Gender benders: The effects of individual differences in HEXACO PI-R measures on gender stereotype processing

Stephanie J. Hammond-Thrasher

Undergraduate Honours Thesis

Department of Linguistics, University of Alberta

Background

Processing language is a complex process beginning with recognizing spoken sounds or written words, and is often facilitated by abstract predictions about social and emotional content. This research examines how humans process the social properties of words embedded in sentences, and how this social content influences the perception of people in discourse. Specifically, the study investigates how individual differences in personality traits affect the perception of gender stereotypes, particularly when an individual is confronted with a clash with typical views of gender roles in society.

I presume that language comprehension draws on general cognitive processes and is not part of a modular, independent structure. That is, language processing relies on skills and knowledge also used for non-linguistic tasks (Diessel, 2019). In fact, evidence suggests that language processing is influenced by neurological systems used in perception, action, and emotion, all of which form a system-wide processing structure (Glenberg et al., 2009). Moreover, research shows that individuals process segmental input from auditory or visual language sources immediately, adapting it into a mental representation of the discourse (Canal et al., 2015). Here, sentences are processed online and in an incremental fashion, enabling individuals to form assumptions about the language they are perceiving in real time (de Hoop & Lamers, 2006).

Situation models

Evidence suggests that comprehenders reproduce a simulated, mental model of what they are processing (e.g. Zwaan & Radvansky, 1998). In fact, sentence-picture verification tasks indicate that participants base their mental situation models on inferences made about character reference, location, and orientation (Connell, 2007). For example, examine the sentences *The eagle was in the sky* and *The eagle was in the nest*. When shown drawings of an eagle with

outstretched wings versus folded wings, participants have a greater chance of reporting that the image matches the former sentence, demonstrating that comprehenders produce simulated, spatial models of sentences (Zwaan et al., 2002). Thus, participants form linguistic representations based on individual experience with and knowledge about the state of affairs in the real world (e.g. Gibbs & Perlman, 2010; Kaup et al., 2012).

As maintained by previous research, situation models are constructed according to the five key dimensions of time, space, protagonist, causation, and motivation (Zwaan et al., 1995). For the purpose of this research, however, I am most interested in changes to the mental representations of characters, particularly their genders. That is, I examine factors affecting inferences about character gender as well as what occurs when this predicted gender clashes with previously-formulated assumptions. Indeed, updating mental models can occur both incrementally, where the model is altered as new information is processed, and globally, where additional information requires a new model replacing a prior one (Kurby & Zacks, 2012). Regardless, shifts in a comprehender's perception of a character, in this case based on gender, can lead to major changes in the interpretation of a described situation, including reevaluating the entire view of the situation, hence restarting the comprehension process.

Despite the potential false flags that accompany forming incorrect linguistics predictions, inferential processing has many potential benefits for comprehension and social interaction. For example, in the sentence *The bass was strummed by the guitarist during the song*, reading times are faster for *guitarist* than if the subject is replaced with *gravedigger* (Pazynski & Kuperberg, 2012). Here, inference-based facilitation is evident, helping individuals make predictions based on semantic information encoded by nouns, allowing commitment to an expected outcome, ultimately increasing confidence and reading times. In fact, inference formation is influenced

both by semantic information and general world knowledge, including emotional and social information evoked by verbs and nouns (Hagoort et al., 2004; Van Berkum, 2018).

Stereotypes

Inference formation during discourse processing relies heavily on stereotype information (Molinaro et al., 2016). One aspect of stereotyping that strongly influences inference is gender. Gender is often encoded socially in a role name, such as *Nurse* or *Doctor*, which stereotypically correspond to females and males, respectively. These gender stereotypes often arise from statistically-driven probabilities, as the strength of association between role names in corpus data positively correlates with the percentages of female employees in those occupations based on census data (Caliskan et al., 2017). Evidence also shows that individuals automatically make inferences regarding character gender upon the presentation of stereotypical gender roles. A study by Oakhill et al. (2005) asked participants to say whether two terms like *Uncle* (definitionally male) and *Nurse* (stereotypically female) could refer to the same person. Here, response times were inhibited when the gender of the stereotype did not match the definitional gender of the paired role name, such as in the example above. This suggests that the gender of a character is incorporated into the mental representation of text even in cases when it is not explicitly stated.

It is vital to note that gender-related inferencing in English has been widely studied through the use of anaphora because it prompts readers to link gender marked pronouns to antecedents in the previous sentential context (Marrville, 2017). For example, a sentence like *The nurse yawned because she/he was tired* is read faster when the pronoun is displayed as *she* than when the pronoun is replaced by *he* (Banaji & Hardin, 1996). That is, this sentential setup forces readers and/or listeners to confront the explicit gender of the role noun, in this case *nurse*, rather

than relying solely upon stereotype assumptions. Thus, the added processing difficulty that occurs when the pronoun gender is incongruent with the stereotype gender manifests through longer reading times.

Individual differences

Because semantic information is closely related to stereotypes, I presume that participant worldview and personality are key factors in social language processing. In line with this, an EEG study by Van Berkum et al. (2009) found that sentences that directly clashed with a participant's value system required additional processing resources as soon as participants encountered a value-inconsistent target word. In the study, participants completed a questionnaire outlining their opinions on socially- and politically- controversial topics such as euthanasia and abortion. Following this, participants read opinionated statements about similar topics. The results concluded that strong value-inconsistent statements took longer to read and elicited both a larger positive ERP effect at 250 ms from the target word as well as a significant negative going wave peaking around 400ms from the target word (N400). For example, participants read a sentence like *Watching TV to relax is fine in my opinion*, where *fine* immediately increases the sentence's processing load for individuals whose value system clashes with the statement, meaning that they take longer to read the sentence. This highlights a link between the neural systems for language and emotion-based valuation and implicates political ideology in social language processing.

