sample differences (Ploog, 2010; Reed, 2017). Under these conditions, impressions of overselectivity in ASD may reflect outlier effects, publication bias, or historical sampling rather than robust population-level differences (Ploog, 2010; Reed, 2017).

In contrast, a quantitative synthesis using diagnostic groups as the main analytic dimension can aggregate effect sizes and estimate the average magnitude of overselectivity in ASD, as compared to TD and other clinical groups (Kelly et al., 2015; Reed, 2017). This approach enables a more robust determination of whether individuals with ASD consistently show a clinical population gap, or whether overselectivity is better understood as a general, capacity-limited outcome distributed across groups. To ensure methodological consistency in this thesis, eligible studies were screened for their testing arrangements, including simultaneous and delayed matching-to-sample formats, as these influence memory demands during assessment (Reynolds & Reed, 2011; Stromer et al., 1993). All analyses used baseline or pre-intervention data to avoid contamination by interventions designed to reduce overselectivity (e.g., differential observing responses or expanded stimulus sets; Dube & McIlvane, 1999; Reynolds & Reed, 2011b). By prioritising studies with extractable element-level performance and variance, this analysis aimed to provide more precise estimates of restricted control and reduce confounding from overall task

difficulty or random error (Dube & Wilkinson, 2014; Reynolds & Reed, 2011b). This thesis, therefore, lays a rigorous quantitative foundation for evaluating both the historical and current clinical relevance of overselectivity.

Scope and Analytic Boundaries

To ensure coherence, this meta-analysis focused on studies that used comparable discrimination paradigms with sufficient variance estimates for calculating effect sizes. Only studies with human participants and baseline or pre-intervention data were included, to prevent intervention strategies from influencing the results (such as differential observing responses, expanded stimulus sets, or other instructional manipulations; Dube & McIlvane, 1999; Reynolds & Reed, 2011b). The analysis targeted compound discrimination procedures (e.g., simultaneous and delayed matching-to-sample tasks with multi-element visual stimuli), because these allow direct measurement of element-level performance (Reynolds & Reed, 2011b; Reed et al., 2012). By prioritising tasks that reported accuracy or selection frequencies for individual elements, rather than just aggregate scores, robust standardised mean differences and sampling variances could be computed.

Notably, the broader overselectivity literature also includes studies embedded within complex intervention packages or using paradigms that do not yield separable element-level data (e.g., certain free-operant procedures or neurophysiological designs; Reed et al., 2012). These were excluded from the main quantitative synthesis to ensure effect sizes were comparable. This decision balanced the inclusion of diverse studies with the need for rigorous and comparable outcomes (Reed et al., 2012; Reynolds & Reed, 2011b). All exclusion reasons were systematically documented, with records kept on task parameters,

population demographics, and outcome definitions (Page et al., 2021; Reynolds & Reed, 2011b). This approach grounded the research in a replicable evaluation of the measurement landscape and maintained a tight focus on the central population comparison. Because the nature of overselectivity may change across the lifespan, participant age was examined as a key secondary moderator.

Age as a Secondary Moderator

Although population differences were the primary focus of this analysis, participant age was also examined as a key secondary factor. This decision was based on evidence that overselectivity follows a non-linear trajectory across the lifespan (Kelly & Reed, 2021; McHugh & Reed, 2007; Reed et al., 2013). In early childhood, especially under 36 months, typically developing children often struggle with tasks that require discrimination between multiple simultaneous cues (Kelly & Reed, 2020; Reed et al., 2013). This finding suggests that a certain degree of narrowed stimulus control is developmentally typical in very young children, rather than being unique to autism (Kelly & Reed, 2020; Reed et al., 2013). Overselectivity generally decreases as children move through the preschool and early school years, with this reduction linked to gains in cognitive flexibility, selective attention, and receptive language (Kelly & Reed, 2021; Wilhelm & Lovaas, 1976). Among individuals with ASD, similar age-related trajectories have been found. Younger children—particularly those with lower verbal abilities—are more likely to display pronounced overselectivity than their older peers (Kelly et al., 2016; Kelly & Reed, 2021; Rieth et al., 2015).

In addition, research indicates that overselectivity often re-emerges later in life, especially among older adults (Kelly et al., 2016; McHugh & Reed, 2007). Studies with ageing populations have found increased overselective responding compared to younger and middle-aged groups, particularly when tasks involve distractors or greater cognitive load (Kelly et al., 2016; McHugh & Reed, 2007). This pattern is consistent with age-related declines in attentional capacity, processing speed, and cognitive flexibility (Kelly et al., 2016; McHugh & Reed, 2007). Altogether, these findings suggest a U-shaped developmental trajectory: narrowed stimulus control is most prominent in very young children and older adults, with broader control emerging in mid-childhood and early adulthood (Kelly & Reed, 2021; Reed et al., 2013). This developmental pattern has important implications for interpreting group differences, as the age structure of samples can exaggerate or obscure true population effects (Kelly & Reed, 2021; Schover & Newsom, 1976). For this reason, the present meta-analysis included participant age as a planned moderator whenever the data permitted. Understanding these developmental effects is essential for interpreting the main findings and for refining the study's research questions.

Research Questions

Building on the theoretical and methodological context outlined above, this thesis addressed the following research questions:

Primary Research Question

1. What is the average magnitude of stimulus overselectivity (defined as the percentage difference between the most-selected and least-selected elements at test) when

comparing individuals with ASD to TD populations on comparable compound discrimination tasks that allow for element-level analysis?

Additional Research Questions

2. To what extent does participant age moderate the magnitude of group differences in overselectivity between ASD and TD populations?

3. Does the publication year of a study moderate the effect size for overselectivity, reflecting the impact of historical changes in diagnostic criteria and sampling practices?

These questions were designed to clarify whether overselectivity represents a distinct and robust characteristic of ASD relative to TD, or whether it reflects a more general, capacity-limited outcome that varies across the lifespan and historical periods. In doing so, the thesis aimed to advance both theoretical understanding and clinical assessment of stimulus control across populations.

Method

Design Review

This meta-analysis investigated whether individuals with ASD demonstrate a significantly higher degree of overselectivity than TD peers on comparable compound discrimination tasks. The primary outcome was the within-group most-least difference (i.e., the percentage-point difference between the element that most strongly controlled responding and the element that controlled it least) at test:

$$Diff = \%Most - \%Least.$$

By anchoring the measure to each participant's distribution of control, rather than a fixed stimulus label (e.g., A vs. B), this approach preserved the signal of restricted control and accounted for individual differences in cue preference.

Protocol and Reporting Standard

The analysis was conducted in accordance with the PRISMA 2020 statement and its explanatory guidelines (Page et al., 2020, 2021). The review protocol, including specific search strings, screening procedures and the pre-specified analysis plan, was developed prior to data extraction to mitigate selection bias. A PRISMA flow diagram (Figure 1) is provided in the Results section to detail the flow of information through the different phases of the analysis.

