Reading is almost entirely a solitary act, and yet there is abundant evidence suggesting that reading, particularly reading fiction, has positive consequences for social cognition. For example, reading fiction improves theory of mind (Dodell-Feder & Tamir, 2018; Mar et al., 2006; Mar, Oatley, & Peterson, 2009) and invites pro-social behaviors (Koopman, 2015). Studies manipulating reading exposure have yielded mixed results on the effects of fiction reading immediately following exposure; some have reported a positive relationship between fiction reading and mentalizing (Kidd & Castano, 2013) while others report failure to replicate these effects (Panero et al. 2016, Samur, Tops & Koole 2018). However, even studies that fail to find an immediate effect of fiction reading report that estimates of fiction reading rates over the lifetime correlate with performance on mentalizing tasks. Moreover, several meta-analyses have found small but reliable effects of fiction reading in both short term manipulations (Dodell-Feder & Tamir, 2018) and in measurements of cumulative fiction reading (Mumper & Gerrig, 2017).

As Dodell-Feder and Tamir (2018) have noted, establishing the link between fiction reading and social cognition does not identify the reason(s) why such a link exists. Understanding the route by which fiction reading affects social cognition is important both for theories of social cognition and for potential applications: if non-social activities such as fiction reading affect social behavior and cognition, then manipulations of experience may inform development of social behaviors and perceptions. Several attempts to address the source of the fiction-social cognition link have focused on a potential role for fiction reading in emotion recognition, which we turn to next.

Fiction and Emotion Processing Accuracy

Readers of fiction are known to infer characters′ emotional states (Gernsbacher, Hallanda & Robertson, 1998). An embodied perspective goes further and suggests that that readers don′t

simply infer characters′ states; they simulate the emotional states while reading (Bal & Veltkamp, 2013; Mar & Oatley, 2008). According to this perspective, readers must "project themselves into the represented events" (Mar & Oatley, 2008, p. 173) of fiction. Empathic improvement from reading fiction depends upon "transportation and transformation" (Bal, Butterman, & Bakkar, 2011, p. 362) of the reader catalyzed by simulation and comprehension of the unfamiliar mental state of fictional characters. In other words, reading fiction causes simulation of characters′ emotions, providing practice being empathic, which in turn improves recognition of all emotions. Consistent with this perspective, some studies have found that the content of fiction text matters: romance, thrillers (Fong, Mullin, & Mar, 2013), and literary fiction (Kidd & Castano, 2013; 2017; Pino & Mazza, 2016) have been claimed to emphasize an exploration of the mental state of fictional characters and to be especially effective in supporting empathy.

Psychological theories that link emotion concepts to simulated emotion experience specifically predict that language experience, and therefore differences in accumulated language experience incurred through reading, is related to the ability to recognize emotion expression. Embodied simulation accounts hold that conceptual knowledge about emotion is grounded in part by sensorimotor and interoceptive experience, which constitute modal rather than abstract representations of emotion concepts (Kavanagh, Niedenthal, & Winkielman, 2012; Niedenthal, 2007). Further, emotion concepts are situated such that specific contexts determine which aspects of the modal representation is used in a given instance of emotion processing. A number of studies support the claim, derived from this account, that emotion concepts exert early effects on the perceptual processing of facial expression (Halberstadt & Niedenthal, 2001; Halberstadt Winkielman, Niedenthal, & Dalle, 2009). Halberstadt, and colleagues (2009) led participants to

conceptualize faces that expressed ambiguous blends of happiness and anger as either one or the other of those discrete emotions. On indicators of both perceptual memory and automatic facial mimicry, the researchers found evidence of biased encoding: ambiguous expressions encoded as "happy" later elicited more automatic smiling than did expressions encoded as "angry," for example.

Relatedly, psychological constructionist theories hold that conceptual knowledge encoded in language is used to provide specific emotional meaning to interoceptive and exteroceptive sensations associated with general affect (Barrett, 2017; Lindquist & Gendron, 2013; Lindquist, MacCormack & Shablack, 2015). Thus, "the scowls, frowns, grimaces, and growls you see over time presumably develop into conceptual knowledge for what *anger* looks like, helping you to make meaning of new instances of facial actions as instances of anger" (Doyle & Lindquist, 2017, p. 62). Several studies demonstrated that impairing individuals' access to emotion word meanings compromises their ability to accurately perceive discrete emotions in facial expression (Lindquist, & Gendron, 2013; Lindquist, Gendron, Barrett & Dickerson, 2014). And in a more recent study, Doyle and Lindquist (2017) showed that learning novel labels for never-before-seen facial actions expressed by "aliens" biased perceptual memory such that participants later recalled new instances of the facial actions as being more similar to the previously learned ones.