Recent evidence shows that stereotype processing is influenced by more than statistically-driven probabilities. In fact, factors such as personality and worldview are implicated in social language processing. A study by Van den Brink et al. (2012) demonstrates that sociocultural violations arising from inconsistent speaker identity and statement type drive

large positive event-related potential (ERP) components that indicate difficulty with semantic processing. For example, larger N400 effects were seen when participants heard the sentence *I cannot sleep without my **teddy bear** in my arms* spoken by an adult male compared to a male child. Participants in this study also completed Baron-Cohen and Wheelwright's (2004) Empathy Quotient Questionnaire (EQQ), where a better ability to empathize predicted greater semantic processing difficulties with the incongruent speaker identity items.

Similarly, Hubert-Lyall and Järvikivi (2020) established a link between individual differences in personality traits and social language processing. Using sentences containing socio-cultural clashes, such as hearing *I sometimes buy my **bras** at Hudson's Bay* spoken by a mature male rather than a mature female, the authors determined that more introverted individuals showed greater pupil dilation when compared to sentences without any clashes. Indeed, larger pupil dilation is associated with greater cognitive and emotional strain. Hence, it is expected that sentences with social information in the form of gender stereotypes would elicit an effect, likely through changes in pupil dilation, reaction times to stimuli, or lower ratings of appropriateness when there is a clash between an individual's stereotype expectations and the reality of the sentence itself.

Thus, examining social language processing necessitates a simultaneous consideration of the individual differences that influence a person's comprehension capacity. Because of the previously discussed implications of personality traits on social language comprehension, I consider personality factors as vital influences for the purpose of this study. That is, research suggests that individuals with different personality profiles exhibit different patterns of the allocation of cognitive and emotional resources during real-time language comprehension. However, there is no systematic research examining the effects of personality traits on the

comprehension of sentences containing explicit gender stereotypes, which are evidently frequent social phenomena with heavy implications on language comprehension.

Current Study

This study investigates the extent to which individual differences in personality traits contribute to gender stereotype processing. That is, participants rated sentences containing gender stereotyped role names for correctness and appropriateness, allowing me to outline differing reactions to stereotype clashes and stereotype compliance. Above, I reviewed evidence that inference formation is influenced by a wide variety of variables, including personality traits and gender stereotype knowledge, yet the relationships between these factors has not been simultaneously and concretely investigated. That is, to increase confidence in the understudied interaction between personality traits and social language processing, systematic, experimental evidence establishing a link between a valid personality measure and the overwhelmingly common phenomenon of gender stereotypes is needed.

Previous research utilizes personality questionnaires such as the NEO Five-Factor Inventory (NEO FFI) (Costa & McCrae, 1992 & Hubert-Lyall & Järvikivi (2020)) and the Empathy Quotient (EQQ) (Baron-Cohen & Wheelwright, 2004) to operationalize personality traits. However, strong critiques of the NEO FFI indicate that both the methodology and traits measured fall prey to questionable assumptions (Block, 1995). That is, the five factors of Extraversion, Openness to Experience, Agreeableness, Conscientiousness, and Neuroticism are operationalized in a way that only accounts for some traits within the scope of human personality. Furthermore, the model is based on lexical hypothesis, which relies on the verbal descriptions of individual differences, thus confounding results with individuals' own pro-social biases. Additionally, while EQQ scores indicate validity in relation to an individual's ability to

empathize, this questionnaire only evaluates a single personality trait. For the purpose of this study, I utilize Ashton and Lee's (2009) HEXACO Personality Inventory-Revised (HEXACO PI-R), which analyzes personality traits within the following six dimensions:

(1) Honesty-Humility, where individuals with high scores avoid manipulating others for personal gain, feel little temptation to break rules, and feel no special entitlement to elevated social status;

(2) Emotionality, where individuals with high scores experience fear and anxiety frequently and feel a need for emotional support from others;

(3) Extraversion, where individuals with high scores feel confident when addressing groups, enjoy social interactions, and withhold judgement of others;

(4) Agreeableness, where individuals with high scores forgive others easily and are able to control their temper;

(5) Conscientiousness, where individuals with high scores have strong organizational skills and strive for perfection in their tasks; and

(6) Openness to Experience, where individuals with high scores are inquisitive, take an interest in unusual ideas or people, and are generally more open to novel situations.

The questionnaire uses a five-point Likert scale to evaluate individuals' responses to 60 different theoretical situations. Hence, the questionnaire avoids the weaknesses of the NEO FFI and EQQ scale, surveying a full spectrum of traits without requisite lexical hypotheses. Regarding gender stereotype processing, I am most interested in traits (3) and (6), as both specifically pertain to an individual's ability to interact socially.

Hence, in this study I propose to answer the following questions:

(1) Is there a difference between Correctness and Appropriateness ratings of sentences containing gender stereotype violations?

(2) Are anaphoric sentences with Incongruent stereotype and pronoun genders rated differently on an Appropriateness scale when compared to sentences with Congruent stereotype and pronoun genders?

(3) Is there a difference between Appropriateness ratings of Female stereotype violations versus Male stereotype violations (i.e., when a Female role noun is Incongruent with a Male pronoun versus when a Male role noun is Incongruent with a Female pronoun)?

(4) To what extent do individual differences in HEXACO PI-R scores affect ratings of Appropriateness of sentences containing gender stereotyped role nouns?