Inclusion and Exclusion Criteria

Studies were eligible for the analysis if they met the following five criteria:

1. Reported original empirical research in a peer-reviewed journal.
2. Used human participants.
3. Trained a differential response (e.g., matching or selection) to compound stimuli in a discrimination-learning procedure.
4. Administered element-level tests suitable for assessing control by each component by either (a) presenting single elements against neutral comparisons or (b) pitting trained elements directly against one another.

5. Reported a behavioural measure (percentage correct or choice proportion) with variance (or a route to derive it) to compute the Most–Least difference and its sampling variance.

The meta-analysis was restricted to baseline or pre-intervention phases to avoid contamination by remediation strategies.

Studies were excluded from the analysis if they met any of the following conditions:

- Non-operant designs: Pavlovian blocking or overshadowing studies that did not match the operant learning definition of overselectivity used in this thesis.
- Stimulus equivalence papers: Research that only utilised recombined compounds without isolating element-level control via single-element probes.
- Publication type: Dissertations and theses were excluded to maintain consistency in peer-reviewed data quality.
- Incomplete data: Papers that lacked the element-level data or variance required for meta-analytic modelling, and where statistics could not be recovered after author contact or digitisation.
- Incompatible paradigms: Single-stimulus "go/no-go" training arrangements or other designs that did not align with the planned Most–Least contrast.

Study Selection and Screening

Two screeners independently completed all screening phases and used Zotero folders to document agreements and disagreements. Disagreements were resolved by

discussion and cross-referencing against the prespecified eligibility criteria using a shared screening spreadsheet.

Information Sources

Systematic searches were conducted across six electronic databases: Web of Science, Cochrane, EBSCOhost, PsycINFO, PubMed (MEDLINE), and Scopus. Hand searches were conducted across the following journals: *Journal of Applied Behavior Analysis*, *Journal of the Experimental Analysis of Behavior*, *Behavioral Interventions*, *Behavior Analysis: Research & Practice*, *Research in Developmental Disabilities*, and *European Journal of Behaviour Analysis*. To identify relevant research that may not have used contemporary terminology in titles or abstracts (e.g., "overselectivity"), supplementary ancestor (reference list) and descendant (citation tracking) searches were also conducted. All records were managed in a shared Zotero group library to ensure synchronisation of deduplication, screening, and disagreement flags between the two screeners.

Search Strategy and Yield

The following base string was used and adapted for each database:

"overselectivity" OR "selective cue use" OR "restricted cue use" OR "selective stimulus control" OR "restricted stimulus control" OR "faulty stimulus control"

Database searches yielded 868 records. After removing 392 duplicates, 476 unique records were screened by title and abstract. This phase identified 208 initial inclusions with a reviewer agreement rate of 81.3%. Following rescreening and discussion, 225 records progressed to full-text review. Full-text assessment yielded 133 articles, with an agreement

rate of 72.9%. Supplemental citation searching identified an additional 25 articles, bringing the total to 158 relevant articles for narrative context. The PRISMA flow diagram (Figure 1) in the Results section details these counts and the final subset of 15 studies (k = 15) eligible for the quantitative synthesis.

Data Items and Coding

For each eligible group (e.g., ASD vs. TD within the same paper), we extracted the following data items:

- Study identifiers: Authors, year, journal.
- Population: Diagnostic status (ASD, typically developing, or 'other' clinical groups)
- Age: Group age means, or the midpoint of a provided age range.
- Task Parameters: Task class (simultaneous discrimination or matching-to-sample) and specific testing arrangement.
- Outcome Data: Element level percentage for the most selected element and for the least selected elements during testing, or a directly reported most–least difference.
- Variance: (1) Mean element-level performance, (2) the standard deviation (*SD*) for each element, and (3) the standard deviation of the within-participant difference (*SDdiff*). Standard error (*SE*) and confidence interval (*CI*) values were converted to *SD* using the group *n*.
- Methodological Notes: Data source (e.g., whether data was digitised from figures or derived from the formula mentioned below) and whether authors supplied additional

statistics.

Variance Extraction and Handling of Missing Data

As the most–least contrast was a paired-sample comparison, the relevant denominator was the variance of the paired difference. We utilised a hierarchical approach to ensure the most precise available estimates were included in the modelling. We prioritised (a) directly reported *SDdiff*, followed by (b) the back-calculation of *SDdiff* from paired *t*-values and mean differences. Where studies provided only element-level *SDs*, we (c) derived *SDdiff* using the formula:

$$SD_{diff} = \sqrt{SD_{Most}^2 + SD_{Least}^2 - 2rSD_{Most}SD_{Least}}$$

If a within-participant correlation (*r*) was not reported, we imputed $r = .50$ as a plausible mid-range value for repeated measures. Prespecified sensitivity analyses reestimated the primary model at $r = .30$ and $r = .60$ to ensure robustness.

For other missing statistics, we converted standard errors (*SE*) or confidence intervals (*CI*s) to *SD*s using the group *n*. We contacted study authors to request original statistics when necessary. When only element-level means and error bars appeared in the figures, we recovered the data using the WebPlotDigitizer and PlotDigitizer software. We calibrated axes, documented error bar types, and cross-checked extracted values against tabled statistics where possible. All conversions and digitisations were recorded alongside each effect size in the primary database.

Data Preparation for Modelling

Because percentage outcomes are bounded between 0 and 100, it can lead to non-constant variance and potential bias. To address this, the most–least difference was transformed and modelled on the logit scale. Sampling variances on the logit scale were obtained via the delta method (Borenstein et al., 2009). One study reported results on a 0–7 scale; this data was converted to percentages (multiplied by 100/7) to ensure a common natural unit across all outcomes. After analysis, fitted means and 95% confidence intervals (*CIs*) were back-transformed to the percentage scale for interpretation.

Statistical Synthesis

A multilevel random-effects meta-analysis with inverse-variance weights was used. Because some articles contained more than one eligible group, the model was fitted with random intercepts for both study and effect to account for the hierarchical dependence of the data. Heterogeneity was summarised with τ^2 and an I^2 proxy to quantify between-study dispersion. Leave-one-study-out diagnostics and forest-style plots were also examined to evaluate the influence of individual data points.

Primary Moderator and Inference

Population was the primary moderator (ASD vs. TD), and a third group comprising 'other' populations was included in exploratory models where data allowed. An age-adjusted model was fitted using centred chronological age as a covariate. Because the number of independent study clusters was small ($k = 15$), CR2 small-sample robust variance estimation (RVE) was used as the primary inferential approach for fixed effects.

This approach adjusted standard errors and used Satterthwaite degrees of freedom to reduce small-sample bias (Tipton, 2015).

Interval Construction for the ASD − TD Difference

To present the ASD–TD difference on the percentage scale, the contrast was computed on the logit scale and simulated from the CR2 robust covariance of the fixed effects. This generated a sampling distribution aligned with the primary inference (Tipton, 2015). The 95% confidence interval (*CI*) was derived from the 2.5th and 97.5th percentiles of the back-transformed simulated values.

