Underinformative Implicature Derivation on the Broader Autism Phenotype

Rylee A. Fennell

Haverford College
Abstract

Underinformative implicatures (UIs) are a type of inference that can give rise to two potential meanings. UIs can be either literal, which do not require the use of social inferences, or pragmatic, which necessitate social inferences. Given that individuals with autistic traits have marked deficits in communicative and social skills, there is reason to believe that these individuals approach deriving UIs differently than typically developing (TD) individuals. The current study is an attempt to replicate van Tiel and Kissine’s (2018) online assessment of how individuals on the broader autism phenotype (BAPs) derive these inferences. The present study utilizes a previously untested technique to assess UI derivation; it employs a graded scale for participants to rate the appropriateness of stimuli rather than a binary response option as all prior research had done (Pijnacker, Hagoort, Buitelaar, Tumisse, & Geurts, 2009; Chevallier, Wilson, Happé, & Noveck, 2010; Su & Su, 2015; Van Tiel & Kissine, 2018). The current study found significant differences between how all BAPs and typically developing (TD) participants respond, but not between how low scoring individuals on the BAP (Low BAPs) compared to high scoring individuals on the BAP (High BAPs) based on autism spectrum quotient scores. The present study found nearly the exact opposite of what van Tiel and Kissine (2018) found in their study. Instead of only High BAPs responding more literally on distributivity implicatures, all BAPs responded more pragmatically on all implicature types except for distributivity. In addition, these results showed evidence of cognitive inflexibility via the tendency to answer using the ends of the scale on lexicalizable implicature types. The current study concludes that there are indeed significant differences in how BAPs respond.
to UIs in comparison to TDs. It is possible that such differences were revealed through the employment of a graded rating scale. These results are highly relevant to making educational spaces more linguistically inclusive for neurodiverse populations.

Keywords: implicature derivation, broader autism phenotype, rating scale, replication
IMPLICATURE DERIVATION ON THE BAP

Underinformative Implicature Derivation on the Broader Autism Phenotype

Contents

Abstract 2

Underinformative Implicature Derivation on the Broader Autism Phenotype 4

Scalar Implicature Derivation ........................................ 6

Diagnosing Autism ..................................................... 7

Implicature Derivation in Individuals with an ASD Diagnosis .......... 8

Implicature Derivation in the Broader Autism Phenotype .............. 10

Flaws in the Existing Literature .................................... 13

The Present Study 18

Method 21

Stimuli ................................................................. 21

Procedure ............................................................ 21

Participants ........................................................... 24

Analysis ............................................................... 26

Results 27

Data Preparation ...................................................... 27

AQ Scores ............................................................ 27

BAP vs. TD ........................................................... 28

High vs. Low BAP .................................................... 34
Scalar Implicature Derivation

Scalar implicatures are inferences that can be considered true from a logical perspective but false from a pragmatic one (Pijnacker et al., 2009). Pragmatic inferences require the enactment of Grice’s maxims of conversation, which are listed below (Grice, 1975).

1. Quality: Be as truthful as possible
2. Relation: Make relevant contributions
3. Manner: Be as clear and concise as possible
4. Quantity: Give only as much information is needed

These maxims dictate the social rules interlocutors abide by in conversation, therefore making them highly relevant to whether one derives a logical or pragmatic inference upon hearing a scalar implicature. For example, the sentence,

1) “Some birds are robins”

is unambiguously true. This is because there is no pragmatic reasoning required in determining if robins are birds because one relies solely on the common ground knowledge, nor does it violate any Gricean maxims. However, the utterance

2) “Some robins are birds”

can be interpreted two ways: literally or pragmatically. The literal interpretation is considered to be true. This is because the meaning of ‘some’ can be adjusted to mean ‘some, and possibly all.’ We see how ‘some’ may or may not mean ‘all’ through examples
where ‘some’ can be either felicitously corrected to mean ‘all’ or denote a subset of the group in question. For example, if a professor entered a classroom and stated,

3) “Some of you passed the exam, in fact you all did”

she would not raise any concern among her students. However if the professor had entered the room and only announced,

4) “Some of you passed the exam”

students would likely panic because this could imply that not everyone passed the exam. Students’ reaction to 4) is considered to be the pragmatic derivation. This is because the students assumed that if the professor could have been more informative, in this case more accurately reporting that a greater number of students passed the exam, she would have. Pragmatic derivations therefore assume that the speaker has observed Grice’s maxim of quantity (Grice 1975). Thus, referring back to the case of 2), if there were a more informative utterance available, such as suggesting that there are some robins that are in fact not birds, the speaker should have said so because it runs counter to common knowledge and could cause confusion. The pragmatic listener would likely judge the scalar implicature of 2) to be false since it is assumed that all robins are birds rather than just a subset of robins that are birds. It is also important to note that scalar implicatures, such as the ones discussed above, are not the only type of quantity implicature. Three additional quantity implicatures will also be addressed in the present study.

Diagnosing Autism

Although Asperger’s Syndrome and Autistic Disorder (commonly known as autism) were previously considered to be separate disorders, they are now both classified as Autism
Spectrum Disorder (ASD). These disorders were classified together in the newest edition of the Diagnostic and Statistical Manual for Mental Disorders (DSM-V). This was done because the only former distinction between autism and Asperger’s, that individuals with Asperger’s did not show evidence of language delays, was found to be inaccurate. Recent research has shown that language delays are not a defining factor in the disorder; rather, the emphasis is now placed on one’s ability to communicate socially regardless of any potential delays in speech (Gernsbacher, Morson, & Grace, 2016; Su & Su, 2015). Upon this discovery, the DSM-V reclassified Asperger’s and Autistic Disorder as ASD to emphasize that this disorder will have an array of presentations, yet are all inherently the same. The DSM-V states that the following are the two criteria for ASD: “persistent deficits in social communication and social interaction across multiple contexts” and “restricted, repetitive patterns of behavior, interests, or activities” (Association et al., 2013). Given that the only difference between these two disorders is a variation in the observable presentation of symptoms rather than the disorder itself, it is possible to compare previous work that assesses Asperger’s vs. Autistic Disorder to studies who utilize the new classification of ASD.

**Implicature Derivation in Individuals with an ASD Diagnosis**

Implicature derivation is of interest in the context of autism because it requires the listener to reason about the speaker’s intentions, which is an inherently social skill. The prevailing theory has therefore been that individuals with an ASD diagnosis would have difficulty drawing pragmatic inferences. Pijnacker et al. (2009) was the first to question whether or not individuals with an ASD diagnosis would struggle more with deriving
pragmatic readings from scalar implicatures than typically developing participants (TDs). The authors found that individuals with high functioning autism (HFA) and Asperger’s diagnoses were able to derive scalar implicatures. This research was conducted using a binary truth value judgment task (TVJT) procedure. Notably, participants with an Asperger’s diagnosis performed significantly better than their HFA and control counterparts (Pijnacker et al. (2009), Fig. 1, p. 6). However, these results could be due to the authors’ separation of patients with Asperger’s vs. HFA. Perhaps if analyses had been conducted combining these two groups, as was done in Su and Su (2015) (which is discussed later in this section), there would be no such differences and thereby mimic more recent results.

These results prompted Chevallier et al. (2010) to attempt to replicate Pijnacker et al.’s (2009) study. Chevallier et al. (2010) argued that, since children have been found to derive implicatures less often than adults, scalars may be an effortful task that could be learned through experience. To control for any learned techniques, the authors used a similar TVJT to Pijnacker et al. (2009) but tested adolescents with an ASD diagnosis instead of adults with HFA or Asperger’s. Chevallier et al. (2010) found different results than Pijnacker et al. (2009): there were no significant differences in implicature derivational tendencies in participants with an ASD diagnosis and TDs. These results offer compelling support for the hypothesis that individuals with an ASD diagnosis do not differ significantly from controls and counters Pijnacker et al. (2009). Chevallier et al.’s (2010) study also supports the notion that there are not significant differences in pragmatic language usage in individuals with an ASD diagnosis.