To explore possible answers to these questions, I performed a rating experiment that examined participants' evaluations of anaphoric sentences on two five-point Likert scales, one asking how Correct they judged the sentence to be and the other asking how Appropriate they judged the sentence to be. For example, participants saw sentences like *The receptionist answered the phone because she heard it ring*, and *The engineer treated the patient because she identified several wounds,* and rated them on both scales. In the first sentence, the gender of the stereotype, *receptionist*, is Congruent with the pronoun, *she*; in the second sentence, the gender of the stereotype, *engineer*, is Incongruent with the pronoun, *she*. No definition of Correctness or Appropriateness was given; I expected participants to view Correctness as a measure of the absence of linguistic errors and Appropriateness as a subjective measure of their own acceptance of the sentence. After rating 32 sentences, participants completed the HEXACO PI-R.

Firstly, I anticipated that Correctness ratings would be high across all items, as the sentences did not contain any linguistics errors. Secondly, I anticipated Incongruence between the stereotype and pronoun genders would correlate with lower Appropriateness ratings overall when compared to Congruent sentences. Thirdly, I predicted that the results might indicate a difference between Appropriateness ratings of Female stereotype violations versus Male stereotype violations, as society inherently encodes a gender hierarchy. Finally, I anticipated that a more Open and Extraverted outlook would correlate with higher Appropriateness ratings of Incongruent sentences overall.

Methodology

Participants

Ethics for this study were approved by the Research Ethics Office at the University of Alberta (PRO00089925) (Appendix A). 91 undergraduate students from introductory Linguistics classes at the University of Alberta participated in the study (67 female, 24 male). All individuals received course credit for their participation. All participants were native speakers of Canadian English with normal or corrected-to-normal vision.

Materials

Role nouns for this study were gathered from a project by Marrville (2017), which cited occupational stereotype norms created by Gygax and Gabriel (2008) and Carreiras et al. (1996). Each sentence (Table I) began with a role noun embedded in a noun phrase (Segment 1) and a verb (Segment 2), followed by a noun phrase and the causal continuation *because* (Segment 3). This separated the following pronoun segment (Segment 4), which either matched or mismatched the stereotypical gender of the role noun, and concluded with an embedded verb phrase. To avoid unintentional ambiguity, each sentence contained only one agent. Filler

sentences followed the same setup, but role nouns were used that did not have any gender stereotype associated with them (Table II). There were no sentences containing semantic or syntactic errors.

Table I: Example of Target Sentence Construction

Segment 1	Segment 2	Segment 3	Segment 4
The receptionist	answered	the phone because	she/he heard it ring.

Note. Item inventories can be found in Appendix B.

Table II: Example of Filler Sentence Construction

Segment 1	Segment 2	Segment 3	Segment 4
The pedestrian	crossed	the road because	she/he heard it ring.

Note. Item inventories can be found in Appendix B.

Procedure

Using Google Forms to ease study participation during the COVID-19 pandemic, each participant saw 32 sentences, eight of which were filler sentences. Of the remaining 24 sentences, 12 were Congruent and 12 were Incongruent. Participants did not read sentences that repeated a role noun-verb pairing to avoid effects from item similarity. Additionally, items were counterbalanced such that each participant was exposed to an equal number of Female stereotypes and Male stereotypes, and thus an equal number of Female and Male pronouns. For each sentence, including fillers, participants rated their Correctness and Appropriateness on two five-point Likert scales. After the completion of the experimental block, participants completed the HEXACO PI-R 60-item questionnaire (Appendix C). Both questionnaires were completed

using Google Forms. Finally, participants completed a language background questionnaire (Appendix D) to ensure their status as native speakers of Canadian English.

Results

Before the analysis, filler items were removed. The results were analysed using linear mixed effects regression models using the package *lme4* (version 1.1-23, Bates et al., 2015) in the R statistical environment (R Core Team, 2019). Rating (on a Likert scale from 1 to 5) was the dependent variable and the fixed effects variables of interest were as follows:

(1) Stereotype Congruence (whether the pronoun gender was congruent versus incongruent with the stereotype gender);

(2) Stereotype Gender (female versus male);

(3) Rating Type (Correctness versus Appropriateness)

(4) HEXACO PI-R Scores, particularly Openness to Experience and Extraversion;

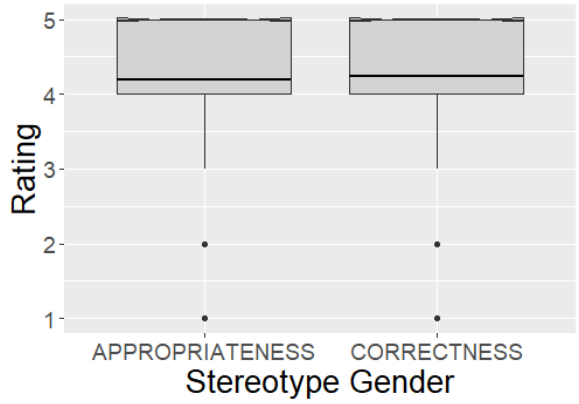
Starting with a maximal model, I used a backward fitting procedure to investigate the contribution of 1-3 to the model fit. I continued by removing non-significant interactions and/or predictors from the model, comparing subsequent models' fit to the data with function anova(). After we arrived at the best model for the manipulated variables, by-subject and by-item random slopes were added (as far as the models converged). HEXACO trait scores were then forward fitted, assessing their contribution to the model fit at each step. In addition to the reported fixed effects below, the final model included by-subject random slopes for Stereotype Congruence and Stereotype Gender.

Correlating rating results with scores from the HEXACO PI-R, we found five key results, for which statistical analyses and results can be found in Appendix E. Firstly, Correctness ratings

were higher overall for all sentences ($t = 2.2077$, $p < 0.05$) when compared to Appropriateness ratings (Figure I).