In addition to linking internal and external-perceptual experience, labels also serve to improve category learning by making categories more distinct (Lupyan, Rakison & McClelland, 2007). For example, imagine two people cooking, and one asks the other to pass the peeler. In the immediate context, the word "peeler" communicates the object requested, but in a broader sense, the use of the label "peeler" also serves as a social and linguistic signal that the object is a

distinct category from objects with other labels, such as "knife" and "grater", even though these objects have similar functions. Similarly, the word "anger" not only serves as a label for the concept, the label also helps the "anger" concept remain distinct from other labeled concepts such as "contempt" and "disgust."

Together, prior work suggest that reading fiction improves emotion recognition via embodied simulation of emotions in fictional characters, via category learning through exposure to emotion labels in fiction, or both. In principle, embodied simulation and learning via labels could independently inform emotion recognition. Fiction passages might describe a character's behaviors that could be simulated by the reader without the presence of any category labels. Conversely, participants may encounter emotion category labels without access engaging in social simulation. Regardless, closer investigation of the nature of emotion language in fiction texts may be informative about the link between fiction and emotion recognition.

Characterizing Beneficial Properties of Fiction

Most research into the fiction-emotion relationship emphasizes the characteristics of readers that moderate the relationship between fiction reading and empathy: participants with high scores on measures associated with fiction reading (Acheson, Wells & MacDonald, 2008), with greater self-rated textual engagement (Bal & Veltkamp, 2013; Johnson, 2012) greater trait openness (Djikic, Oatley, & Moldoveanu, 2013), and fewer depressive symptoms (Koopman, 2015) show the strongest relationships between fiction reading and empathy. Only a few studies have sought to characterize the properties of fiction text, as distinct from other types of text, that yield positive emotion processing outcomes. Koopman (2016) manipulated literary foregrounding -- holistically defined as original phonological, semantic, and grammatical features such as metaphor and imagery -- and showed texts with more foregrounding support

empathy to a greater degree than texts with less foregrounding. Using a small sample of experimental texts, Kidd, Ongis, & Castano (2016) claimed reflexive markers, operationalized as the proportion of words associated with discussion of mental states, partially mediated the effect of fiction reading on empathy. While some research has concluded that "the effect of [literary fiction] across experiments may not be easily reduced to superficial literature characteristics" (p. 380, Kidd & Castano, 2013), researchers have increasingly called for a close analysis of fiction literature to establish why fiction reading supports empathy (Kidd & Castano, 2017).

Large-scale language corpora have made it possible to characterize fiction and other genres of text via analyses of a genre's statistical patterns, derived from many million or multi-billion-word corpora. These statistical differences across genres have been linked to linguistic behavior, including vocabulary development (e.g. Goodman, Dale, & Li, 2008) and the comprehension and production of complex syntax (e.g. Montag & MacDonald, 2015). By characterizing natural language, large and representative corpora warrant claims about the relationship between language experience in everyday life and emotion processing abilities. To establish the qualities of fiction texts that are broadly relevant for emotion theory, statistical analyses of the qualities of emotion content, applied over large corpora of fiction and non-fiction texts, may aid in identifying the cause of fiction's impact on empathy.

Aims of the Present Study

The present work combines corpus analyses and behavioral methods to clarify how reading experience with emotion category labels affects emotion recognition. We focused on emotion category labels because they may have an important role for emotion recognition, as discussed above.

First, we conducted corpus analyses of fiction and non-fiction genres to quantify how frequently emotion category labels are used in an emotive sense in fiction and non-fiction texts. If fiction uniquely supports emotion recognition abilities, then emotion category labels should be used in an emotive sense more often in fiction than non-fiction. Second, in two experiments, we measured experience with reading fiction and emotion recognition abilities. If fiction reading experience supports emotion recognition, then participants with greater fiction reading experience should show greater recognition of emotions.

Corpus analyses

Corpus. For the following analyses, we employed the Corpus of Contemporary English (COCA; Davies, 2008), a growing body of English language across different genres. The corpus includes text tagged from various genres. For analysis, we split the corpus by Fiction, Spoken, and Other genres. Descriptions of all genres may be found in the COCA documentation (https://www.english-corpora.org/coca/help/texts.asp). Fiction texts included short-stories and plays from literary and popular magazines for both children (e.g. *Scholastic Scope*) and adults (e.g. *The New Yorker*) and first chapters of fiction books. The Spoken category includes transcriptions of unscripted conversations from popular television and radio programs (e.g. *Good Morning America*, *Jerry Springer*, *All Things Considered*). The Other category collapsed across remaining genres, including newspapers (e.g. *Associated Press*), academic journals (*e.g. Stanford Law Review*), and popular magazines (e.g. *Smithsonian*). At the time of analysis, the corpus was composed of roughly 560+ million words.

Procedure. Simple (anger, joy, surprise, disgust, sadness, fear) and complex emotion category labels (amusement, despair, relief, anxiety, pleasure, irritation, interest, pride), along

with their contexts spanning the 5 preceding and 5 following tokens were extracted from the corpus by lemma. Due to the large number of extracted tokens, one-twentieth of each emotion category label was randomly selected for analysis. A total of 20,172 emotion category labels and their contexts were analyzed. In reported statistics, errors in extraction from the corpus resulting in incomprehensible strings (n = 22) were removed from the dataset, leaving a total of 20,150 analyzed cases.