Primary Analysis Dataset

The primary age-adjusted ASD–TD model included 41 distinct effects from 15 studies. For this, an "effect" was defined as a within-group most–least difference from a unique sample, task, or condition; several studies contributed more than one effect under this definition. This dataset was a prespecified subset of the larger meta-eligible pool, restricted to those effects that provided both the ASD–TD contrast and usable chronological age information for the covariate. Effects that did not meet these criteria were excluded from the primary inferential model.

Sensitivity and Robustness Checks

Five prespecified sensitivity checks were conducted to assess the stability of the primary findings:

1. The primary model used an imputed within-participant correlation of $r = .50$. The model was re-estimated at $r = .30$ and $r = .60$, and on the subset of effects with directly available *SD*'s, to confirm robustness.
2. The model was refitted after excluding the 24 effects recovered via digitisation software to evaluate the influence of the extraction methods used.
3. Base population contrasts were compared across unadjusted and age-adjusted models to assess whether age and developmental stages had an impact on the findings.
4. The study era was explored as a supplementary covariate and via a Population × Year interaction to evaluate the effect of historical diagnostic shifts.
5. Where data allowed, supplementary models located other clinical populations (e.g., Down syndrome, Rett syndrome) relative to the ASD and TD fitted means.

Risk of Bias and Study Limitations

Risk of bias considerations were tailored to the specific parameters of laboratory-based discrimination tasks. Each study was coded for features that could distort population estimates of restricted stimulus control. These features included the use of single-element probes during testing to isolate control by individual components, the availability of element-level performance data (rather than aggregate accuracy scores), the reporting of variance statistics suitable for meta-analytic modelling, and the clear separation of baseline performance from post-intervention data.

These features were used to guide the interpretation of the results and the implementation of prespecified sensitivity analyses, such as excluding digitised effects. A single numeric rating for study quality was not computed, in accordance with PRISMA

2020 recommendations for avoiding arbitrary weights in meta-analysis (Page et al., 2020, 2021).

Small-Study Effects

Where the number of available study clusters allowed ($k \geq 10$), small-study effects were screened with an Egger-type regression test and a visual inspection of a funnel plot. In this context, "small-study effects" refers to the potential tendency for smaller studies with lower statistical power to show larger population estimates. Any observed asymmetry was interpreted cautiously, given the dataset's dispersion and hierarchical structure.

Ethical Considerations

This analysis used previously published empirical data and de-identified files shared by study authors. No new data collection from human participants occurred, so ethics approval was not required. Any original data files shared by authors were stored securely and used only for the stated analytic purpose of this thesis.

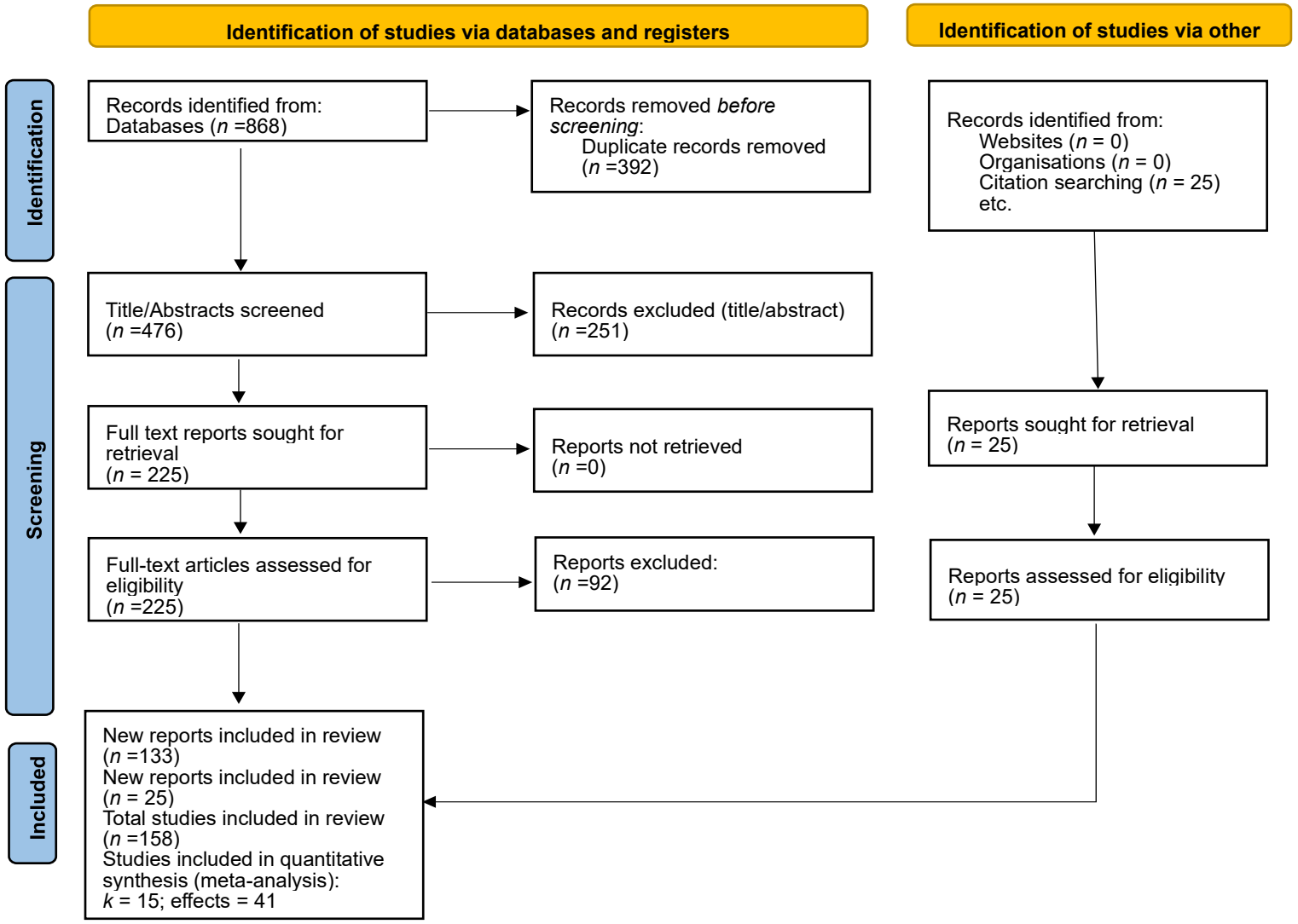
Results

Study Identification and Scope

In total, 868 records were identified through database and hand searches. After removing duplicates, 476 unique records were screened by title and abstract, and 225 full-text articles were assessed for eligibility. Of these, 158 were included for narrative context, and 15 studies (k = 15) with extractable element-level variance were eligible for quantitative synthesis, contributing 41 distinct effects to the primary model. The PRISMA flow diagram (Figure 1) summarises the screening and inclusion process.