Figure 1

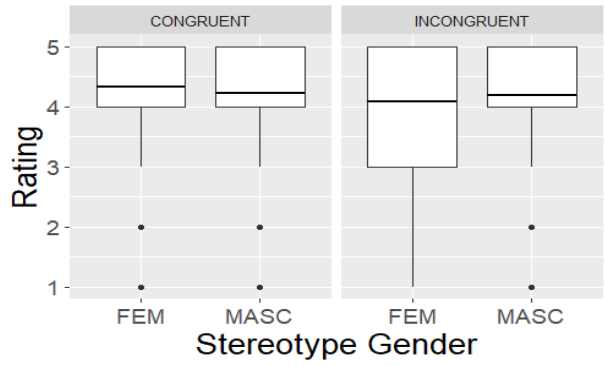
Participant Ratings of Appropriateness versus Correctness



Secondly, Incongruent sentences were rated slightly lower in Appropriateness ($t = -4.8746$, $p < 0.001$) when compared to Appropriateness ratings of Congruent sentences (Figure II). Importantly, the third result indicates an interaction between Stereotype Gender and Congruence. That is, Incongruent sentences were correlated with significantly lower Appropriateness ratings when the Stereotype Gender was Female ($t = 3.2172$, $p < 0.001$), but the effect of Incongruence was significantly less pronounced when the stereotype gender was Male (Figure II).

Figure II

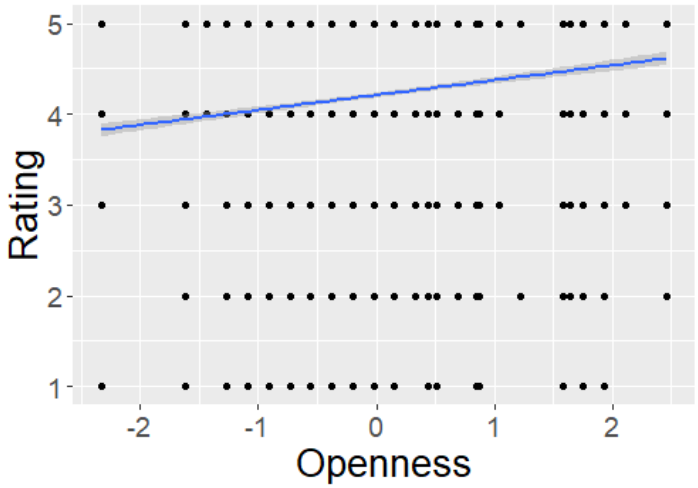
Participant Ratings of Congruent versus Incongruent Sentences by Stereotype Gender



The fourth result indicates a correlation between individual differences in Openness to Experience scores and Appropriateness ratings. That is, higher Openness to Experience scores correlated with higher Appropriateness ratings ($t = 3.0112$, $p < 0.01$) when compared to individuals with lower Openness to Experience scores (Figure III).

Figure III

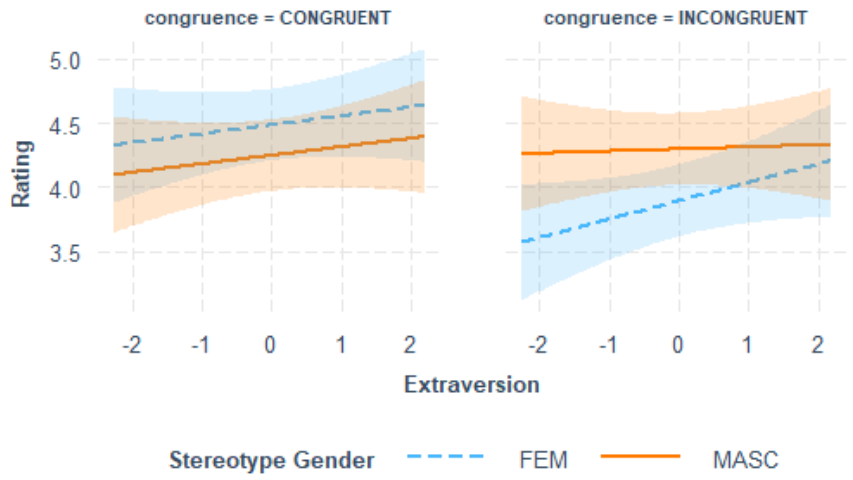
Individual Differences in Openness to Experience Scores by Ratings of Appropriateness



Finally, the fifth result indicates a three-way interaction between stereotype gender, stereotype Congruence, and individual differences in Extraversion scores. That is, Incongruent, Female stereotypes were rated significantly lower in Appropriateness by participants with low Extraversion scores ($t = -2.148$, $p < 0.05$) when compared to Incongruent, Male stereotypes. Interestingly, this effect did not hold for Congruent sentences (Figure IV). There were no main effects or interactions between other HEXACO PI-R scores and manipulated variables.

Figure IV

Individual Differences in Extraversion Scores by Ratings of Appropriateness According to Congruent versus Incongruent Conditions



Note. Stereotype Gender is indicated by dashed versus smooth lines.

Discussion

The results of this rating study indicate not only that reactions to gender stereotype clashes affect real-time language processing, but also that individual differences in certain personality measures mediate these effects quite noticeably. Indeed, the allocation of cognitive and emotional resources for language comprehension appears to depend on an individual's personality profile. Moreover, the degree to which a clash in stereotype gender affects language processing seems to also depend on the typical gender of the stereotype itself. This suggests that certain gender roles are more cemented than others and hence have more influence on social interactions. Below, I will discuss in detail the implications of the different outcomes from this project, along with their repercussions for language processing in a social context.