Because individual words are highly ambiguous, it was necessary to hand-code the sentence contexts to determine which examples of the target words were truly labeling emotions. For each token, research assistants were tasked with identifying whether the emotion category label was used in an emotive sense (e.g. *cry of* ***relief***) or not (e.g. *hurricane* ***relief*** *fund*). In instances where research assistants were uncertain about the emotive content of the emotion category label in context, one experimenter (SS) provided a judgment.

For each combination of emotions, corpus genres, and raters, a total of 20 contexts were randomly selected to be coded again by a third research assistant to calculate inter-rater reliability. In bins without a total of 20 contexts, all cases were selected, resulting in a total of 985 cases that were coded by both coder 1 and 3 or coder 2 and 3.

Results. Emotive ratings were analyzed with respect to the corpus in which the emotive category label occurred. We fit a binomial mixed effects regression, predicting emotive rating (1, 0) from emotion type (-0.5 = Simple; 0.5 = Complex), corpus genre (-0.5 = non-fiction; 0.5 = fiction), and their interaction along with a by-emotion random intercept. Note, the coding for corpus genre collapsed across Other and Spoken genres, as we were specifically interested in comparing fiction to other genres, though means of each genre are reported for comparison. There are three main findings of interest. First, complex emotions were used in an emotive sense

(M = .56) less frequently than simple emotions (M = .96), overall, as indicated by a significant effect of emotion type, b = -2.47, $X^2(1) = 10.31$, $p < .01$. Second, fiction corpora employed emotion category labels in an emotive sense more frequently (M = .98) than spoken (M = .70) or other corpora (M = .66), as indicated by a significant effect of corpus genre, b = 2.83, $X^2(1) = 537.64$, $p < .001$. Third, simple emotions were used in an emotive sense across all of fiction (M = .99), spoken (M = .93), and other corpora (M = .95), while complex emotions were used in an emotive sense consistently only in fiction corpora (M = .97) and not spoken (M = .46) or other corpora (M = .49), as indicated by a significant interaction between emotion type and corpus genre, b = 1.95, $X^2(1) = 65.05$, $p < .001$. This pattern is illustrated in *Figure 1*.

Contexts coded by two raters matched in rating for most cases (M = .81). Cohen′s kappa indicated significant overlap between rater 1 and rater 3 (κ = .77) and fair overlap between rater 2 and rater 3 (κ = .33).