Su and Su (2015) attempted to further generalize the findings of Pijnacker et al.
(2009) and Chevallier et al. (2010) by testing 4-15 year-old Mandarin speakers with varying diagnoses of ASD and Asperger’s using a binary TVJT. The authors found an insignificant difference between participants with an ASD-like diagnosis and TDs. These results further strengthen the argument that deriving scalar implicatures is not a learned strategy in individuals with an ASD diagnosis, and that it may be an intact function. In addition, it also supports Su and Su’s (2015) attempt to generalize prior findings because they have now shown that a different age group with a very different native language demonstrate the same patterns as participants who are both older and speak English. Unlike Pijnacker et al. (2009), Su and Su (2015) chose to group participants with an ASD-adjacent diagnoses as one experimental group. The work presented in this study supports the DSM-V’s modification of diagnosing of HFA and Asperger’s disorder as a single disorder [Su & Su 2015]. This is because there were found to be no differences between the ASD-like group and the TDs regardless of their more specific diagnoses.

Implicature Derivation in the Broader Autism Phenotype

A final rendition of the hypothesis that there are differences in how individuals with an ASD diagnosis and TDs interpret quantity implicatures was conducted by van Tiel and Kissine (2018). However, three important changes from the existing literature were introduced. Firstly, this study analyzed data from individuals on the broader autism phenotype (BAP) instead of individuals with a confirmed ASD diagnosis. Individuals on the BAP have not received a formal ASD diagnosis, but are reported to have high levels of autistic traits as established by the Autism Spectrum Quotient (AQ). The AQ was recently reassessed by [Ruzich et al., 2015] and found to be a reliable and valid measure of autistic
traits. The authors opted to use a non-clinical population in order to both further generalize their results and reach a broader population through an online paradigm.

Verifying that individuals have an ASD diagnosis online is nearly impossible, but measuring how many autistic traits one displays can be accomplished through a reliable and valid scale. Secondly, the authors conducted their study using online experiment databases rather than in a laboratory as in the previous studies. Specifically, the authors recruited subjects from Prolific, Mechanical Turk, Autism Research, WrongPlanet, and Reddit.

The third change van Tiel and Kissine (2018) introduced was testing three other quantity implicature types in addition to scalars: conditional, exhaustivity, and distributivity implicatures. Examples of these implicatures can be seen in Figure 1. This was done in order to account for three concerns regarding quantity implicature derivations. The authors suggested four factors contribute to the likelihood of an individual on the BAP succeeding on this task. Firstly, the researchers suggested that BAPs would still be able to infer a pragmatic meaning of a scalar implicature using the social motivation theory of autism. This theory argues that individuals with autistic traits would still make the proper derivation with enough contextual clues even if it necessitates that one to take on the perspective of another person. Next, they proposed that scalars are lexicalizable. This is to say that the literal meaning can be remembered rather than deriving the pragmatic reading. The authors argued that since autism or autistic tendencies do not impact lexical knowledge, there should be no difference between a neurotypical individual and someone on the spectrum in this regard. A third factor that could influence how BAPs derive quantity implicatures is due to the potential of selective pragmatic impairment. If individuals with autistic traits struggle with pragmatic interpretations, they may have learned
compensatory strategies to assist them in reaching pragmatic derivations. Such an example would be learning to utilize context to aid in difficult derivations. Van Tiel and Kissine (2018) also claim that scalar derivation is too simple of a process that enables the listener to reject stronger options. Finally, the authors proposed that alternatives of scalars are inherently contextualized in their meaning which can clue listeners in to their derivation (Van Tiel and Kissine (2018), p. 7). A replication of van Tiel and Kissine’s (2018) hypotheses table for all four implicature types can be seen in Table 1.

<table>
<thead>
<tr>
<th>Scalar</th>
<th>Distributivity</th>
<th>Conditional</th>
<th>Exhaustivity</th>
<th>Soc mot</th>
<th>Lex</th>
<th>Sel imp</th>
<th>Alt</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Soc mot: social motivation theory; Lex: lexicalizability; Sel imp: selective pragmatic impairment; Alt: alternatives approach (van Tiel and Kissine, 2018, p. 15).

Van Tiel and Kissine (2018) hypothesized that BAPs should have a more difficult time deriving quantity implicatures as they increase in complexity and contextually dependent alternatives. The researchers found that only distributivity implicatures were significantly more challenging for BAPs. Van Tiel and Kissine (2018) reasoned that distributivity implicatures required more complex analysis than simply reasoning about possible alternatives (as could be done with the other implicatures). This paper specifically has prompted the present study to further investigate these results. This research was also motivated by changing the technique used during the experiment itself. However, the current study agrees with van Tiel and Kissine (2018) that distributivity implicature
derivations specifically could pose additional challenges for BAPs due to their increased reliance on social awareness via perspective taking.

**Flaws in the Existing Literature**

**Binary Judgments.** All of the previously mentioned literature test implicature derivation using binary TVJT. These tasks operate by asking participants to observe a stimulus, in these cases an implicature, and decide whether it is true or false. While the simplicity of this design is desirable, it may not capture the full range of complexity needed to derive an implicature. Sikos, Kim, and Grodner (2019) argue that by coercing an awkward or inappropriate decision, one cannot determine if the implicature derivation or the decision making process is causing the challenge. Take the example of 2) “Some robins are birds.” Accepting that this is a true statement may be awkward for pragmatic responders because it violates the maxim of quantity. These individuals interpret selecting ‘true’ is allowing underinformativeness when it is generally assumed that the interlocutor could be more informative in this case. Thus, ‘false’ as their only alternative. While one can argue that ‘false’ is an acceptable answer when operating under the maxim of quantity, it could also be claimed that the literal response still holds merit because it abides by the maxims of quality. Hence there is no ‘correct’ answer provided in a binary TVJT task which likely causes confusion for the participant. Current research supports the notion that one’s tolerance for the literal interpretation can vary depending on the speaker’s social attributes for binary judgments, but not for graded ones. This indicates that, when judgment abilities are less accessible, binary judgments are affected and could result in unreliable data (Sikos, Kim, & Grodner 2019).
Additionally, one cannot discern anything about the participants’ reasoning. For example, a participant may derive the pragmatic inference but find the literal one to be acceptable as well and report it over the pragmatic one. Prior studies are therefore not measuring one’s ability to derive pragmatic implicatures, but their willingness to do so. Such intricacies may be overlooked in the framework of existing research, and thus necessitate an explanation as to why implicature derivation is difficult for TDs, individuals who have an ASD diagnosis, and individuals who fall on the BAP. There is also reason to believe that individuals with higher levels of autistic traits may utilize a graded judgment differently from TDs, which could not be explored in the binary template. Justification for this belief begins with understanding what associative learning is and how individuals with an ASD diagnosis use it.