Stereotype Violations

It does not seem that individuals view stereotype violations as a significant factor when forming judgements about the correctness of sentences. Indeed, results from this study indicate

that gender stereotype violations only affect judgements of sentence appropriateness. This indicates that a gender stereotype clash is not considered an error in the same way that syntactic errors are considered completely incorrect; stereotype violations are accepted as correct regardless of the sentential context, but not necessarily as socially appropriate. This result reflects previous findings, where syntactic and semantic violations elicit larger processing effects when compared to socio-cultural clashes; yet, these socio-cultural violations still elicit more effort than sentences without clashes (Hubert-Lyall & Järvikivi, 2020). Thus, gender stereotype violations are deemed correct despite their context, yet individuals may question their overall degree of appropriateness.

Furthermore, the significant interaction between stereotype gender and congruence suggests that female gender roles are more mentally, and thus more socially, cemented when compared to male gender roles. That is, female stereotype violations correlated with lower ratings of sentence appropriateness than male stereotype violations. This indicates that the gender of a particular stereotype mediates the degree of surprise that individuals experience when encountering a stereotype clash. In regards to this study, this manifested as lower appropriateness ratings; however, according to previous research, this affects the time it takes to process stereotype clashes as well as the usage of cognitive and emotional language processing resources (Van den Brink et al., 2012). Thus, the increased social weight of female gender roles is implicated in real-time language processing speed and capabilities, regardless of the specific role name in question.

Personality Traits

The results of this study implicate individual differences across two personality dimensions in social language processing. Firstly, it appears that people who are more open to

experience are less surprised by gender stereotype violations than people who are less open to experience, as demonstrated by the correlation between high Openness to Experience scores with higher ratings of appropriateness. Interestingly, this suggests that individuals who are more adventurous and imaginative are more likely to accept language inputs that clash with common social views of gender roles. Indeed, this increases confidence in the previous finding that an individual's own value system influences how they process social statements (Van Berkum et al., 2009). That is, this project suggests that individuals who are curious about their environment and wish to experience novel situations process clashes with common social knowledge with more ease. Emphatically, this highlights that individuals who are closed off and who lack the spirit of inquiry may be more prone to embracing gender stereotypes and thus may be less accepting of non-conformity and even discrimination. In regards to language processing, individuals who are less open to experience certainly appear to question clashes with gender stereotypes when compared to those who are more open-minded.

Finally, the three-way interaction between stereotype gender, stereotype congruence, and individual differences in Extraversion scores indicates that individuals who are more introverted tend to hold more cemented views of female gender roles. That is, individuals who are more extraverted and thus more enthusiastic about social interactions were more likely to accept female gender stereotype violations as appropriate when compared to those possessing less extraverted traits. Importantly, this result suggests that extraverts excel at expressing social acceptance, as gender stereotypes hold less influence on their comprehension and judgements of clashes with typical female gender roles. This result is the first of its kind, emphasizing both the increased strength of female gender stereotypes compared to male gender stereotypes as well as the inability of introverted individuals to think critically about non-conformity within society.

Future Research

Evidently, the understudied relationship between gender stereotype processing and individual personality profiles requires future research to increase confidence in the interactions identified in this paper. There are several innovative directions possible for these studies. Firstly, the correlation between HEXACO PI-R traits and individual political views may yield interesting implications regarding links between moral and political decision making and their effects on language comprehension. That is, using a political ideology questionnaire, such as one created by Grenier et al. (n.d.), in a similar experimental paradigm will further illuminate the relationship between individual differences and social language processing.

Furthermore, this concept of individual personality and affective profiling will be strengthened by the use of a more subconscious measure, such as the Disgust Scale - Revised (Haidt et al., 2012). This questionnaire assesses the strength of an individual's behavioural immune system, which relates to how strongly an individual reacts to uncommon situations with disgust. For example, when an individual reads the statement *It would not upset me at all to watch a person with a glass eye take the eye out of the socket*, a high rating of immediate disgust indicates a weaker behavioural immune response, as the individual is unable to process the statement without a conflicting emotional reaction. This automatic measure of an individual's emotional decision making is an important component of individual differences, as it illuminates a key social language processing mechanism which likely has implications on gender stereotype comprehension. Hence, a similar experimental paradigm that includes the Disgust Scale - Revised will explicate this relationship. Finally, future research occurring in a time without COVID-19 in-person research restrictions will necessitate a pupillometry or self-paced reading or listening paradigm in order to investigate the time course of these effects.

References

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*, 340-345.

Banaji, M. R. & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163-175.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software, 67*(1), 1–48.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. Psychological Bulletin, 117(2), 187–215. https://doi.org/10.1037/0033-2909.117.2.187

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora necessarily contain human biases. *Science, 356*, 183–186.

Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: Self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology, 6.*

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotype gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology: Section A, 49*(3), 639–663.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*(3), 476–485.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and*

the NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.

de Hoop, H. & Lamers, M. (2006). Incremental distinguishability of subject and object. *Case, Valency and Transitivity*, 269–287.

Diessel, H. (2019). *Cognitive Processes and Language Use. In The Grammar Network: How Linguistic Structure Is Shaped by Language Use* (pp. 23-40). Cambridge: Cambridge University Press.

Gibbs R. W. Jr. & Perlman, M. (2010). Language understanding is grounded in experiential simulations: A response to Weiskopf. S*tudies in History and Philosophy of Science Part A, 41*(3), 305-308

Glenberg, A. M., Webster, B. J., Mouilso, E., Havas, D., & Lindeman, L. M. (2009). Gender, emotion, and the embodiment of language comprehension. *Emotion Review, 1*(2), 151–161.

Gygax, P. M. & Gabriel, U. (2008). Can a group of musicians be composed of women? Generic interpretation of French masculine role names in the absence and presence of feminine forms. *Swiss Journal of Psychology, 67*(3), 143–151.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441.

Hubert Lyall, I. & Järvikivi, J. (2020). Listener's Personality Traits Predict Changes in Pupil Size During Auditory Language Comprehension. Submitted for publication.