Figure 1

Proportion of simple and complex emotions coded emotive or not by corpus genre

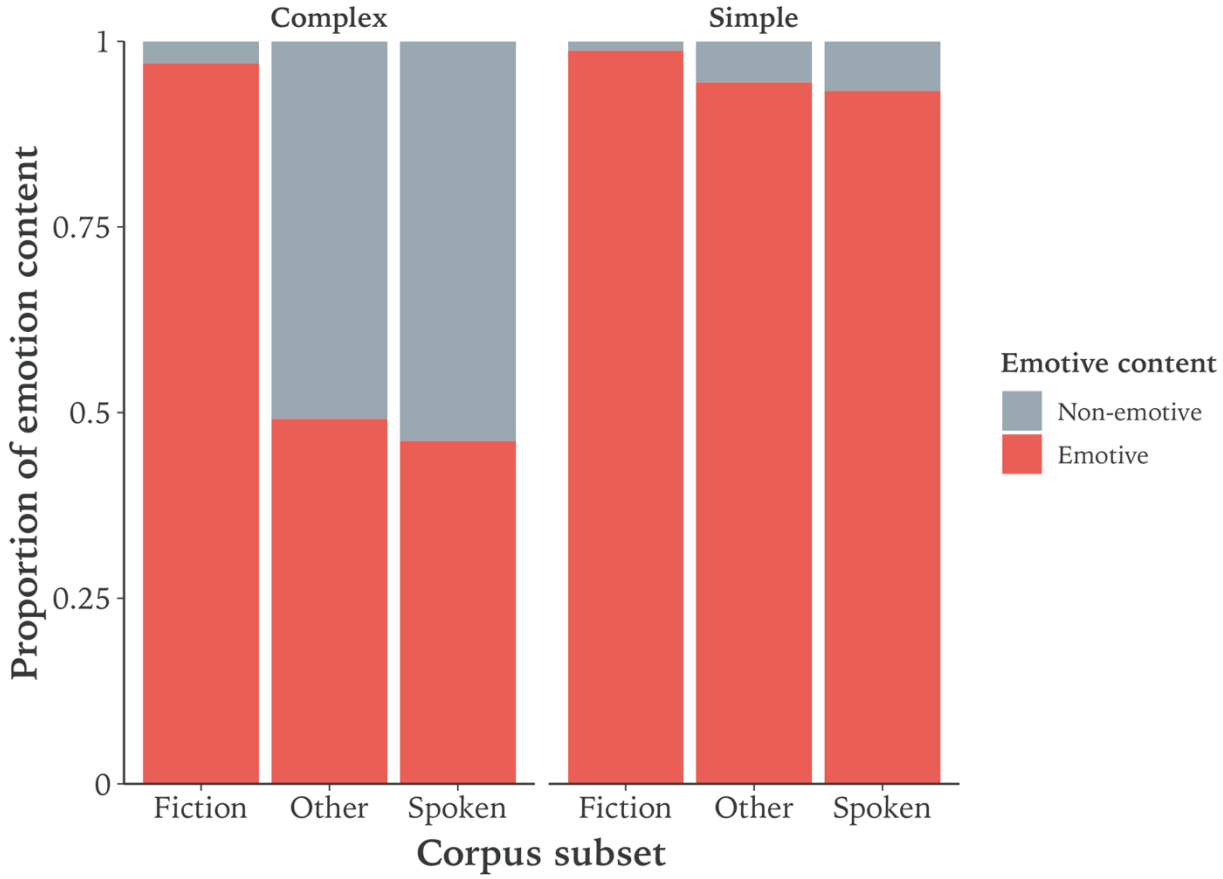

Fig. 1 Corpus analyses split by corpus genre. Fiction texts employed emotion category labels more often in emotive contexts (red) than in non-emotive contexts (gray). Complex emotions were used most regularly in an emotive sense in fiction texts, but not other genres. Simple emotions were used in emotive senses across all genres.

Discussion. These data suggest that complex emotion information is available from fiction corpora. Unlike simple emotions, which are used in an emotive sense equally across fiction, spoken language, and other corpus genres, complex emotions are used in an emotive

sense predominantly in fiction corpora. As a whole, these data complement many previous studies suggesting that fiction is a strong source of emotion information (e.g. Kidd & Castano, 2013). If these differences in the treatment of emotion category labels shape emotion concepts, then experience with fiction texts should predict emotion recognition abilities for complex emotions and less so for simple emotions.

Experiment 1

In this experiment, we tested the extent to which recognition of simple and complex emotions are differentially predicted by experience with fiction reading. We employed two tasks: the Author Recognition Task (ART; Acheson, Wells, & MacDonald, 2008) and the shortened version of the Geneva Emotion Recognition Test (GERT-S; Schlegel & Scherer 2016). The ART is presumed to assess fiction language experience through knowledge of authors, most of whom are fiction authors. If experience with fiction is particularly important source of complex emotion information, as suggested by our corpus analyses, and language experience affects emotion recognition abilities, then more experience with fiction should correspond to better recognition of complex emotions.

This experiment and the following experiment were approved by the University of Wisconsin-Madison I.R.B. and all participants gave their informed consent prior to participation. All data and analyses are available online (https://osf.io/79tmf/files/).

Participants. A total of 134 University of Wisconsin-Madison undergraduate students (M_{age} = 18.75; .70 female) received course credit for their participation. Each student indicated native experience with English, having spoken English in the home before the age of 5.

Materials. Videos of actors expressing emotions were taken from the GERT-S. Videos were shot from the chest up, capturing posture, movement, facial expression, and vocalizations

of actors, though no real language was expressed in any video. Actors expressed one of 14 different emotions, corresponding to the simple (anger, joy, surprise, disgust, sadness, fear) and complex (amusement, despair, relief, anxiety, pleasure, irritation, interest, pride) emotions identified in the corpus analyses. Participants judged 3 videos for each emotion, for a total of 42 total judgments from each participant.

We used the Author Recognition Task (Acheson, et al., 2008), a common assessment of fiction reading used in a number of studies of fiction reading and social cognition.

Procedure. Participants were seated in individual booths in a room with up to 5 other participants and the experimenter. Following completion of informed consent, participants entered a Qualtrics survey requesting demographics information. Participants then completed the ART followed by the GERT-S.

In the ART, participants were instructed to identify real authors and reject non-author, foil names. Participants were presented with a grid of names and asked to click on the names of real authors and ignore names of non-authors. Participants were told that points would be subtracted for selecting non-authors. They did not receive feedback on their selections.

At the beginning of the GERT-S, participants were requested to wear provided headphones. Participants were then given brief instructions about the nature of the task and provided instructions for each of the 14 emotion category labels that were employed in the task. During each trial, participants watched a video of an actor. Following completion of the video, the participant immediately labelled the emotion expressed by the actor using 1 of 14 emotion category labels arranged in a circle. Prior to beginning the task proper, participants were provided the opportunity to practice on one trial and given an option to practice another trial. The experiment progressed automatically until completion at which time the participant was informed

of their accuracy on the GERT-S.