**Associative Learning and ASD.** People learn in many different ways. Sapey-Triomphe et al. (2018) examined how individuals on the autism spectrum differ from TDs in the learning strategies they employ. The authors found in their associative learning study that individuals with an ASD diagnosis often use what is referred to as the lookup-table strategy (LUT). Rather than regularizing information to form categories that can be further specialized when needed, individuals on the spectrum memorize items individually based on what they have observed and do not infer any underlying rules in the behaviors (Sapey-Triomphe, Sonié, Hénaff, Mattout, & Schmitz, 2018). Thus, when learning how to interpret a UI, individuals with high levels of autistic traits may have previously observed other speakers; they then make a pragmatic derivation using the LUT strategy to mirror that pragmatic derivation in their own speech despite not drawing the inference themselves. Conversely, TDs often use the interpolation (INT) strategy, wherein
individuals may attempt to identify implicit rules in interaction and reason through new stimuli to find the most appropriate way to represent the new input.

However, autistic traits can take a variety of forms; while some individuals may struggle to use the INT strategy, others may be more successful. It therefore would be most prudent to consider how individuals falling at different places on the Autism Spectrum Quotient (AQ) will utilize associative learning. There is reason to believe that individuals who still possess a significant number of autistic traits but do not score as highly as other participants will show a different pattern of responses than those who score the highest. More moderately scoring participants may prefer the literal derivation of UIs as opposed to higher scoring participants because they are less frequently relying on the LUT strategy. Instead, these participants may attempt to identify implicit rules in interaction and reason through new stimuli to find the most appropriate way to represent the new input as TDs do. What these individuals do differently from TDs is that when they draw their inferences, they are not incorporating the social information, such as understanding Grice’s maxim of quantity, into their decision making. Individuals with less extreme levels of autistic traits may draw a literal derivation if they use the INT strategy because they fail to integrate social information into their decision making. This is supported by Norbury et al. (2010), who found that children with ASD did not improve after multiple exposures to a visual associative learning task learning nonce words, unlike their TD counterparts. While both groups were able to succeed to some degree in the task, TD participants attended more to the speaker’s face and less to remembering the exact phonological representation of the nonce word, while the children with an ASD diagnosis did the opposite and did not show signs of improvement.}\hhref{Norbury, Griffiths, & Nation}
This strengthens the theory that individuals with autistic traits may not integrate new information if it has a social basis, particularly if they can employ other learning tactics instead. These results support the hypothesis that individuals with moderate levels of autistic traits will favor the literal interpretation over the pragmatic because it avoids utilizing social cues while still showing evidence of reasoning through the stimulus.

Evidence that individuals on the autism spectrum can use the INT strategy comes from research on memory and ASD (Williams, Goldstein, & Minshew, 2006). The authors found that children ages 8-16 with ASD had lower rates of success with complex visual and verbal information, but had intact associative learning abilities, verbal working memory, and recognition skills. This suggests that there may be differences in the organization of memory abilities in individuals with an ASD diagnosis. The discovery that individuals with ASD have poor complex visual and verbal skills in comparison to TDs suggests that when they are presented with a situation as complex as a social interaction using UIs, social cues are the first to be eliminated, therefore reallocating their cognitive resources to deriving meaning through associative learning. However, if the meaning is contingent on the social context, it is possible that the meaning will be lost in the process. This is highly relevant in the case of UIs. If social information is likely to be eliminated when one encounters a complex stimuli, it is likely that the participant with at least moderate levels of autistic traits draws a literal derivation because they were not able to attend to both the sentence and the social inference.

However, not all of the literature supports the idea that individuals with an ASD diagnosis have challenges with social inferencing (Hani, Gonzalez-Barrero, & Nadig, 2013). The authors examined children’s (ages 3-6) using social cues to learn words and their
ability to use referential understanding. Referential understanding is defined here as the ability to use a newly learned label to identify an object of the same category (i.e., “dog” does not solely mean one’s own dog, but all animals falling under the same general description) (Hani et al., 2013, p. 974). The authors also investigated differences in parent communication styles between the two groups, though this is of secondary interest for my purposes. (Hani et al., 2013) contrary to the previous studies, that children with ASD were able to use social referencing to label novel objects. These results offer a basis for individuals with an ASD diagnosis to use social referencing at an early age. While word learning itself is not relevant to the current study, these results demonstrate that alternative learning styles, including those that necessitate social interaction, can be employed by children with an ASD diagnosis. Therefore, if the current study does not find any significant results between BAPs and TDs, it may just be that no such differences exist.

Inflexibility in ASD. Another hallmark of autism is cognitive inflexibility. Granader et al. (2014) proposed that executive function (EF) deficits are the primary cause of this inflexibility. EF is defined here as impulse control, problem solving, planning and executing goals, cognitive flexibility, and working memory (Granader et al., 2014). Cognitive inflexibility in ASD often takes the form of emphasis on routine adherence, sameness, and unchanging behavioral and thought patterns (Granader et al., 2014). To examine how well individuals with an ASD diagnosis utilize EF, the researchers administered the Behavioral Rating Inventory of Executive Functioning (BRIEF) to the parents of 411 children ages 5-18. The authors found that the “shift” subscale, which measures cognitive flexibility (here defined as attempting the same challenge repeatedly despite it not working), was significantly higher in the ASD group compared to controls in
everyday situations (Granader et al., 2014). This suggests that once someone with an ASD diagnosis has made up their mind, they are significantly more likely to stay with that decision regardless of if it appropriate or not.

Evidence of cognitive inflexibility relates to the current study in how individuals with an ASD diagnosis will utilize a scale, which was not provided to them in earlier research. Once a participant has come to a conclusion about a UI, either using LUT in individuals with the highest scores or INT but omitting social cues in the moderately scoring participants, it is likely that they will not reason through the scale to pick the most appropriate answer. For many individuals, a sentence like 2) “Some robins are birds” is neither right nor wrong because they interpret it flexibly. This is to say that while one makes the pragmatic derivation, the TD participant is also able to understand the literal derivation. However, individuals with higher levels of autistic traits may not possess this ability to reason flexibly either as a result of using associative learning or EF deficits.

When using a 7-point scale to rank UIs, this may manifest itself in BAPs primarily using the extremes of the scale (1, 2, 6, and 7) significantly more than the control group because they are unable to see beyond the initial answer they came up with. In other words, BAPs will adhere so strongly to their initial interpretation of the UI (whether it is pragmatic or literal) to the point where they are unable to reason through any of the midpoints on the scale.

The Present Study

The present study examines two key hypotheses. Firstly, it explores how the response options in the previous study affect results. Van Tiel and Kissine (2018) only provided
binary true/false response options, which could obscure results by not accounting for the difficulty choosing between two awkward choices or associative learning. This thesis utilizes a 7-point Likert scale that asks how appropriate the sentence describes the accompanying image rather than the traditional binary judgment.

**Hypothesis 1.** BAPs will utilize the far ends of the scale significantly more than TDs on target stimuli as a result of associative learning or cognitive inflexibility.

In the case of associative learning, the mid-range values of the scale will likely not be used because once an implicature is learned associatively, it is no longer an inference. Therefore, there is no need to use the middle of the scale. In the case of cognitive inflexibility, BAPs struggle to both deviate from previously acquired knowledge and to view something as simultaneously correct and incorrect. Details about which participants will use associative learning or have cognitive inflexibility is provided in Hypothesis 2. The current hypothesis would predict that BAPs will more frequently select 1, 2, 6, and 7 compared to their typically developing individuals.

Secondly, the present study argues that individuals with varying levels of autistic traits prefer pragmatic or literal derivations of lexicalizable underinformative implicatures. Previous work hypothesized that individuals with an ASD diagnosis should prefer the literal derivation as a result of social deficits that are commonly associated with ASD (Pijnacker et al., 2009; Chevallier et al., 2010; Su & Su, 2015). However, the prior studies generally did not find any significant differences between individuals with an ASD diagnosis and TDs until van Tiel and Kissine (2018) expanded their study’s scope to include different types of implicatures. While van Tiel and Kissine (2018) found that BAPs exhibited significant differences in their derivation of distributivity implicatures in
comparison to TDs, there is reason to believe that all of the previous literature has not considered why individuals with autistic traits make the derivations they do. The current study hopes to address this gap in the literature.