Figure 1

PRISMA Flow Diagram of Included Studies



Primary Meta-Analytic Model

The primary outcome was the within-group most–least (%) difference during testing after compound training. We fitted a multilevel model with random intercepts for study and effect, using CR2 small sample inference with Satterthwaite degrees of freedom (*df*) to reduce small-sample bias (Tipton, 2015; see Table 1).

Table 1

Primary Age-Adjusted Model (Logit Scale, CR2 Inference)

Coefficient	*β*	*SE*	*t (Satt)*	*df (Satt)*	*p (CR2)*
ASD vs TD	0.715	0.264	2.713	5.79	.036
Age (centred)	0.006	0.006	1.041	1.34	.450

Note. CR2 = small-sample robust tests; age is centred; random intercepts for Study and Effect. *β* = coefficient; *SE* = standard error; *t* = test statistic; *df* = degrees of freedom; *p* = probability value.

Heterogeneity and Variance

Random-effects variances on the logit scale were τ^2Study = 0.144 and τ^2Effect = 0.084. The approximate multilevel I^2 partition was Total ≈ 0.407, Study ≈ 0.257, and Effect ≈ 0.150, indicating that roughly 41% of variability reflected true differences, distributed across both studies and individual effects within studies.

Robustness and Covariates

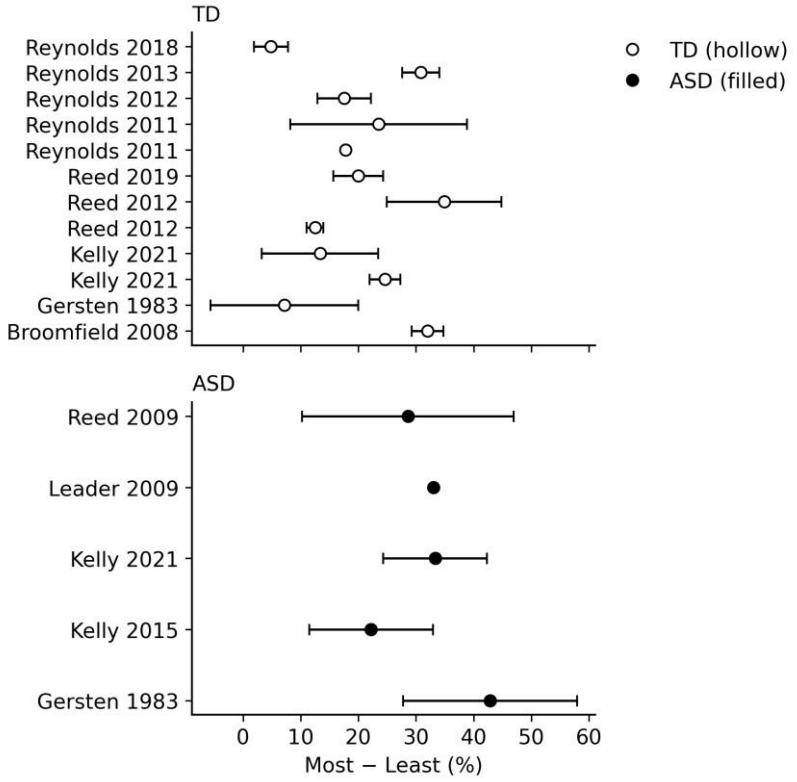
The model confirmed that the primary population gap between individuals with ASD and TD peers remained statistically significant even after adjusting for publication year (*β* = 0.682, *SE* = 0.249, *t* (5.42) = 2.74, *p* = .037). Notably, the slope for publication year itself was not statistically significant (*β* = −0.014, *SE* = 0.014, *p* = .405), and the age covariate remained non-significant (*p* = .311). These results indicated that the magnitude of restricted stimulus control was not independently determined by the year of study publication.

Population × Publication Year

A secondary model analysed the interaction between population and publication year to determine if the ASD–TD difference shifted across different diagnostic eras. The interaction was not statistically significant ($\beta = -0.012$, $SE = 0.042$, t (1.95) = −0.30, p = .793). Importantly, when this interaction was accounted for, the simple ASD effect was no longer statistically significant (p = .110).

Figure 2

Forest Plot of Study-Level Most–Least (%) Effects by Population, with Pooled Means



Note. Each point is a study-level most − least (%) difference with a 95% *CI* (inverse-variance aggregation within study when multiple rows were present). Some studies contributed both a TD and an ASD point, yielding 17 rows across both panels, corresponding to 15 unique studies in the primary model.

Descriptive Analysis: Three-Group Comparison

We conducted a three-group descriptive analysis to locate "other" clinical populations relative to the TD and ASD groups. Back-transformed means were TD = 21.02%, ASD = 32.18%, and Other = 49.30%. As the "other" category aggregated heterogeneous populations and procedures it was not included in the primary inferential model.

Sensitivity to Imputed Correlation

Refitting the age-adjusted model under $r = .30$ and $r = .60$ yielded the same ASD inference as the primary model ($\beta = 0.715$, $SE = 0.264$, $t\,(5.79) = 2.713$, $p = .036$), confirming that the population estimate was robust to the choice of imputed within-participant correlation.

Unadjusted Model Summary

To evaluate the base contrast without controlling for developmental maturation, we fitted an unadjusted model that excluded participant age as a covariate (see Table 2). This model, which did not include the chronological age adjustment, remained statistically significant ($\beta = 0.626$, $SE = 0.224$, $t\,(5.47) = 2.80$, $p = .035$).

A summary of all supplementary and sensitivity analyses, including publication year and interaction effects, is provided in Table 2. Adding publication year as a covariate did not alter the primary ASD conclusion ($p = .037$; see Table 2). Importantly, when the interaction between population and publication year was included, the simple effect of population was no longer significant ($p = .110$; see Table 2).

Influence Diagnostics

Leave-one-study-out diagnostics were examined to determine whether the ASD–TD contrast depended on any single study. Single-study omissions produced *CR2* p-values in the .021–.129 range, indicating that no single study determined the ASD–TD contrast. Omitting the oldest pre-2000 study shifted the test into the non-significant range within that band; the

direction of the effect did not change. This finding confirmed that the population estimate was sensitive to the inclusion of historical diagnostic samples, a point further explored in the Discussion.

Table 2

Summary of Supplementary and Sensitivity Analyses (Logit Scale, CR2)

Model	***β* (ASD vs TD)**	***SE***	***t (Satt)***	***df (Satt)***	***p***	**Conclusion**
Publication Year covariate	0.682	0.249	2.740	5.42	.037	Significant
Population × Year interaction (ASD main effect)	0.665	0.318	2.090	3.70	.110	Not significant
Population × Year interaction (interaction term)	-0.012	0.042	-0.300	1.95	.793	Not significant
Sensitivity: r = .30 / .60	0.715	0.264	2.713	5.79	.036	Robust to r
Excluding digitised effects	0.557	0.285	1.958	6.42	.095	Direction unchanged; statistical power reduced
Unadjusted (no Age)	0.626	0.224	2.797	5.47	.035	Significant

Note. All statistics use CR2 standard errors with Satterthwaite degrees of freedom. β = coefficient; *SE* = standard error; t = test statistic; *df* = degrees of freedom; p = probability value.