Results. Accuracy on each trial of the GERT-S (0 = incorrect, 1 = correct) was regressed on the emotion type presented in the video (-0.5 = Simple, 0.5 = Complex), participant ART score (square-rooted and mean-centered), and the interaction between the two. By-item and by-participant random intercepts as well as a by-item random slope for participant ART score and a by-participant random slope for emotion type were included in the reported model. See *Figure 2* for average participant recognition accuracy per emotion type and smoothed trend.

Participants were not better at recognizing one type of emotion (simple/complex) over another, as indicated by similar rates of recognition for simple and complex emotions and a non-significant effect of emotion type, $X^2(1) = 0.49$, $p = .48$. Furthermore, participants who had higher ART scores were not better at recognizing emotions, overall, as indicated by a non-significant effect of ART score, $X^2(1) = 2.34$, $p = .13$. However, the interaction between the emotion type and ART score was significant, $b = 4.24$, $X^2(1) = 4.24$, $p < .05$, indicating that participants who had higher ART scores were better able to recognize complex emotions.

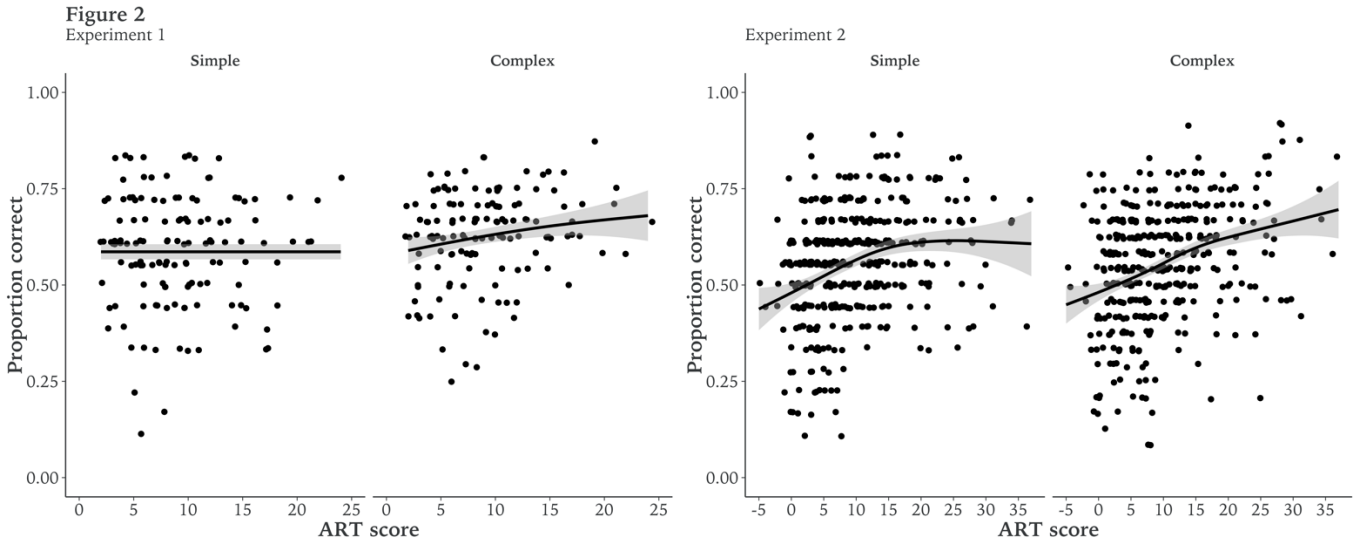

Fig 2 Performance in Experiment 1 and Experiment 2. Lines depict smoothed trend line estimated from raw data using a generalized additive model; bands are standard error. Dots indicate average performance in condition for each participant, jittered for visualization. Note, the x-axis scale for Experiment 2 represents a larger range of ART scores than Experiment 1.

Discussion. These results accord neatly with the previously reported corpus analyses. Participants with higher ART scores were better able to recognize complex emotions. These results suggest that fiction reading experience supports emotion recognition, and that this support may be specific to the emotions that are uniquely treated in an emotive sense in fiction texts.

Experiment 2

Experiment 2 sought to replicate Experiment 1 to assess reliability with a larger and more diverse sample. The experiment was thus modified to be run on Amazon's Mechanical Turk. Design and analyses were pre-registered using the Open Science Framework (https://osf.io/jc95w; Foster & Deardorff, 2017).

Participants. A sample of 400 native English speakers (spoken English in the home before the age of 7) from the United States were compensated $2.50 for their participation. In addition to location and language requirements, participants were required to be between the ages of 18 and 25 (M_{age} = 21.90; .52 female). We retained an age range similar to that in Experiment 1 because ART score is positively correlated with age; a longer history of reading provides more opportunity for encountering author names on the test.

Data was collected in an initial batch of 10 participants, which was checked by an experimenter to ensure proper saving and storage. An additional 390 participants were then

collected following this step, and data was analyzed only following collection of all 400 participants.