**Hypothesis 2a.** The highest scoring individuals on the AQ will answer scalar and exhaustivity (lexicizable implicatures) pragmatically because they will have primarily utilized associative learning (thereby memorizing that “some” is solely encoded to mean “not all” and does not require the speaker to make any pragmatic inferences of their own). This would predict that the highest scoring participants will primarily respond using the lower end of the scale (1, 2, and 3) to indicate that they find the underinformative implicature to be an inappropriate description of the accompanying image.

**Hypothesis 2b.** Alternatively, those with significant, yet fewer, autistic traits will favor literal derivation on lexicalizable implicatures because they are drawing their own inferences rather than using associative learning, but are forgoing the social component in the derivation. This would predict that moderately scoring participants would primarily respond using the lower end of the scale (5, 6, and 7) to indicate that they find the underinformative implicature to be an appropriate description of the accompanying image.

Van Tiel and Kissine (2018) also examined the relationship between vocabulary skills as a measure of language competence to follow up Pijnacker et al.’s (2009) and Chevallier et al.’s (2010) assessment. While Pijnacker et al. (2009) and Chevallier et al. (2010) found vocabulary size to be a significant predictor of a subject’s likelihood of responding pragmatically, van Tiel and Kissine only found this relationship to be marginally significant ($p = .09$) for distributivity implicatures only; all others were not significant. The current study hypothesizes that there will be no relationship between vocabulary size and likelihood
to respond pragmatically, although this relationship will not be explored in depth. It is only performed to execute a complete replication of van Tiel and Kissine (2018).

**Method**

**Stimuli**

All participants will see scalar, distributivity, conditional, and exhaustivity quantity implicatures. Each participant will see 15 statements per sentence type. 5 of these sentences will be control questions that necessitate a ‘completely appropriate’ response, 5 will be control questions that necessitate a ‘completely inappropriate’ response, and an additional 5 will be the target underinformative implicatures. Examples of each implicature type, control questions, and stimuli can be found in Figure 1 (Van Tiel & Kissine, 2018). Participants each saw a total of 60 stimuli. Each sentence was paired with a corresponding image composed of different combinations of four colors and three shapes (red, yellow, blue, green; circle, triangle, square). It is important to note that it is appropriate for the figure to describe the controls as ‘true’ or ‘false.’ This is because there is no derivation needed; these questions are unambiguously correct or incorrect and should not result in any differences among the experimental and control groups. In the event that they are, non-acceptance testing will be performed to ensure that participants are still reliably differentiating between true and false stimuli, but are perhaps making a few errors or find a certain stimulus more ambiguous than it was intended to be.

**Procedure**

Participants who signed up to complete the study on Prolific were directed to the current study, which was run through Ibex Farm. The experiment can be accessed via
Upon clicking the link to the study, participants were presented with a consent form. This form explained the risks and benefits of participation, which were potential eye strain from looking at a computer and contributing to the field of linguistics, respectively. They were also shown a brief description of what the study entails and a reminder of participants’ rights when taking the study. No identifiable personal information is recorded, and anonymity is ensured. By selecting “yes” at the bottom of the screen, participants acknowledge that they have reviewed and understood the terms of participation and are over 18 years of age.
Participants were then brought to a form asking for their Prolific identification code, which was used to ensure their anonymity. They were also asked for their age, gender identity, native language, and highest level of education reached to establish basic demographics. Participants were then asked if they experience any form of colorblindness, as this could affect their ability to distinguish the color of the shapes in the experiment. They were also asked to report if they have ever been diagnosed with an autism spectrum disorder or if they believe that they have an autism spectrum disorder. There were also two attention checking questions.

From there, participants were directed to the explicit instructions of the study. Specifically, subjects were told that they would be asked to use a scale of 1-7 to rank how “appropriate” a description was of the accompanying image. They were informed that a rating of ‘1’ means that the statement is “completely inappropriate” while a rating of ‘7’ is “completely appropriate.” In order to encourage participants not to overthink each judgments, participants were advised to try their best to pick the answer that feels the most natural to them and not to spend too long on one question. Once participants selected that they have read and understood these directions, they were able to begin the experiment. They are shown two practice trials before beginning the randomized test trials. After making their judgments for the 60 test stimuli, participants took the AQ followed by the vocabulary test. They were then thanked for their participation and their results were submitted to the server, where they will be automatically anonymized and recorded.
Participants

30 TD and 26 BAP participants were recruited through the online service Prolific (Prolific, 2019). The present study’s participant number mirrors van Tiel and Kissine’s (2018) study very closely; the original paper assessed 36 TDs and 27 BAPs. Subjects self-identified as either TD or on the BAP. Only participants who self-identified as being on the BAP were considered to be part of the experimental group, regardless of their AQ score. Similarly, TDs who had higher AQ scores were still grouped as TD due to their personal identification preferences. However, this is not believed to limit the statistical power of the present study, as the TD and BAP groups were found to be significantly different from each other despite any variation within each group. Further information about AQ scores for TDs and BAPs can be found in Table 2.

Within the BAP group, 16 are considered to be Low BAPs (BAPs with an AQ score under 30), while 10 are considered to be High BAPs (BAPs with an AQ score of 30 or higher). 13 TD and 14 BAP participants identified as female, and 17 TDs and 11 BAPs identified as male. No other gender identities were reported. TD subjects had a mean age of 34.9 (SD = 11, range 19-54), while experimental participants had a mean age of 32.9 (SD = 8.3, range 21-51). TDs took an average of 29.4 minutes to complete the study (SD = 13 minutes, range = 11-78 minutes). BAPs took an average of 27.9 minutes to complete the study (SD = 10, range = 17-78 minutes). The average education level of both groups was a bachelor’s degree (range = high school-PhD). Of the BAPs, 3 reported being diagnosed with ASD as a child, 4 reported being diagnosed with ASD as an adult, 3 reported being in the process of obtaining an ASD diagnosis, 14 reported that they
identified as being on the autism spectrum but have not received a diagnosis, and 1 elected not to answer. Given that this is an online study, there is no way to verify the answers provided. All subjects were compensated at a rate of $8/hour for their time.

Van Tiel and Kissine (2018) did not provide demographic data by group; rather, they reported this information as one large group that consists of both TDs and BAPs. The authors found a mean age of 34 (SD = 11, range = 18-64). 32 participants identified as male (no other gender identities were reported). The average time to complete the study was 29 minutes (SD = 9 minutes). The authors did not specify the Autism Spectrum Quotient (AQ) scores of each group; they only reported that the average AQ score was 27 with a range of 7-46 (Van Tiel & Kissine, 2018, p. 17). The authors did not report the average education achieved by participants or the range in their findings. From what can be assessed here, there appears not to be significant differences between the population recruited in the present study and that of van Tiel and Kissine’s (2018) study.

The only potential difference could come from the rate of individuals who affirmed that they had received an ASD diagnosis. Van Tiel and Kissine (2018) reported that 19 individuals reported that they received an ASD diagnosis, while 41 did not. In the present study, 3 participants reported being diagnosed with ASD as a child, 4 reported being diagnosed with ASD as an adult, 3 reported that they are in the process of receiving an ASD diagnosis, 14 self-identify as being on the autistic spectrum, and 1 did not provide an answer. This may affect how participants rate sentences in the experiment.
Analysis

The current study used a two-way ANOVA to analyze results. Specifically, it has examined the interaction between three dependent variables, which are the ratings participants provide for each question. The independent variables are the inference types, truth values, AQ scores, and the interaction between them. It also assessed non-acceptance rates and tendencies towards extremes. Each analysis was conducted both for BAPs vs. TDs and Low BAP vs. High BAP. In the case of model non-convergence, the current study would adopt the procedures in Bates, Mächler, Bolker, and Walker (2014). However, no such measures were needed for the present study.