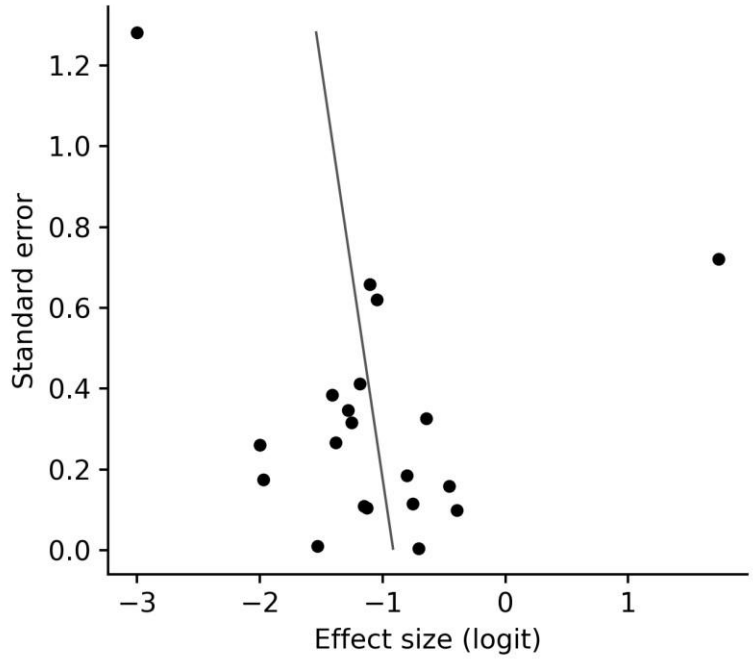
Small-Study Effects

To screen for potential publication bias or "small-study effects" (the tendency for smaller studies with lower statistical power to systematically report larger effect sizes than larger studies), we conducted an Egger-type regression on study-level aggregated effects.

This test evaluates funnel plot asymmetry as a proxy for reporting bias. The analysis did not indicate a clear asymmetry in the distribution of findings (z = –0.902, p = .367; see Figure 3).

Figure 3

Funnel Plot (Egger-Type Check)



Note. Each point represents a study-level aggregate (inverse-variance weighted within study). The vertical line marks the pooled effect; dotted lines show symmetric pseudo-95% bounds for orientation.

The findings above provide a quantitative synthesis of group differences and relevant moderators. Their interpretation and implications are considered in the following section.

Discussion

Summary of Primary Findings

This thesis provided a rigorous quantitative analysis of OS in individuals with autism spectrum disorder (ASD) relative to typically developing (TD) populations. The age-adjusted multilevel meta-regression identified a statistically significant population gap (p = .036). The

results indicated a 14.20% difference in overselectivity, with back-transformed means showing higher overselectivity among individuals with ASD (35.16%) compared to TD participants (20.96%). Although both populations exhibited some degree of overselectivity, individuals with ASD demonstrated a systematic narrowing of stimulus control that was significantly more pronounced. This outcome aligns with the view that overselectivity is a general, capacity-limited process across populations (Reed, 2006, 2012). The 14.20% gap remained stable across participant ages (p = .450). These findings are consistent with behavioural perspectives that identify restricted stimulus control as a feature of the ASD profile (Lovaas et al., 1971; Reed et al., 2012).

A critical finding was the inherent fragility of this population effect. Influence diagnostics showed that older ASD data heavily anchored the TD contrast. Omitting the single pre-2000 study (Gersten, 1983), or accounting for the Population × Year interaction shifted the model into the non-significant range (p = .110–.129). When only the 14 recent studies published from 2000 onwards were analysed, the population difference no longer reached statistical significance (p = .129). This suggests the observed gap may be specific to earlier diagnostic periods. As a result, the generalisability of the population effect to contemporary samples is limited, and these findings must be interpreted with caution.

This fragility has important practical implications. Clinicians and researchers should be cautious when interpreting the ASD–TD gap in overselectivity as a robust or universal feature of the current autism population (Lord & McGee, 2001; Rieth et al., 2015). Rather than assuming a persistent, clinically meaningful difference, these results indicate that the extent of restricted stimulus control in ASD is highly sensitive to historical context, changing diagnostic criteria, and sample characteristics. For clinical assessment and intervention, it is essential to avoid generalising older findings to present-day populations without considering recent diagnostic changes, intervention exposure, and functional profiles (Mason et al., 2021;

Reynolds & Reed, 2011b). For researchers, these findings highlight the need to replicate population-level contrasts using contemporary samples and standardised, element-level measurement. The observed effect is best viewed as provisional, reinforcing the need for ongoing critical evaluation as diagnostic and intervention practices evolve.

Historical Context and Data Fragility

The exploratory analysis of publication year revealed a key limitation. While the model initially suggested that publication year did not significantly moderate effect size variability (p = .405), this result is constrained by the limited historical data available for modelling. Sensitivity analyses suggested that the publication era may be as influential as participant age in determining the magnitude of overselectivity. When the interaction between population and publication year was included, the simple population difference was no longer statistically significant (p = .110). Influence diagnostics also showed that the primary ASD–historical data heavily anchored the TD contrast. Omitting the single study by Gersten (1983) shifted the meta-analytic test into the non-significant range (p = .129). This fragility implies that major diagnostic shifts, especially the transition from the DSM-III to the DSM-5 (American Psychiatric Association, 2013; Rieth et al., 2015), may be altering the extent to which OS characterises the contemporary ASD population. Rather than viewing the association as stable, it is more accurate to conclude that these findings are sensitive to the historical composition of the data pool. Further research using standardised element-level metrics is needed to evaluate how broadened diagnostic boundaries and increased early intervention have influenced the ASD–overselectivity link over time.

Stimulus Control Hierarchies

Bickel's (1986) more nuanced theoretical account offers a valuable framework for interpreting the patterns observed in the current results. Instead of viewing overselectivity as a simple lack of learning, Bickel posited that it reflects a stimulus control hierarchy, where

behaviour is influenced by a range of stimulus–response relations, including stimulus position, context, and even non-reinforced (S−) elements (Bickel, 1986; Bickel et al., 1984). This perspective is clinically important as it suggests overselective responding is not random but is governed by the wrong cues.

The current findings are consistent with this hierarchical view. By applying the most–least difference metric, this study quantified the degree of restricted stimulus control and identified heterogeneity in overselectivity across studies and effects (Mason et al., 2021). While the analytic approach did not directly determine which specific cues dominated responding for individual participants, the observed variation underscores the importance of considering not only whether overselectivity is present, but also how it may manifest differently across individuals and contexts (Bickel, 1986; Sidman, 1980).