Materials. The GERT-S and ART were the same as used in Experiment 1 with an addition of some attention checks. Attention checks were added to the ART and to the end of the experiment to filter out participants responding randomly. The ART attention check required participants to click one box labeled "Please indicate this box is an author" placed randomly in the list of names. Participants were instructed of the presence of this box, though they were not told where the box was located. The attention check at the end of the experiment entailed a definition matching task. Participants were required to match definitions of the 14 emotion category labels employed in the GERT-S to the words. Each participant saw a 15th definition that read "Please choose the response 'pride'."

Procedure. The procedure for this experiment was the same as Experiment 1, though no interaction with the experimenter was possible due to method of administration.

Results. Following our pre-registered analysis plan, a total of 13 participants were removed for responding randomly, leaving 387 participants in the final analyses. Data were analyzed in the same way as Experiment 1. Participants with higher ART scores were more likely to recognize emotions correctly than participants with lower ART scores, as indicated by a significant effect of ART score, $b = 0.24$, $X^2(1) = 56.74$, $p < .001$. However, participants were not better at recognizing one type of emotion moreso than another, as indicated by a non-significant effect of emotion type, $X^2(1) = 0.01$, $p = .92$. Finally, participants with higher ART scores were not better at recognizing complex emotions over simple emotions, as indicated by a non-significant interaction between ART score and emotion type, $X^2(1) = 1.19$, $p = .28$. See *Figure 2* for a visual summary of these results.

Discussion. These results further suggest that fiction reading experience supports emotion recognition. While the predicted differences in recognition by an interaction of emotion type and ART score were not supported in this sample, as they were in Experiment 1, these findings are broadly in line with the literature suggesting that long-term fiction reading supports emotion recognition, overall (e.g. Panero et al., 2016).

There are several potential reasons why Experiment 2 did not find an interaction between ART score and emotion type as did Experiment 1. All participants from Experiment 1 were drawn from the relatively homogenous undergraduate population of the University of Wisconsin-Madison. In contrast, participants in Experiment 2 were drawn from the relatively heterogeneous population of the Mechanical Turk workforce of the United States. The ART employed in Experiments 1 and 2 was validated using an undergraduate sample (Acheson, Wells, & MacDonald, 2008), but there is significant variation in language experience across cultures of the United States. Indeed, application of the ART to different cultures requires careful selection of author names to create a valid and meaningful measure of reading experience (e.g. Chen & Fang, 2015; Lee et al., 2019; Masterson & Hayes, 2007). One measure of fiction reading experience may therefore capture many different kinds of language experiences between samples, even for participants with the same ART score. In Experiment 1, these experiences mapped neatly onto the patterns in COCA, and in Experiment 2, they did not. If cultural differences between participants undermined the interaction, then more sensitive measures of language experience and corresponding analyses of culturally-specific language corpora would prove useful in characterizing how statistical properties of language affect emotion recognition.

Relatedly, control of moderating factors may unmask the interaction in Experiment 2. Previous studies have demonstrated control for factors like gender, age, and education does not diminish the overall relationship between ART score and emotion recognition (Kidd & Castano, 2013), though more subtle relationships may exist. For example, participants with a college education select fear and disgust labels more often in emotion recognition tasks than participants with no college education (Trauffer, Widen, & Russell, 2013). If ART score correlates with education – a likely assumption – then the more educated participants in the heterogeneous sample of Experiment 2 may have recognized simple and complex emotions at similar rates. Gender may also be relevant. Women are more likely to accurately recognize subtle emotions (Hoffman et al., 2010) and negatively valanced emotions (Thompson & Voyer, 2014) than men, and Summers (2013) suggested women have a stronger proclivity to read fiction books over other books. Environmental differences between Experiment 1and 2 may have also attributed to the discrepancies between the two studies. Participants in Experiment 1 completed the study in a quiet room at individual booths with over-ear headphones, while participants in Experiment 2 completed the study online in an environment of their choosing. It could be that distractions may have impacted performance on subtle aspects of the GERT, such as recognition of complex emotions. Additional measurement and control of factors correlated with both ART score and emotion recognition abilities, as well as environmental factors that may influence performance on the GERT, could help clarify the emotion-fiction link.

Finally, the interaction in Experiment 1 may reflect a Type I error. Experiment 2 was specifically designed to replicate the finding of Experiment 1 in a larger and more diverse sample to establish the generalizability and reliability of the findings, and the failure to replicate the

critical interaction should invite a critical look at the corpus analyses and the Experiment 1. However, rather than suggesting that statistical properties of language have no bearing on emotion recognition, these results should prompt alternative approaches to examining the statistical properties of language and their relationship with emotion. Suggested studies are taken up in the General Discussion.