Data was discarded from participants who exhibit a non-acceptance rate of under 70%, have any form of colorblindness, were under the age of 18, or reported that English was not their first language or were not fluent in English.

Additional analyses will be conducted outside the scope of this thesis for future publication. Data gathered for the present study will be analyzed using linear mixed effects regression. The independent variables are the four implicature types, participant scores on the AQ, and the interaction between implicature types and AQ scores on both hypotheses. In one set of models, the entire AQ scale will be used. In the second set, only the social and communicative subscales will be used. This is because these two scales are shown to be relevant to ASD. The following is the model that will be used to conduct additional analyses for each implicature type:

\[
\text{sf.choice.model} \leftarrow \text{glmer(\text{TrueJudgment AQ * InferenceType + VerbInteraction + (ASQ * InferenceType | SUBJECT) + (AQ * InferenceType | ITEM)}, family = \text{binomial})}
\]
data = under[under$InferenceType %in% c("si," "fc"),]

The dependent variables in these later analyses are the number of pragmatic answers (i.e., those answered false on the binary task or 1, 2, or 3 on graded judgments) participants select and how frequently participants use the ends of the scale (1, 2, 6, and 7). The independent variables are the four implicature types tested in van Tiel and Kissine (2018) and AQ scores. Future work will also adopt Bates’ (2014) non-convergence procedures as in this thesis Bates et al. (2014).

**Results**

**Data Preparation**

One participant on the BAP and two TD participants were excluded from analysis for having error rates over 30%. Upon obtaining the data, it became apparent that using a standard 10% error exclusion rate was inappropriate for the current study because it did not take into consideration that individuals would struggle with control stimuli. While responses to what should have been patently true and false answers occasionally varied, participants’ likelihood to reliably differentiate between true and false (i.e., non-acceptance rates) has proven to be an accurate measure. Thus, the following analyses will use this criterion to exclude participants. This will be discussed in the following section.

**AQ Scores**

The average TD AQ score was 17.6 (range 8-29, SD = 6.20), while the average BAP score was 27.9 (range 17-39, SD = 6.51). The subscale scores for TD and BAP subjects can be found in Table 2. All subscale scores between BAPs and TDs were significantly different except for the imagination subscale ($p = .283$). The average Low BAP score was 25.5
Typically developing subjects and subjects on broader autism phenotype’s tendency to reject a sentence.

(range 17-29, SD = 3.97). While the average High BAP score was 34.1 (range 30-39, SD = 2.76). Table 3 lists the subscale scores for Low BAP and High BAP subscale scores. Low BAPs and High BAPs were significantly different on all subscale measures except attention to detail (p = .219).

While significant differences within the AQ subscales are potentially interesting, assessing how these scores affect ratings on both implicature types and truth values is too complex for the purpose of the current study. The present study will therefore only examine total AQ scores for the purpose of grouping participants. It is suggested that future research address how particular subscale scores are associated with acceptability judgments.

**BAP vs. TD**

The following sub-section will report how the entire BAP compares to TDs in regards to the present study’s hypotheses. It will address these hypotheses in the order in which they are presented in the method.
Table 2
*BAP vs. TD AQ Subscale Scores*

<table>
<thead>
<tr>
<th>AQ Subscale</th>
<th>BAP Mean</th>
<th>SE</th>
<th>TD Mean</th>
<th>SE</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Com</td>
<td>5.31</td>
<td>.383</td>
<td>2.43</td>
<td>.361</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Soc</td>
<td>6.50</td>
<td>.458</td>
<td>3.13</td>
<td>.395</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Atd</td>
<td>5.77</td>
<td>.210</td>
<td>4.37</td>
<td>.370</td>
<td>.003*</td>
</tr>
<tr>
<td>Asw</td>
<td>7.38</td>
<td>.434</td>
<td>4.47</td>
<td>.367</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Total</td>
<td>28.6</td>
<td>1.33</td>
<td>17.6</td>
<td>1.13</td>
<td>&lt;.001*</td>
</tr>
</tbody>
</table>

Com = communication; soc = social skills; atd = attention to detail; asw = attention switching; img = imagination.

**Tendency towards extremes.** Hypothesis 1. states that individuals on the BAP should use the ends of the scales, or “extreme” ratings, as a result of cognitive inflexibility. Data for each implicature type can be seen in Table 3. Only scalar implicatures showed significant evidence of extreme responses ($p = .013$). A trend towards significance was observed in all other implicatures as well as the overall rating of the four inference types. This suggests that when an implicature is highly lexicalizable, BAPs are more likely to use the ends of the scale regardless of its truth value. This is supported by van Tiel and Kissine’s (2018) proposition that items that can be lexicalized are easier for participants on the BAP to interpret, despite this not being shown in their own data. In terms of the present study, this data also supports BAPs lacking the cognitive flexibility to consider the middle of the scale on difficult stimuli.

**Ratings.** Individual implicature ratings were first calculated for the entire BAP in comparison to TDs. These results can be seen in Figure 3. UI responses from both participants did not behave as predicted. Specifically, TD acceptance rates are substantially more literal (i.e., accepting of the UI) than had been anticipated, while BAPs
were trending towards more pragmatic. These results directly oppose those found in Pijnacker et al. (2009), Chevallier et al. (2010), and Su and Su (2015), all of whom saw both TD subjects and subjects with an ASD diagnosis overwhelmingly respond pragmatically to scalar UIs. While van Tiel and Kissine (2018) found more moderate results in their data (both Low BAPs and High BAPs responded pragmatically roughly 50% of the time on all implicatures, aside from low BAPs favoring the pragmatic interpretation significantly more for distributivity implicatures), there was still no trend towards literal or pragmatic responding for either population. Contrary to what van Tiel and Kissine (2018) had expected, distributivity implication derivations do not seem to add any additional challenges for individuals on the BAP despite necessitating perspective taking. True control questions were much more uniform; all participants found control true statements to be true in all four implicature types.

Table 3
Low BAP vs. High BAP AQ Subscale Scores

<table>
<thead>
<tr>
<th>AQ Subscale</th>
<th>Low BAP</th>
<th>High BAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>Com</td>
<td>4.50</td>
<td>.428</td>
</tr>
<tr>
<td>Soc</td>
<td>5.63</td>
<td>.584</td>
</tr>
<tr>
<td>Atd</td>
<td>5.56</td>
<td>.273</td>
</tr>
<tr>
<td>Asw</td>
<td>6.38</td>
<td>.539</td>
</tr>
<tr>
<td>Img</td>
<td>2.56</td>
<td>.329</td>
</tr>
<tr>
<td>Total</td>
<td>24.6</td>
<td>1.31</td>
</tr>
</tbody>
</table>

Com = communication; soc = social skills; atd = attention to detail; asw = attention switching; img = imagination.

Participants’ ratings on different UI implicature types were all significant except for distributivity implicatures. Explicit results are shown in Table 5. Conditional \((p = .016)\), scalar, \((p = .028)\), exhaustive \((p = .024)\), and overall combined \((p = .025)\) implicatures
were significant. Means on these implicatures were always at least one point higher for TDs than BAPs, suggesting that BAPs are more harshly judging UIs than TDs. Distributivity implicatures, however, were not significant \((p = .521)\). This is the exact opposite of the results found by van Tiel and Kissine (2018) (p. 22); the authors found no significant results for any implicature rating differences between the two groups aside from distributivity, which was found to become increasingly literal as one’s AQ score increased.