Kaup, B., Zwaan, R. A., & Lüdtke, J. (2007). The experiential view of language comprehension: How is negation represented? In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (p. 255–288).

Lawrence Erlbaum Associates Publishers.

Kurby, C. A. & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition, 40*(5), 812–826.

Marrville, C. (2017). *Gender and dominance in action: World view and emotional affect in language processing and use*. Department of Linguistics, University of Alberta.

Molinaro, N., Su, J. J., & Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language, 155-156,* 36–43.

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition, 33*(6), 972–983.

Paczynski, M. & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language, 67*, 426–448.

R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Van Berkum, J. J. (2018). Language comprehension, emotion, and sociality: Aren't we missing something? In S. A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics* (2nd edn). OUP.

Van Berkum, J. J. A., Holleman, B., Nieuwland, M. S., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science, 20*(9), 1092–1099.

Van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M.,

Buitelaar, J. K., & Hagoort, P. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social cognitive and affective neuroscience, 7*(2), 173-183.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 386–397.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological science, 13*(2), 168-171.

Appendix A

Figure VI

Letter of Ethics Approval

UNIVERSITY OF ALBERTA

RESEARCH ETHICS OFFICE

308 Campus Tower
Edmonton, AB, Canada T6G 1K8
Tel: 780.492.0459
uab.ca/reo

Notification of Approval

Date:	October 21, 2019	
Study ID:	Pro00089925	
Principal Investigator:	Stephanie Hammond-Thrasher	
Study Supervisor:	Juhani Järvikivi	
Study Title:	Gender benders: Stereotypes, synthesized speech, and inferential processing	
Approval Expiry Date:	Monday, October 19, 2020	
Approved Consent Form:	Approval Date 10/21/2019	Approved Document Consent Form .pdf

Thank you for submitting the above study to the Research Ethics Board 2. Your application has received a delegated review and been approved on behalf of the committee.

Any proposed changes to the study must be submitted to the REB for approval prior to implementation. A renewa report must be submitted next year prior to the expiry of this approval if your study still requires ethics approval. If you do not renew on or before the renewal expiry date, you will have to re-submit an ethics application.

Approval by the Research Ethics Board does not encompass authorization to access the staff, students, facilities or resources of local institutions for the purposes of the research.

Sincerely,

Ubaka Ogbogu, LLB, BL, LLM, SJD
Chair, Research Ethics Board 2

Note: This correspondence includes an electronic signature (validation and approval via an online system).

Appendix B

Table III

Items for List 1

Sentence	Type	Stereotype Gender	Congruence	Item Number
The housekeeper washed the sheets because she noticed they were dirty	Stereotyped	F	congruent	01
The florist clipped the flowers because he saw they had grown	Stereotyped	F	incongruent	02
The beautician applied the eyeliner because she thought it was pretty	Stereotyped	F	congruent	03
The receptionist answered the phone because he heard it ringing	Stereotyped	F	incongruent	04
The dancer waved at the crowd because she was proud of the performance	Stereotyped	F	congruent	05
The nanny drove the van because he was taking the kids out	Stereotyped	F	incongruent	06
The model departed from the house because she was going to a meeting	Stereotyped	F	congruent	07
The prostitute ran to the curb because he had a client	Stereotyped	F	incongruent	08
The librarian flipped through the book because she was looking for a certain page	Stereotyped	F	congruent	09
The dietician recommended a supplement because he knew it was good for health	Stereotyped	F	incongruent	10
The opera singer practiced scales because she had a performance that night	Stereotyped	F	congruent	11
The maid left early because he finished the task	Stereotyped	F	incongruent	12
The cheerleader messed up because she was distracted by a friend	Stereotyped	F	congruent	13

The nurse took a long nap because he had worked a long shift	Stereotyped	F	incongruent	14
The secretary searched for a pen because she needed to write a note	Stereotyped	F	congruent	15
The babysitter abandoned the kids to bed because he needed to take a call	Stereotyped	F	incongruent	16
The boss paid with the company card because he was entertaining a client	Stereotyped	M	congruent	17
The butcher wrapped the meat because she was serving a customer	Stereotyped	M	incongruent	18
The carpenter picked up the hammer because he needed to drive in nails	Stereotyped	M	congruent	19
The farmer cooked a big feast because she was entertaining the neighbours	Stereotyped	M	incongruent	20
The golfer threw the club because he was frustrated	Stereotyped	M	congruent	21
The judge ordered the maximum sentence because she strived for justice	Stereotyped	M	incongruent	22
The plumber dislodged the obstruction because he used a strong solvent to loosen it	Stereotyped	M	congruent	23
The soldier saluted the officers because she believed in the freedom they fought for	Stereotyped	M	incongruent	24
The firefighter ran into the building because he wanted to save the dog	Stereotyped	M	congruent	25
The pilot announced the landing because she was approaching the destination	Stereotyped	M	incongruent	26
The engineer went over the plans because he maintained safety standards	Stereotyped	M	congruent	27
The sailor raised the sail because she was departing on a trip	Stereotyped	M	incongruent	28
The surgeon closed the incision because he had completed the surgery	Stereotyped	M	congruent	29

The president smiled at the crowd because she wanted to appear friendly	Stereotyped	M	incongruent	30
The police officer examined the case file because he suspected foul play	Stereotyped	M	congruent	31
The truck driver swerved off the road because she saw a deer	Stereotyped	M	incongruent	32
The pedestrian looked both ways because she was about to cross the street	Filler	N/A	N/A	33
The concert goer drank some water because he was thirsty	Filler	N/A	N/A	34
The spectator sat down because she was about to watch the game	Filler	N/A	N/A	35
The neighbour called the police because he heard a loud noise	Filler	N/A	N/A	36
The student arrived late because she had missed the bus	Filler	N/A	N/A	37
The musician tuned the guitar because he noticed it was out of tune	Filler	N/A	N/A	38
The individual looked at the book cover because she wanted to read it	Filler	N/A	N/A	39
The person laughed at the joke because he thought it was funny	Filler	N/A	N/A	40

Note. Half of the participants were randomly assigned to List 1. Sentences were presented in a randomized order. Each sentence was presented with two Likert scales from one to five, one asking for a rating of Correctness and the other for a rating of Appropriateness.