General discussion

These experiments take a first step toward empirically linking emotion recognition with properties of natural language experience. Our corpus analysis showed that words labeling a complex emotion are used particularly often in an emotive sense in fiction but not in other genres, whereas words with a simple emotion meaning are used in an emotive sense equally often across all three genres. In Experiment 1, participants with more fiction reading experience, as indexed by ART scores, were better able to recognize complex emotions than were participants with lower fiction reading experience. In Experiment 2, participants with higher ART scores were better able to recognize emotions overall compared to those with lower scores, but the specific effect for complex emotions seen in Experiment 1 was not reliable. Together, these findings suggest that fiction reading experience is related to emotion recognition abilities and that statistical properties of language experience may contribute to emotion knowledge.

These results may have implications for the role of emotion category labels in emotion recognition, in that we show fiction may increase readers′ exposure to some kinds of category labels. Category labels help people learn emotion categories and distinguish a given category from others (Doyle & Lindquist, 2017). Access to that label supports emotion recognition (Lindquist, & Gendron, 2013; Lindquist, et al., 2014). Our corpus analyses ground this phenomenon in natural experience; different corpus genres exhibit statistical differences in their

use of emotion category labels, which may, in turn, affect the formation of coherent emotion concepts. While researchers have previously analyzed the properties of small samples of fiction (Koopman, 2016) or experimentally manipulated text (Kidd, Ongis, & Castano, 2016), the corpus analyses employed here warrant claims about natural language experience. Identifying this statistical property over large text corpora is critical given the importance of long-term measures of fiction reading experience in emotion recognition (e.g. Panero et al., 2016; Samur et al., 2018).

The holistic measure of semantic sense we employ approximates the relationship between an emotion category label and an emotion concept. Cognitive research into properties of word meaning could guide future research and more nuanced measures. For example, statistical properties like word frequency, contextual diversity, and semantic distinctiveness each impact word recognition and comprehension (Johns, Gruenenfelder, Pisoni, & Jones, 2012); likewise, contextual distinctiveness may capture some variation in expression of the emotion concept that informs emotion recognition.

Higher order statistical properties of language use will likely be particularly important to make more progress understanding how emotion recognition is shaped by fiction reading experience. In social interactions, emotion recognition requires the integration of facial expression, posture, and tone, much of which is implied but not always present in fiction text. Distributional models of semantics derived solely from statistical properties of text corpora capture some experiential and visual semantic features (see Lewis, Zettersten, & Lupyan, 2019 in response to Kim, Ellis, & Bedny, 2019). A similar learning mechanism may support the formation of emotion concepts through language in embodied and constructionist perspectives.

Corpus analyses beyond fiction texts investigating fine-grained corpus sub-genres may further qualify the link between language experience and empathy. For example, researchers have argued that literary fiction is a particularly beneficial source of emotion content as compared to genre fiction (e.g. Kidd & Castano, 2013). An analysis of the properties of emotion category labels in literary vs genre fiction may clarify why fiction reading experience improves empathy. Language patterns vary dramatically between genre (e.g. Johns & Jamieson, 2018). Corpus analyses may prove useful for qualifying how other types of language experience – in television shows, movies, children's literature, spoken language, and more – may correspond to the formation of specific emotion concepts.

Distributional statistical analyses of natural language could prove useful in further defining individual and cross-cultural differences in emotion recognition (Elfenbein et al., 2007; Jack et al., 2012; Jackson et al., 2019). Language structure varies dramatically between cultures (e.g. Johns & Jamieson, 2019), and emerging research suggests that these linguistic differences may correspond to cultural variation in emotion concepts (Jackson et al., 2019). Behavioral data employing culturally sensitive language experience measures and emotion recognition tasks would enable a parallel to be drawn between linguistic patterns and emotion concepts.

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior research methods*, *40*(1), 278-289. https://doi.org/10.3758/BRM.40.1.278

Bal, P. M., Butterman, O. S., & Bakker, A. B. (2011). The influence of fictional narrative experience on work outcomes: A conceptual analysis and research model. *Review of General Psychology*, *15* (4), 361-370. https://doi.org/10.1037/a0025416

Bal, P. M., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PloS one*, *8*(1), e55341. https://doi.org/10.1371/journal.pone.0055341

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* , *42*(2), 241-251. https://doi.org/10.1017/S0021963001006643

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. Social Cognitive and Affective Neuroscience, 12, 17–23. http://dx.doi.org/10.1093/scan/nsx060

Chen, S.-Y., & Fang, S.-P. (2015). Developing a Chinese version of an Author Recognition Test for college students in Taiwan. *Journal of Research in Reading, 38*(4), 344–360. https://doi.org/10.1111/1467-9817.12018

Davies, M. (2008). The corpus of contemporary American English (COCA): 560 million words, 1990-present. https://www.english-corpora.org/coca/

Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Reading other minds: Effects of literature on empathy. *Scientific Study of Literature*, *3*(1), 28-47. https://doi.org/10.1075/ssol.3.1.06dji

Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. *Journal of Experimental Psychology: General, 147*(11), 1713–1727. https://doi.org/10.1037/xge0000395

Doyle, C. M., & Lindquist, K. A. (2017). *Language and emotion: Hypotheses on the constructed nature of emotion perception.* In J.-M. Fernández-Dols & J. A. Russell (Eds.), *Oxford series in social cognition and social neuroscience. The science of facial expression* (p. 415–432). Oxford University Press.