*Figure 3.* Typically developing subjects and subjects on broader autism phenotype’s mean implicature ratings with standard errors.
Given that there are minimal differences between the current population and the population tested in the original study, this is likely not due to individual differences.

There was also a significant effect between ratings, group identity (BAP or TD), and implicature type for ‘false’ stimuli. Further details can be seen in Table 6. These data further suggest BAPs are responding more harshly than TDs on all questions as a result of cognitive inflexibility. Truth value and group identity were examined for an interaction of
Table 4
*BAP vs. TD Extreme Responses*

<table>
<thead>
<tr>
<th>Implicature Type</th>
<th>BAP Mean</th>
<th>SD</th>
<th>TD Mean</th>
<th>SD</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>distributivity</td>
<td>.723</td>
<td>.376</td>
<td>.693</td>
<td>.327</td>
<td>.075</td>
</tr>
<tr>
<td>conditional</td>
<td>.608</td>
<td>.351</td>
<td>.740</td>
<td>.311</td>
<td>.141</td>
</tr>
<tr>
<td>scalar</td>
<td>.492</td>
<td>.389</td>
<td>.727</td>
<td>.295</td>
<td>.013 *</td>
</tr>
<tr>
<td>exhaustive</td>
<td>.523</td>
<td>.404</td>
<td>.673</td>
<td>.334</td>
<td>.134</td>
</tr>
<tr>
<td>overall</td>
<td>.587</td>
<td>.298</td>
<td>.708</td>
<td>.223</td>
<td>.086</td>
</tr>
</tbody>
</table>

BAP vs. TD tendency towards responding using the extremes of the scale, defined as answering with a score of ‘1’, ‘2’, ‘6’, or ‘7’. Overall represents the average of all four implicature types.

ratings. There was a significant effect of group identity ($p = .018$) and a significant interaction between truth value and group identity ($p = .006$). This also suggests that BAPs are responding more pragmatically on UIs.

**Non-acceptances.** Figure 2 reports the non-acceptance rates of participants, while Table 7 reports the significant values per implicature type. Non-acceptances are defined as a participant’s willingness to reject a sentence. TDs responded unexpectedly on control false items, particularly conditional falses and distributivity falses. Difficulty may have arisen from the stimuli itself or a general complexity associated with two uncommon sentence structures.

BAPs were much more rigid in their willingness to accept false statements ($M = .96$, $SD = .06$) in comparison to TDs ($M = .89$, $SD = .09$). This supports the hypothesis that these individuals as a whole prefer the archetypal form of a given utterance. This is to say that while TDs reasoned through control items based on their own judgments (which will be discussed in the next section), BAPs overwhelmingly answered “correctly” on true/false questions. This is potentially the result of cognitive inflexibility. If the subject is unable to
reason through the statement, they may instead lean towards answering more harshly for the sake of simplicity.

This was also found to be the case for UIs. BAPs continued their pattern of responding significantly more pragmatically on all implicature types except on distributivity implicatures ($p = .171$). $P$ values for the other implicature types can be found in Table 8. Further analyses will be needed to confirm if this trend mimics the true/false pattern seen in van Tiel and Kissine’s (2018) to show the accuracy of the replication. Given the current data, it is likely that adding a rating scale makes participants more sensitive to stimuli. However, this falls outside the scope of the current paper.

**High vs. Low BAP**

The following analyses compare differences within the BAP population. As previously discussed, these groups are divided into Low BAP and High BAP depending on one’s AQ score. Given the lack of significance in the following findings, all tables for Low BAP vs. High BAP are located in the appendix.

**Tendency towards extremes.** Table A1 displays the extreme responses rate
between Low BAPs and High BAPs. None of these values were significant or showed signs of trending towards significance (overall $p = .880$). This indicates that these two sub-populations are not exhibiting a difference in pragmatic or literal responses.

**Ratings.** Figure 4 is a boxplot of the average ratings on all implicature types and truth values, outliers, and the inter-quartile range. As hypothesized, High BAPs trended towards rating UIs lower (i.e., less acceptable/more pragmatically), than Low BAPs, who in turn trended more towards scoring UIs higher (i.e., more acceptable/literally). However, this trend is marginal at best.

Low BAPs and High BAPs did not significantly differ on how they rated different UI implicatures (overall $p = .666$). However, there is a marginal trend towards significance for scalar and exhaustivity UIs, which fits in line with hypothesis 2, which suggested that Low BAPs and High BAPs will differ on items that are lexicalizable because of their ability to be memorized. Additional details can be found in Table A2.

The effects of truth value on group identity through ratings is also insignificant, despite a trend towards significance on ‘false’ items (overall $p = .761$, false $p = .198$). The same holds true for the interaction of truth condition ($p = .450$) and the interaction between group and truth condition ($p = .736$). Tables A3 and A4, respectively, provide
Table 7

<table>
<thead>
<tr>
<th>Implicature Type</th>
<th>BAP Mean</th>
<th>SD</th>
<th>TD Mean</th>
<th>SD</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>distributivity</td>
<td>.492</td>
<td>.450</td>
<td>.407</td>
<td>.381</td>
<td>.171</td>
</tr>
<tr>
<td>conditional</td>
<td>.446</td>
<td>.413</td>
<td>.213</td>
<td>.315</td>
<td>.020 *</td>
</tr>
<tr>
<td>scalar</td>
<td>.400</td>
<td>.375</td>
<td>.187</td>
<td>.344</td>
<td>.031 *</td>
</tr>
<tr>
<td>exhaustive</td>
<td>.523</td>
<td>.427</td>
<td>.273</td>
<td>.377</td>
<td>.024 *</td>
</tr>
<tr>
<td>overall</td>
<td>.465</td>
<td>.315</td>
<td>.270</td>
<td>.296</td>
<td>.020 *</td>
</tr>
</tbody>
</table>

BAP and TD average tendency to reject false statements. ‘Overall’ indicates the average score of a group’s response on all UI sentences.

Additional data.

**Non-acceptances.** Finally, there were no significant differences between Low BAPs and High BAPs on any non-acceptance measures with no trends towards significance (overall $p = .808$). This supports the notion that once subjects are over a threshold of autistic traits, they behave similarly regardless of how far above their score is from said threshold. More information can be found in Table A5.

**Vocabulary Scores**

Differences in TD ($M = 76.6$, $SD = 13.1$, $SE = 2.39$) and BAP ($M = 82.7$, $SD = 11.0$, $SE = 2.15$) performance on the vocabulary assessment was only marginally significant ($p = .067$). This suggests that it is potentially possible for BAPs to have an advantage taking the experiment due to a better command of the English language. Even if this were to be the case, it only strengthens the argument that BAPs have decreased cognitive flexibility in comparison to TDs because even with an advantage in vocabulary, BAPs were still unable to reason through individual items. Further details can be seen in Table A5.

Low BAPs ($M = 83.1$, $SD = 8.85$, $SE = 2.21$) and High BAPs ($M = 82.0$, $SD = 14.2$, $SE$
Table 8
*BAP vs. TD Interaction Effects of Truth Value and Group on Ratings*

<table>
<thead>
<tr>
<th>Truth Value</th>
<th>BAP</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>UI</td>
<td>4.45</td>
<td>1.37</td>
</tr>
<tr>
<td>F</td>
<td>1.72</td>
<td>.575</td>
</tr>
<tr>
<td>T</td>
<td>6.86</td>
<td>.2148</td>
</tr>
</tbody>
</table>

BAP and TD average score on each truth condition. The effect of the group was significant (*p* = .018 *), as was the effect of the interaction of the truth value and group (*p* = .006 *).