Methodological Contributions and Data Constraints

A key contribution of this project is methodological. By defining the outcome as the most–least difference at test, we aligned the metric directly with the behavioural construct of restricted stimulus control (Bickel et al., 1984). This choice addressed common issues with relying only on overall accuracy or categorical cut-offs, which can obscure within-participant variability (Mason et al., 2021; Reynolds & Reed, 2011a). Logit-scale modelling corrected for statistical bias in bounded outcomes, and CR2 robust inference with Satterthwaite adjustments ensured valid conclusions, even with limited study clusters (Tipton, 2015). Despite these strengths, a substantial reporting gap remains, as only 15 of 158 relevant articles provided data suitable for our inclusion. The loss of statistical significance when digitised data were excluded ($p = .095$) underscores the field's reliance on incomplete or nonstandardised reporting. This combination of methodological advances and data limitations highlights the need for universal adoption of standardised element-level metrics and variance reporting in behaviour-analytic research. Routine reporting of the standard deviation of the

paired difference would greatly enhance the statistical power and cumulative value of meta-analytic estimates.

Methodological Limitations and the Reporting Gap

This study faced several methodological limitations that shape the interpretation of its findings. A substantial proportion of effect sizes came from digitised figures. Although this technique is considered reliable (Burda et al., 2017; Rohatgi, A., 2021), removing these 24 effects increased the standard error of the population estimate. As a result, while the direction of the ASD–TD gap remained consistent, the model no longer reached statistical significance under the reduced sample size (p = .095; see Table 2). This sensitivity analysis demonstrates that, although the primary finding did not result from extraction, digitised data were essential to achieve sufficient statistical power. This highlights the critical need for more standardised element-level reporting in future behavioural research. When researchers omit the specific metrics required for meta-analytic modelling, they restrict the field's ability to build a cumulative understanding of overselectivity. Future research should prioritise standardising the reporting of three key values: (1) mean element-level performance for all components of a compound stimulus, (2) the standard deviation (*SD*) for each element, and (3) either the within-participant correlation (*r*) or the standard deviation of the paired difference (*SDdiff*). Standardising these metrics will enable more precise meta-analyses and support stronger evidence-based approaches to remediation. The need for reporting consistency is not just technical, but it also directly impacts theoretical progress. Without transparent and complete data, it remains difficult to resolve ongoing debates about the mechanisms underlying overselectivity.

Theoretical and Mechanistic Implications

Despite these methodological constraints, the present results still inform key theoretical debates about the mechanisms responsible for overselectivity in ASD. Although a

14.20% gap was identified, it is important to recognise that meta-analytic models do not directly test processes. As such, these results do not provide a definitive conclusion on whether attention during learning or retrieval at test is the primary driver for OS. Nevertheless, the observed pattern of OS is consistent with stimulus control hierarchies (Bickel, 1986; Bickel et al., 1984). Rather than indicating a simple lack of learning, the results suggest that individuals with ASD may be guided by a heterogeneous set of relations. These could include stimulus position and context, which are often hidden when only aggregate accuracy scores are considered.

To move beyond group-level descriptions, future research should use experimental designs that can disentangle attentional and retrieval-based mechanisms. For example, pairing eye-tracking during learning with revaluation logic at test could clarify whether individuals with ASD fail to observe all relevant cues during training, or whether encoded cues are later suppressed during retrieval (Dube et al., 2010; Leader et al., 2009). Following compound training, revaluation procedures—such as extinguishing the overselected stimulus and observing if control shifts to previously underselected cues—could provide evidence for latent learning (Broomfield et al., 2008; Reed et al., 2009). Combining behavioural data with physiological measures such as pupillometry, may clarify whether attention and performance-based mechanisms are independent or interactive (Anderson et al., 2006). Integrative experimental studies of this kind may offer a more definitive test of overselectivity mechanisms in ASD.

Additionally, both participant age and publication year were modelled as linear covariates due to data constraints, which may have obscured more complex, non-linear patterns. Age likely follows a U-shaped developmental trajectory, and the influence of diagnostic shifts—such as the transition from DSM-III to DSM-5—is similarly non-linear and not fully captured by current models. The lack of meta-eligible historical data prevents

robust evaluation of how these diagnostic changes have shaped population estimates (American Psychiatric Association, 2013; Kelly & Reed, 2021; McHugh & Reed, 2007; Reed et al., 2013). As additional high-quality studies are published, future meta-analyses should model both developmental and historical factors using non-linear approaches.

Future Directions in Behavioural Research

To clarify the behavioural mechanisms underpinning overselectivity, future studies should integrate task-analytic instruction with revaluation logic within a single experimental design. For example, extinguishing a dominant (overselected) cue to assess whether previously underselected elements gain control would provide a more rigorous test of stimulus control than accuracy scores alone (Dube & Wilkinson, 2014; Reed, 2011). Such research will help determine whether observed clinical differences reflect a fundamental failure of initial attention or a relative failure of retrieval and stimulus control at the time of choice.

Clinical Implications

These findings offer concrete guidance for practitioners in educational and Applied Behaviour Analysis settings. Assessment and instruction should move beyond overall accuracy to systematically probe the degree of control exerted by each stimulus element. For example, if a learner responds correctly during a picture-to-word matching task but relies solely on the initial letter or a single visual feature, this pattern suggests OS is present (Birnie-Selwyn & Guerin, 1997; Dube & McIlvane, 1999). While restricted OS appears in both ASD and typically developing populations, the effect is more pronounced among individuals with ASD, potentially impeding the acquisition of complex skills (Birnie-Selwyn & Guerin, 1997; Reed et al., 2013). When element-level probing reveals that a single cue dominates, clinicians should implement evidence-based strategies. These may include differential observing responses (DOR), requiring learners to interact with each relevant

component (Dube & McIlvane, 1999; Farber et al., 2017), or using revaluation procedures such as withholding reinforcement for a dominant cue and providing training that highlights underselected elements (Leader et al., 2009; Reed et al., 2009). Increasing the requirement for observing responses (e.g., from FR1 to FR10) can also broaden control by ensuring adequate engagement with all stimuli (Doughty & Hopkins, 2011). Through such targeted interventions, practitioners can expand the range of stimulus control, thereby increasing learning opportunities for individuals affected by overselectivity.

Conclusion

This thesis identified a statistically significant 14.20% population gap in overselectivity between individuals with ASD and their typically developing peers (p = .036). However, this effect proved fragile, being heavily influenced by a single pre-2000 study (Gersten, 1983). When this data point was omitted, or when population and publication year were considered together, the group difference was no longer statistically significant (p = .129). Furthermore, the scope of this meta-analysis was limited by a pervasive reporting gap: only 15 of 158 relevant articles (9.5%) provided extractable element-level metrics for modelling. To advance the field, future research must consistently report mean element-level performance, the *SD* for each element, and the *SD* of the paired difference (*SDdiff*). Standardising these metrics is essential for tracking how the ASD–overselectivity relationship evolves as diagnostic boundaries change. Ultimately, understanding and remediating overselectivity will require both methodological advances and innovative, mechanism-driven experimental designs. Addressing these challenges will enable the field to better support the learning and development of individuals affected by restricted stimulus control.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Anderson, C. J., Colombo, J., & Shaddy, J. (2006). Visual scanning and pupillary responses in young children with autistic spectrum disorder. Journal of Clinical and Experimental Neuropsychology, 28(8), 1238–1256. https://doi.org/10.1080/13803390500376790

Bickel, W. K. (1986). A microanalysis of the controlling stimulus-response relations engendered during the assessment of stimulus overselectivity. *The Psychological Record, 36*(2), 225–238.