Table IV

Items for List 2

Sentence	Type	Stereotype Gender		
The housekeeper washed the sheets because he noticed they were dirty	Stereotyped	F	incongruent	01
The florist clipped the flowers because she saw they had grown	Stereotyped	F	congruent	02
The beautician applied the eyeliner because he thought it was pretty	Stereotyped	F	incongruent	03
The receptionist answered the phone because she heard it ringing	Stereotyped	F	congruent	04
The dancer waved at the crowd because he was proud of the performance	Stereotyped	F	congruent	05
The nanny drove the van because she was taking the kids out	Stereotyped	F	congruent	06
The model departed from the house because he was going to a meeting	Stereotyped	F	incongruent	07
The prostitute ran to the curb because she had a client	Stereotyped	F	congruent	08
The librarian flipped through the book because he was looking for a certain page	Stereotyped	F	incongruent	09
The dietician recommended a supplement because she knew it was good for health	Stereotyped	F	congruent	10
The opera singer practiced scales because he had a performance that night	Stereotyped	F	incongruent	11
The maid left early because she finished the task	Stereotyped	F	congruent	12
The cheerleader messed up because he was distracted by a friend	Stereotyped	F	incongruent	13
The nurse took a long nap because she had worked a long shift	Stereotyped	F	congruent	14

The secretary searched for a pen because he needed to write a note	Stereotyped	F	incongruent	15
The babysitter abandoned the kids because she needed to take a call	Stereotyped	F	congruent	16
The boss paid with the company card because she was entertaining a client	Stereotyped	M	incongruent	17
The butcher wrapped the meat because he was serving a customer	Stereotyped	M	congruent	18
The carpenter picked up the hammer because she needed to drive in nails	Stereotyped	M	incongruent	19
The farmer cooked a big feast because he was entertaining the neighbours	Stereotyped	M	congruent	20
The golfer threw the club because she was frustrated	Stereotyped	M	incongruent	21
The judge ordered the maximum sentence because he strived for justice	Stereotyped	M	congruent	22
The plumber dislodged the obstruction because she used a strong solvent to loosen it	Stereotyped	M	incongruent	23
The soldier saluted the officers because he believed in the freedom they fought for	Stereotyped	M	congruent	24
The firefighter ran into the building because she wanted to save the dog	Stereotyped	M	incongruent	25
The pilot announced the landing because he was approaching the destination	Stereotyped	M	congruent	26
The engineer went over the plans because she maintained safety standards	Stereotyped	M	incongruent	27
The sailor raised the sail because he was departing on a trip	Stereotyped	M	congruent	28
The surgeon closed the incision because she had completed the surgery	Stereotyped	M	incongruent	29
The president smiled at the crowd because he wanted to appear friendly	Stereotyped	M	congruent	30
The police officer examined the case	Stereotyped	M	incongruent	31

file because she suspected foul play				
The truck driver swerved off the road because he saw a deer	Stereotyped	M	congruent	32
The pedestrian looked both ways because he was about to cross the street	Filler	N/A	N/A	33
The concert goer drank some water because she was thirsty	Filler	N/A	N/A	34
The spectator sat down because he was about to watch the game	Filler	N/A	N/A	35
The neighbour called the police because she heard a loud noise	Filler	N/A	N/A	36
The student arrived late because he had missed the bus	Filler	N/A	N/A	37
The musician tuned the guitar because she noticed it was out of tune	Filler	N/A	N/A	38
The individual looked at the book cover because he wanted to read it	Filler	N/A	N/A	39
The person laughed at the joke because she thought it was funny	Filler	N/A	N/A	40

Note. The remaining half of the participants were randomly assigned to List 2. Sentences were presented in a randomized order. Each sentence was presented with two Likert scales from one to five, one asking for a rating of Correctness and the other for a rating of Appropriateness.

Appendix C

Figure VII

Language Background Questionnaire

Directions:

Below you will answer a series of questions about your language background and experience. Please try to respond to each question appropriately, even if you are not completely sure of your response.

Questionnaire:

1. What is your gender? (select one of the following)
    a. Male
    b. Female
    c. Non-binary
    d. Prefer not to say
2. What is your age? _____
3. Do you consider yourself a native speaker of English? (if no, please specify your native language)
    a. Yes
    b. No _____
4. How would you rate your English reading proficiency on a scale from one to five, where one is Beginner and five is Expert? _____
5. How would you rate your English writing proficiency on a scale from one to five, where one is Beginner and five is Expert? _____

Note. Only native speakers of English were included in analysis, specifically those with English reading and writing proficiency ratings between four and five.

Appendix D

Figure VIII

HEXACO PI-R 60-Item Self Report Form

Directions:

Below, you will find a series of statements about you. Please read each statement and decide how much you agree or disagree with that statement, filling in the blank with the appropriate number based on the following scale:

5 = strongly agree
4 = agree
3 = neutral (neither agree nor disagree)
2 = disagree
1 = strongly disagree

Please answer every statement, even if you are not completely sure of your response.