= 4.49) once again did not significantly differ (*p* = .805). This suggests that language skills do not affect the performance of either group in relation to each other. Additional analyses regarding how vocabulary scores predict ratings for each implicature type (as van Tiel and Kissine (2018) performed) are too intricate for the purposes of the present study, but could be an avenue for future research.

**Discussion**

The present study found unexpected yet significant results between BAPs and TDs, and a lack of significant differences between Low BAPs and High BAPs. The current findings provide evidence for Hypothesis 1, which proposed that BAPs will use the ends of the scale more frequently than TDs as a result of cognitive inflexibility or associative learning. BAPs responded pragmatically significantly more than TDs. This is likely the result of their elevated levels of cognitive inflexibility. The significant interaction between truth value and participant group suggests that individuals on the BAP are interpreting UIs differently than TDs. This is strengthened by the fact that a three-way interaction between all three independent variables is significant as well. Thus, BAPs rated sentences,
regardless of inference type truth value, more harshly (i.e., less willing to rate a sentence using 5, 6, or 7) than TDs. In addition, TDs were more accepting of all statements (included falses, as discussed in the context of non-acceptance) than BAPs.

This finding contradicts the results found in all previous studies wherein TDs responded overwhelmingly pragmatically. The tendency for TDs to respond more literally was even found to be the case on control items, wherein TDs were far more forgiving on ‘false’ questions than participants on the BAP. Such response patterns, particularly on the part of TDs, may be an intrinsic flaw in the questions themselves because all participants struggled with conditional and distributivity false control questions. However, two things are important to note at this point. Firstly, these questions were taken directly from van Tiel and Kissine’s (2018) study without any modification. Secondly, van Tiel and Kissine’s (2018) study did not show any evidence of difficulty on control stimuli. Thus, it is possible that providing more response options has highlighted an underlying challenge in these questions that render them inappropriate as control stimuli. It is also possible that the provision of a scale causes all subjects to be more sensitive to how they truly interpret the sentence than could be seen in a binary response system.

Hypothesis 1 can also be explored through examining the non-acceptance rates because rejecting statements. BAPs more frequently rejected both UIs and potentially challenging false questions. The overall trend in more pragmatic scoring may be a result of participants on the BAP exhibiting cognitive inflexibility in a different way than previously predicted. BAPs may more frequently reject false and underinformative stimuli as a reaction to knowing what “should” be done rather than trying to parse each example as TDs did. This likely explains why TDs answered “incorrectly” more frequently. BAPs are
more likely to consider a sentence unacceptable by not allowing ambiguity to affect their judgments as a result of the complexity of the task. In the case of UIs, the added factor of social information likely impacts their willingness to accept a statement because it is not factored into the decision making process (Norbury et al., 2010). While BAPs are not responding in the way the current study had predicted, it is clear that BAPs are more limited in their ability to flexibly consider any stimulus.

A product of this inflexibility emerges through differences in BAP vs. TD tendency to utilize extreme responses. Because scalars show significant evidence of this and all other implicatures show a trend towards significance, we must consider how lexicalizability affects BAPs’ interpretation of UIs. This not only offers support for Hypothesis 1, but also demonstrates the extent to which BAPs will rate a statement they consider to be incorrect. Scalars are thus likely the only truly significant value because they are the easiest to lexicalize. Lexicalizing scalars turns the UI into a simple true or false question, thereby erasing the need for any middle ground. Given that there is a trend towards significance for all other items, perhaps there are two response patterns co-occurring within the population. Participants may truly be responding using extremes in a way that does not relate to one’s level of autistic traits, and is instead a result of individual preference. Many participants may be using the extremes as a result of simplifying the UI either through associative learning or cognitive inflexibility. Others, however, may be unable to make a decision as a result of not being able to reason through it due to mental inflexibility, and thereby select ‘4’ as a neutral response as a way to avoid addressing the complexity of the sentence. Thus, the results show an indication of significant difference between BAPs and TDs in terms of responding using extremes, but it is not fully realized. This theory requires
further testing, but offers a potential avenue for future research.

Hypothesis 2 was not supported by the results of the present study, which posited that Low BAPs would rate UIs more literally as a result of cognitive inflexibility, while High BAPs would rate UIs more pragmatically as a result of associative learning. Low BAPs and High BAPs were nearly identical in their performances, with no significant differences between them. The current study accepts the null hypothesis for Hypotheses 2a and 2b, which is that there are no significant differences between Low BAPs and High BAPs as a result of associative learning or cognitive inflexibility. It should be noted that the present results for Low BAPs and High BAPs contradict van Tiel and Kissine’s (2018) findings, which demonstrated a significant trend for Low BAPs to respond more pragmatically on distributivity UIs than High BAPs.

It is also pertinent to examine how the present study relates to the existing body of literature regarding how individuals with autistic traits derive quantity implicatures. The results of this study counter those found in Chevallier et al. (2010), Su and Su (2015), and van Tiel and Kissine (2018). This is because the present study found significant differences in how individuals on the BAP respond to both control and target stimuli in comparison to TDs, which was not the case in any of the studies listed. The results found in the current study are potentially the result of methodological improvements made in the current study. However, the present study cannot conclusively refute the findings of Pijnacker et al. (2009). Given that this thesis’ results show a trend towards significance between Low BAPs’ and High BAPs’ UI ratings, it is possible that this data could be significant if statistical power could be enhanced via recruiting more participants. Therefore, the present study offers preliminary evidence that there could be a difference between response
behaviors in Low BAPs and High BAPs and thereby support Pijnacker et al.’s (2009) findings of UI derivation distinctions within the autism spectrum, but such conclusions cannot be made at this time.

Limitations

One limitation of the present study was the inability to verify demographic data provided by participants as a result of conducting a web-based study. This is particularly evident in the inability to verify where one falls on the autism spectrum. However, this was also the case for van Tiel and Kissine’s (2018) study, so it should not be considered a strong reason to doubt the validity of the present findings. A second limitation was the number of TD participants with unusually high AQ scores. There were 8 TD participants who fell within one standard deviation of the Low BAP average AQ score. Thus, this research is potentially confounded by the fact that a substantial portion of the TD population possesses higher levels of autistic traits than the general public. This is likely the result of conducting the study online, as earning monetary compensation through a web-service requires little social interaction and therefore may be a favorable opportunity for individuals with higher levels of autistic traits. An additional limitation is the population size. While the current study was only one participant away from reaching the number of BAPs in van Tiel and Kissine’s (2018) study, this is still too small of a population to generalize to the entire broader autism phenotype. Future research should make a conscious effort to recruit as many participants as possible to enhance statistical power. A final limitation is the diagnostic rate in the present sample; van Tiel and Kissine (2018) procured 19 subjects who reported an ASD diagnosis, while the current study only
reported 3. Although this variation is difficult to control, future research should take additional care to ensure that there as many self-reported ASD diagnoses as possible.