Bickel, W. K., Stella, M. E., & Etzel, B. C. (1984). A reevaluation of stimulus overselectivity: Restricted stimulus control or stimulus control hierarchies. *Journal of Autism and Developmental Disorders, 14*(2), 137–157. https://doi.org/10.1007/BF01438786

Birnie-Selwyn, B., & Guerin, B. (1997). Teaching children to spell: Decreasing consonant cluster errors by eliminating selective stimulus control. *Journal of Applied Behavior Analysis, 30*(1), 69–91. https://doi.org/10.1901/jaba.1997.30-69

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Wiley.

Broomfield, L., McHugh, L., & Reed, P. (2008). Re-emergence of under-selected stimuli, after the extinction of over-selected stimuli in an automated match to samples procedure. *Research in Developmental Disabilities, 29*(6), 503–512. https://doi.org/10.1016/j.ridd.2007.09.001

Brown, A., Cariveau, T., Ellington, P., & Platt, D. F. (2022). Overselectivity during reading related tasks for children at risk for reading failure. *Learning and Motivation*, *80*, 101849.

Cook, R. A., Anderson, N., & Rincover, A. (1982). Stimulus overselectivity and stimulus control: Problems and procedures. In R. L. Koegel, A. Rincover, & A. L. Egel (Eds.), *Educating and understanding autistic children* (pp. 90–105). College Hill Press.

Dickson, C. A., Wang, S. S., Lombard, K. M., & Dube, W. V. (2006). Overselective stimulus control in residential school students with intellectual disabilities. Research in Developmental Disabilities, 27(6), 618–631. https://doi.org/10.1016/j.ridd.2005.07.004

Doughty, A. H., & Hopkins, M. N. (2011). Reducing stimulus overselectivity through an increased observing-response requirement. *Journal of Applied Behavior Analysis, 44*(3), 653–657. https://doi.org/10.1901/jaba.2011.44-653

Dube, W. V. (2009). Stimulus overselectivity in discrimination learning. In P. Reed (Ed.), *Behavioural theories and interventions for autism* (pp. 23–46). Nova Science Publishers.

Dube, W. V., Dickson, C. A., Balsamo, L. M., O'Donnell, K. L., Tomanari, G. Y., Farren, K. M., Wheeler, E. E., & McIlvane, W. J. (2010). Observing behavior and atypically restricted stimulus control. *Journal of the Experimental Analysis of Behavior, 94*(3),

297–313. https://doi.org/10.1901/jeab.2010.94-297

Dube, W. V., Farber, R. S., Mueller, M. R., Grant, E., Lorin, L., & Deutsch, C. K. (2016). Stimulus overselectivity in autism, Down syndrome, and typical development. *American Journal on Intellectual and Developmental Disabilities, 121*(3), 219–235. https://doi.org/10.1352/1944-7558-121.3.219

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., & Fowler, T. R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin, 17*(1), 8–14.

Dube, W. V., & McIlvane, W. J. (1997). Reinforcer frequency and restricted stimulus control. *Journal of the Experimental Analysis of Behavior, 68*(3), 303–316. https://doi.org/10.1901/jeab.1997.68-303

Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis, 32*(1), 25–33. https://doi.org/10.1901/jaba.1999.32-25

Dube, W. V., & Wilkinson, K. M. (2014). The potential influence of stimulus overselectivity in AAC: Information from eye-tracking and behavioural studies of attention with individuals with intellectual disabilities. *Augmentative and Alternative Communication, 30*(2), 172–185. https://doi.org/10.3109/07434618.2014.904924

Dinsmoor, J. A. (1985). The role of observing and attention in establishing stimulus control. Journal of the Experimental Analysis of Behavior, 43(3), 365–381. https://doi.org/10.1901/jeab.1985.43-365

Fabio, R. A., Pergolizzi, G., Nucita, A., Iannizzotto, G., & Caprì, T. (2021). The role of a virtual avatar in attention and memory tasks in Rett syndrome. *BMC Neurology, 21*(1), Article 1. https://doi.org/10.1186/s12883-021-02212-w

Farber, R. S., Dickson, C. A., & Dube, W. V. (2017). Reducing overselective stimulus control with differential observing responses. *Journal of Applied Behavior Analysis, 50*(1), 87–105. https://doi.org/10.1002/jaba.363

Fields, L., & Spear, J. (2012). Measuring joint stimulus control by complex graph/description correspondences. *The Psychological Record, 62*(3), 413–430.

Gersten, R. M. (1983). Stimulus overselectivity in autistic, trainable mentally retarded, and non-handicapped children: Comparative research controlling chronological (rather than mental) age. *Journal of Abnormal Child Psychology, 11*(1), 61–75. https://doi.org/10.1007/BF00912178

Grey, H. C., Morris, C., Perrin, J., & Oliveira, J. P. (2024). Evaluating the effects of compound stimuli on stimulus control during match-to-sample procedures. *Behavior Analysis in Practice, 17*(4), 1211–1215. https://doi.org/10.1007/s40617-024-00990-6

Gutowski, S. J., & Stromer, R. (2003). Delayed matching to two-picture samples by individuals with and without disabilities: An analysis of the role of naming. *Journal of Applied Behavior Analysis, 36*(4), 487–505. https://doi.org/10.1901/jaba.2003.36-487

Hedbring, C., & Newsom, C. (1985). Visual overselectivity: A comparison of two instructional remediation procedures with autistic children. *Journal of Autism and Developmental Disorders, 15*(1), 9–22. https://doi.org/10.1007/BF01837895

Kelly, M. P., Leader, G., & Reed, P. (2015). Stimulus over-selectivity and extinction-induced recovery of performance as a product of intellectual impairment and autism severity. *Journal of Autism and Developmental Disorders, 45*(10), 3098–3106. https://doi.org/10.1007/s10803-015-2471-x

Koegel, R. L., & Schreibman, L. (1977). Teaching autistic children to respond to simultaneous multiple cues. *Journal of Experimental Child Psychology, 24*(2), 299–311. https://doi.org/10.1016/0022-0965(77)90008-X

Koegel, R. L., Schreibman, L., Britten, K., & Laitinen, R. (1979). The effects of schedule of reinforcement on stimulus overselectivity in autistic children. *Journal of Autism and Developmental Disorders, 9*(4), 383–397. https://doi.org/10.1007/BF01531446

Leader, G., Loughnane, A., McMoreland, C., & Reed, P. (2009). The effect of stimulus salience on over-selectivity. *Journal of Autism and Developmental Disorders, 39*(2), 330–338. https://doi.org/10.1007/s10803-008-0626-y

Lord, C., & McGee, J. P. (Eds.). (2001). *Educating children with autism*. National Academy Press.