Questionnaire:

1. I would be quite bored by a visit to an art gallery.
2. I plan ahead and organize things, to avoid scrambling at the last minute.
3. I rarely hold a grudge, even against people who have badly wronged me.
4. I feel reasonably satisfied with myself overall.
5. I would feel afraid if I had to travel in bad weather conditions.
6. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.
7. I'm interested in learning about the history and politics of other countries.
8. I often push myself very hard when trying to achieve a goal.
9. People sometimes tell me that I am too critical of others.
10. I rarely express my opinions in group meetings.
11. I sometimes can't help worrying about little things.
12. If I knew that I could never get caught, I would be willing to steal a million dollars.
13. I would enjoy creating a work of art, such as a novel, a song, or a painting.
14. When working on something, I don't pay much attention to small details.
15. People sometimes tell me that I'm too stubborn.
16. I prefer jobs that involve active social interaction to those that involve working alone.
17. When I suffer from a painful experience, I need someone to make me feel comfortable.
18. Having a lot of money is not especially important to me.
19. I think that paying attention to radical ideas is a waste of time.
20. I make decisions based on the feeling of the moment rather than on careful thought.
21. People think of me as someone who has a quick temper.
22. On most days, I feel cheerful and optimistic.
23. I feel like crying when I see other people crying.

24. I think that I am entitled to more respect than the average person is.
25. If I had the opportunity, I would like to attend a classical music concert.
26. When working, I sometimes have difficulties due to being disorganized.
27. My attitude toward people who have treated me badly is “forgive and forget”.
28. I feel that I am an unpopular person.
29. When it comes to physical danger, I am very fearful.
30. If I want something from someone, I will laugh at that person's worst jokes.
31. I’ve never really enjoyed looking through an encyclopedia.
32. I do only the minimum amount of work needed to get by.
33. I tend to be lenient in judging other people.
34. In social situations, I’m usually the one who makes the first move.
35. I worry a lot less than most people do.
36. I would never accept a bribe, even if it were very large.
37. People have often told me that I have a good imagination.
38. I always try to be accurate in my work, even at the expense of time.
39. I am usually quite flexible in my opinions when people disagree with me.
40. The first thing that I always do in a new place is to make friends.
41. I can handle difficult situations without needing emotional support from anyone else.
42. I would get a lot of pleasure from owning expensive luxury goods.
43. I like people who have unconventional views.
44. I make a lot of mistakes because I don’t think before I act.
45. Most people tend to get angry more quickly than I do.
46. Most people are more upbeat and dynamic than I generally am.
47. I feel strong emotions when someone close to me is going away for a long time.
48. I want people to know that I am an important person of high status.
49. I don’t think of myself as the artistic or creative type.
50. People often call me a perfectionist.
51. Even when people make a lot of mistakes, I rarely say anything negative.
52. I sometimes feel that I am a worthless person.
53. Even in an emergency I wouldn’t feel like panicking.
54. I wouldn’t pretend to like someone just to get that person to do favors for me.
55. I find it boring to discuss philosophy.
56. I prefer to do whatever comes to mind, rather than stick to a plan.
57. When people tell me that I’m wrong, my first reaction is to argue with them.
58. When I’m in a group of people, I’m often the one who speaks on behalf of the group.
59. I remain unemotional even in situations where most people get very sentimental.
60. I’d be tempted to use counterfeit money, if I were sure I could get away with it.

Figure IX

HEXACO PI-R 60-Item Scoring Key

Individual scores should be computed as numbers between one and five. Questions marked with R are reverse-keyed items; for these statements, responses should be reversed prior to computing scores. For example, a score of five on a reverse-keyed statement becomes a one when keyed correctly.

The following breakdown indicates to which factor each statement correlates:

1. Honesty-Humility: 6, 12R, 18, 24R, 30R, 36, 42R, 48R, 54, 60R
2. Emotionality: 5, 11, 17, 23, 29, 35R, 41R, 47, 53R, 59R
3. Extraversion: 4, 10R, 16, 22, 28R, 34, 40, 46R, 52R, 58
4. Agreeableness: 3, 9R, 15R, 21R, 27, 33, 39, 45, 51, 57R
5. Conscientiousness: 2, 8, 14R, 20R, 26R, 32R, 38, 44R, 50, 56R
6. Openness to Experience: 1R, 7, 13, 19R, 25, 31R, 37, 43, 49R, 55R

Factor scale scores should be computed as means across all items within the factor.

Appendix E

Table V

Statistical Analyses of All Variables with Openness to Experience

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	3.3764	0.3181	101.9928	10.6156	0.000	***
congruenceINCONGRUENT	-0.2035	0.0417	178.4934	-4.8746	0.000	***
genderMASC	-0.0677	0.1238	34.1922	-0.5468	0.588	
ratingtypeCORRECTNESS	0.0499	0.0226	5506.495	2.2077	0.0273	*
openness	0.271	0.09	89.3207	3.0112	0.0034	**
congruenceINCONGRUENT:genderMASC	0.147	0.0457	5513.959	3.2172	0.0013	**

Note. $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***

Table VI

Statistical Analysis of All Variables with Extraversion

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	4.418	0.142	48.459	31.076	0.000	***
congruenceINCONGRUENT	-0.591	0.176	31.196	-3.353	0.002	**
genderMASC	-0.238	0.173	29.111	-1.373	0.18	
extrav	0.07	0.075	51.754	0.938	0.353	
ratingtypeCORRECTNESS	0.041	0.029	3069.5	1.427	0.154	
openness	0.185	0.066	48.008	2.81	0.007	**
congruenceINCONGRUENT:genderMASC	0.638	0.242	28.005	2.635	0.014	*

congruenceINCONGRUENT:extrav	0.074	0.058	86.269	1.283	0.203	
genderMASC:extrav	-0.004	0.048	121.96	-0.074	0.941	
congruenceINCONGRUENT:genderMASC:extrav	-0.124	0.058	3069.5	-2.148	0.032	*

Note. $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***