Conclusions

This study found significantly different results from van Tiel and Kissine’s (2018) study. Van Tiel and Kissine (2018) had found that High BAPs were significantly less likely to respond pragmatically on only distributivity UIs than Low BAPs and TDs. The present study found no significant differences between Low BAPs and High BAPs, which also runs counter to what van Tiel and Kissine (2018) found. The current research found the reverse; BAPs rated conditional, exhaustivity, and scalar UIs significantly less literally than TDs, with distributivity UIs being the only insignificant value. These unexpected results should be considered alongside the significance of BAPs’ scalar extreme responses and the trend towards significance for all other implicature types, and their elevated non-acceptance rates. When examined collectively, these results suggest that there is an effect of cognitive inflexibility for BAPs that prevents them from considering all response options in the same way as TDs. There was also no distinguishable difference between Low BAPs and High BAPs, suggesting that once a threshold is met for autistic traits, subjects on the BAP respond similarly. These results were likely a product of providing participants with a graded response scale instead of binary options. This is because using a graded scale allowed participants to reflect their ability to derive UIs rather than their willingness to do so. This study provides relevant and implementable results for individuals seeking to create linguistically inclusive spaces for neurodiverse populations such as schools, work environments, and personal connections. These results also offer insight into how
individuals with autistic traits interpret information at varying levels of complexity and informativity, which is relevant in all domains of life.

**Future Research**

To reiterate the above AQ Subscales discussion, it is highly surprising that there are a number of significant differences between Low BAP and High BAP AQ Subscale scores, yet nearly no significant differences in how Low BAPs and High BAPs responded to the experimental stimuli. Future work may choose to seek why this is. For the purposes of publishing these findings, more complex analyses, specifically those included in the latter portion of the analysis section, will be conducted to more explicitly determine the effects of one’s rating score, vocabulary score, AQ and sub-scale scores, and more. Future work may also address the tendency for BAPs’ preference in choosing more neutral ratings when presented with UIs. Additional research may also pursue replicating other studies using a graded scale rather than a binary TVJT to contextualize the results found in the current study. Further investigation of the present results may also opt to replicate the existing body of literature using a rating scale to specifically examine the marginally significant trend for Low BAPs and High BAPs to respond differently from each other. If such data were reported, one could simultaneously support Pijnacker et al.’s (2009) findings that there is variation within response behaviors on the autism spectrum and refute the findings of Chevallier et al. (2010), Su and Su (2015), and van Tiel and Kissine (2018) that no such distinction exists.
Acknowledgments

I’d like to thank Professor Dan Grodner for working with me for the duration of this project. I’d also like to thank my advisor, Professor Amanda Payne, my second reader, Professor Shizhe Huang, and my student readers, Anya Capps and Travis Herringshaw for all their guidance and feedback throughout this process. I am grateful to Haverford and Swarthmore Colleges for funding this research. An additional thanks to Gerrit Farren for his help with programming. A final thank you to my family – none of this would have been possible without your love and support.
References


Appendix

Data Summary

Table A1
Low BAP vs. High BAP Extreme Responses

<table>
<thead>
<tr>
<th>Implicature Type</th>
<th>Low BAP Mean</th>
<th>Low BAP SD</th>
<th>High BAP Mean</th>
<th>High BAP SD</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>distributivity</td>
<td>.750</td>
<td>.403</td>
<td>.680</td>
<td>.343</td>
<td>.653</td>
</tr>
<tr>
<td>conditional</td>
<td>.613</td>
<td>.354</td>
<td>.600</td>
<td>.365</td>
<td>.932</td>
</tr>
<tr>
<td>scalar</td>
<td>.513</td>
<td>.406</td>
<td>.460</td>
<td>.378</td>
<td>.745</td>
</tr>
<tr>
<td>exhaustivity</td>
<td>.500</td>
<td>.386</td>
<td>.560</td>
<td>.450</td>
<td>.721</td>
</tr>
<tr>
<td>overall</td>
<td>.594</td>
<td>.272</td>
<td>.575</td>
<td>.351</td>
<td>.880</td>
</tr>
</tbody>
</table>

Low BAP vs. High BAP tendency towards responding using the extremes of the scale, defined as answering with a score of ‘1’, ‘2’, ‘6’, or ‘7’. Overall represents the average of all four implicature types.

Table A2
Low BAP vs. High BAP Rating Scores

<table>
<thead>
<tr>
<th>Implicature Type</th>
<th>Low BAP Mean</th>
<th>Low BAP SD</th>
<th>High BAP Mean</th>
<th>High BAP SD</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>distributivity</td>
<td>4.51</td>
<td>2.18</td>
<td>4.22</td>
<td>1.90</td>
<td>.730</td>
</tr>
<tr>
<td>conditional</td>
<td>4.48</td>
<td>1.86</td>
<td>4.52</td>
<td>2.09</td>
<td>.955</td>
</tr>
<tr>
<td>scalar</td>
<td>4.35</td>
<td>1.87</td>
<td>4.96</td>
<td>1.18</td>
<td>.366</td>
</tr>
<tr>
<td>exhaustivity</td>
<td>4.10</td>
<td>1.62</td>
<td>4.72</td>
<td>1.81</td>
<td>.372</td>
</tr>
<tr>
<td>overall</td>
<td>4.36</td>
<td>1.47</td>
<td>4.61</td>
<td>1.264</td>
<td>.666</td>
</tr>
</tbody>
</table>

Low BAP and High BAP average responses on each implicature type. Overall indicates the average score of a group’s response on all UI sentences.
Table A3

*Low BAP vs. High BAP Effects Between Ratings, Group, and Implicature Type*

<table>
<thead>
<tr>
<th>Truth Value</th>
<th>impType Effect</th>
<th>Group Effect</th>
<th>Interaction Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>UI</td>
<td>.598</td>
<td>.666</td>
<td>.598</td>
</tr>
<tr>
<td>F</td>
<td>&lt;.001*</td>
<td>.166</td>
<td>.198</td>
</tr>
<tr>
<td>T</td>
<td>.154</td>
<td>.901</td>
<td>.761</td>
</tr>
</tbody>
</table>

UI = underinformative; T = true; F = false; impType = implicature type.

Table A4

*Low BAPs vs. High BAPs Interaction Effects of Truth Value and Group on Ratings*

<table>
<thead>
<tr>
<th>Truth Value</th>
<th>Low BAP Mean</th>
<th>Low BAP SD</th>
<th>High BAP Mean</th>
<th>High BAP SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>UI</td>
<td>4.36 1.47</td>
<td>4.61 1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1.59 .488</td>
<td>1.92 .670</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>6.87 .211</td>
<td>6.86 .233</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Low BAP and High BAP average score on each truth condition. Neither the effect of the group ($p = .450$) nor effect of the interaction of the truth value and group ($p = .736$) were significant.

Table A5

*Low BAP vs. High BAP Non-Acceptances*

<table>
<thead>
<tr>
<th>Implicature Type</th>
<th>Low BAP Mean</th>
<th>Low BAP SD</th>
<th>High BAP Mean</th>
<th>High BAP SD</th>
<th>$p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>distributivity</td>
<td>.450</td>
<td>.482</td>
<td>.560</td>
<td>.409</td>
<td>.555</td>
</tr>
<tr>
<td>conditional</td>
<td>.400</td>
<td>.407</td>
<td>.520</td>
<td>.434</td>
<td>.482</td>
</tr>
<tr>
<td>scalar</td>
<td>.413</td>
<td>.410</td>
<td>.380</td>
<td>.333</td>
<td>.835</td>
</tr>
<tr>
<td>exhaustivity</td>
<td>.550</td>
<td>.435</td>
<td>.480</td>
<td>.434</td>
<td>.693</td>
</tr>
<tr>
<td>overall</td>
<td>.453</td>
<td>.333</td>
<td>.485</td>
<td>.301</td>
<td>.808</td>
</tr>
</tbody>
</table>

Low BAP and High BAP average tendency to reject false statements. Overall indicates the average score of a group’s response on all UI sentences.