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autism spectrum disorders. In M. Rice & S. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 7–29). Lawrence Erlbaum Associates.

Lovaas, O. I., Schreibman, L., Koegel, R. L., & Rehm, R. (1971). Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology, 77*(3), 211–222. https://doi.org/10.1037/h0031015

Mason, D., Capp, S. J., Stewart, G. R., Kempton, M. J., Glaser, K., Howlin, P., & Happé, F. (2021). A meta-analysis of outcome studies of autistic adults: Quantifying effect size, quality, and meta-regression. *Journal of Autism and Developmental Disorders, 51*(9), 3165–3179. https://doi.org/10.1007/s10803-020-04784-2

McHugh, L., & Reed, P. (2007). Age trends in stimulus overselectivity. *Journal of the Experimental Analysis of Behavior, 88*(3), 369–380. https://doi.org/10.1901/jeab.2007.88-369

McHugh, L., & Wood, R. (2013). Stimulus over-selectivity in temporal brain injury: Mindfulness as a potential intervention. *Brain Injury, 27*(13–14), 1595–1599. https://doi.org/10.3109/02699052.2013.834379

Mueller, M. M., Olmi, D. J., & Saunders, K. J. (2000). Recombinative generalization of within-syllable units in prereading children. *Journal of Applied Behavior Analysis, 33*(4), 515–531. https://doi.org/10.1901/jaba.2000.33-515

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Perez, W. F., Endemann, P., Pessôa, C. V. B. B., & Tomanari, G. Y. (2014). Assessing stimulus control in a discrimination task with compound stimuli: Evaluating testing procedures and tracking eye fixations. *The Psychological Record, 65*(1), 83–88. https://doi.org/10.1007/s40732-014-0092-1

Ploog, B. O. (2010). Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of*

Autism and Developmental Disorders, 40(11), 1332–1349. https://doi.org/10.1007/s10803-010-0990-2

Ploog, B. O., Brooks, P. J., Scharf, A., & Aum, S. (2014). Perception of the prosody and content of sentences in an unfamiliar language in children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 8*(7), 775–787. https://doi.org/10.1016/j.rasd.2014.03.014

Reed, P. (2011). Comparator mechanisms and autistic spectrum conditions. In T. R. Schachtman & S. R. Reilly (Eds.), *Associative learning and conditioning: Human and animal applications* (pp. 168–190). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199730223.003.0009

Reed, P. (2017). Over-selectivity is related to autism quotient and empathizing, but not to systematizing. *Journal of Autism and Developmental Disorders, 47*(2), 405–414. https://doi.org/10.1007/s10803-016-2990-3

Reed, P. (2019a). Previous mindfulness experience interacts with brief mindfulness induction when reducing stimulus overselectivity. *Applied Cognitive Psychology, 33*(4), 558–564. https://doi.org/10.1002/acp.3474

Reed, P. (2019b). Unpredictability reduces over-selective responding of individuals with ASD who have language impairments. *Research in Autism Spectrum Disorders, 57*, 35–45. https://doi.org/10.1016/j.rasd.2018.10.006

Reed, P., Altweck, L., Broomfield, L., Simpson, A., & McHugh, L. (2012a). Effect of observing-response procedures on overselectivity in individuals with autism spectrum

disorders. *Focus on Autism and Other Developmental Disabilities, 27*(4), 237–246. https://doi.org/10.1177/1088357612457986

Reed, P., Savile, A., & Truzoli, R. (2012b). Event related potential analysis of stimulus overselectivity. *Research in Developmental Disabilities, 33*(2), 655–662. https://doi.org/10.1016/j.ridd.2011.11.012

Reed, P., Reynolds, G., & Fermandel, L. (2012c). Revaluation manipulations produce emergence of underselected stimuli following simultaneous discrimination in humans. *The Quarterly Journal of Experimental Psychology, 65*(7), 1345–1360. https://doi.org/10.1080/17470218.2012.656663

Reed, P., Broomfield, L., McHugh, L., McCausland, A., & Leader, G. (2009). Extinction of over-selected stimuli causes emergence of under-selected cues in higher-functioning children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(2), 290–298. https://doi.org/10.1007/s10803-008-0629-8

Reed, S. R., Stahmer, A. C., Suhrheinrich, J., & Schreibman, L. (2013). Stimulus overselectivity in typical development: Implications for teaching children with autism. *Journal of Autism and Developmental Disorders, 43*(6), 1249–1257. https://doi.org/10.1007/s10803-012-1658-x

Reynolds, G., & Reed, P. (2011a). Effects of schedule of reinforcement on over-selectivity. *Research in Developmental Disabilities, 32*(6), 2489–2501. https://doi.org/10.1016/j.ridd.2011.07.011

Reynolds, G., & Reed, P. (2011b). The strength and generality of stimulus over-selectivity in simultaneous discrimination procedures. *Learning and Motivation, 42*(2), 113–122. https://doi.org/10.1016/j.lmot.2010.12.001

Reynolds, G., & Reed, P. (2013). Effect of a surprising downward shift in reinforcer value on stimulus over-selectivity in a simultaneous discrimination procedure. *Learning and Motivation, 44*(1), 31–45. https://doi.org/10.1016/j.lmot.2012.05.007

Reynolds, G., & Reed, P. (2018). The effect of stimulus duration on over-selectivity: Evidence for the role of within-compound associations. *Animal Learning & Cognition, 44*(3), 293–308. https://doi.org/10.1037/xan0000175

Rieth, S. R., Stahmer, A. C., Suhrheinrich, J., & Schreibman, L. (2015). Examination of the prevalence of stimulus overselectivity in children with ASD. *Journal of Applied Behavior Analysis, 48*(1), 1–14. https://doi.org/10.1002/jaba.165

Schneider, H. C., & Salzberg, C. L. (1982). Stimulus overselectivity in a match-to-sample paradigm by severely retarded youth. *Analysis and Intervention in Developmental Disabilities, 2*(4), 273–304. https://doi.org/10.1016/0270-4684(82)90025-8

Schover, L. R., & Newsom, C. D. (1976). Overselectivity, developmental level, and overtraining in autistic and normal children. *Journal of Abnormal Child Psychology, 4*(3), 289–298. https://doi.org/10.1007/BF00917765

Sidman, M. (1980). A note on the measurement of conditional discrimination. Journal of the Experimental Analysis of Behavior, 33(3), 285–297. https://doi.org/10.1901/jeab.1980.33-285

Strand, S. C. (1991). Attentional aspects of classroom behavior and discrimination learning.

Research in Developmental Disabilities, 12(3), 229–241. https://doi.org/10.1016/0891-4222(91)90009-H

Tipton, E. (2015). Small sample adjustments for robust variance estimation with metaregression. *Psychological Methods, 20*(3), 375–393. https://doi.org/10.1037/met0000039

Wilhelm, H., & Lovaas, O. I. (1976). Stimulus overselectivity: A common feature in autism and mental retardation. *American Journal of Mental Deficiency*, *81*(1), 26–31.