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7*(1), 131–146. https://doi.org/10.1037/1528-3542.7.1.131

Fong, K., Mullin, J. B., & Mar, R. A. (2013). What you read matters: The role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(4), 370 - 376. https://doi.org/10.1037/a0034084

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association : JMLA*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Gernsbacher, M. A., Hallada, B. M., & Robertson, R. R. W. (1998). How automatically do readers infer fictional characters' emotional states? *Scientific Studies of Readin: The Official Journal of the Society for the Scientific Study of Reading* , *2*(3), 271–300. https://doi.org/10.1207/s1532799xssr0203_5

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Halberstadt, J. (2005). Featural Shift in Explanation-Biased Memory for Emotional Faces. *Journal of Personality and Social Psychology, 88*(1), 38–49. https://doi.org/10.1037/0022-3514.88.1.38

Halberstadt, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology, 81*(4), 587–598. https://doi.org/10.1037/0022-3514.81.4.587

Halberstadt, J., Winkielman, P., Niedenthal, P. M., & Dalle, N. (2009). Emotional conception: How embodied emotion concepts guide perception and facial action. *Psychological Science*, *20*(10), 1254-1261. https://doi.org/10.1111/j.1467-9280.2009.02432.x

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta psychologica*, *135*(3), 278–283. https://doi.org/10.1016/j.actpsy.2010.07.012

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109

Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517-1522. https://doi.org/10.1126/science.aaw8160

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of*

the Acoustical Society of America, *132*(2), EL74–EL80. https://doi.org/10.1121/1.4731641

Johns, B. T., & Jamieson, R. K. (2018). A large-scale analysis of variance in written language. *Cognitive Science, 42*(4), 1360–1374. https://doi.org/10.1111/cogs.12583

Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences, 52*(2), 150–155. https://doi.org/10.1016/j.paid.2011.10.005

Lee, H., Seong, E., Choi, W., & Lowder, M. W. (2019). Development and assessment of the Korean Author Recognition Test. *Quarterly Journal of Experimental Psychology*, *72*(7), 1837–1846. https://doi.org/10.1177/1747021818814461

Lewis, M., Zettersten, M., & Lupyan, G. (2019). Distributional semantics as a source of visual knowledge. *Proceedings of the National Academy of Sciences*, *116*(39), 19237-19238. https://doi.org/10.1073/pnas.1910148116

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342* (6156), 377–380. https://doi.org/10.1126/science.1239918

Kidd, D., & Castano, E. (2017). Different stories: How levels of familiarity with literary and genre fiction relate to mentalizing. *Psychology of Aesthetics, Creativity, and the Arts, 11*(4), 474–486. https://doi.org/10.1037/aca0000069

Kidd, D., Ongis, M., & Castano, E. (2016). On literary fiction and its effects on theory of mind. *Scientific Study of Literature, 6*(1), 42–58. https://doi.org/10.1075/ssol.6.1.04kid

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213-11222. https://doi.org/10.1073/pnas.1900952116

Koopman, E. M. E. (2015). Empathic reactions after reading: The role of genre, personal factors and affective responses. *Poetics*, *50*, 62-79. https://doi.org/10.1016/j.poetic.2015.02.008

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. Emotion Review, 5, 66–71. http://dx.doi .org/10.1177/1754073912451351

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*(2), 375–387. https://doi.org/10.1037/a0035293

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology, 6,* Article 444. https://doi.org/10.3389/fpsyg.2015.00444

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science, 3*(3), 173–192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality, 40*(5), 694–712. https://doi.org/10.1016/j.jrp.2005.08.002

Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes, *Communications*, *34*(4),

407-428. doi: https://doi.org/10.1515/COMM.2009.025

Masterson, J., & Hayes, M. (2007). Development and data for UK versions of an author and title recognition test for adults. *Journal of Research in Reading, 30*(2), 212–219. https://doi.org/10.1111/j.1467-9817.2006.00320.x

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General, 144*(2), 447–468. https://doi.org/10.1037/xge0000054

Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 109–120. https://doi.org/10.1037/aca0000089

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930

Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *Journal of Personality and Social Psychology, 111*(5), e46–e54. https://doi.org/10.1037/pspa0000064

Pino, M. C., & Mazza, M. (2016). The use of "literary fiction" to promote mentalizing ability. *PloS one*, *11*(8), e0160254.

Samur, D., Tops, M., & Koole, S. L. (2018). Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013). *Cognition and Emotion, 32*(1), 130–144. https://doi.org/10.1080/02699931.2017.1279591

Schlegel, K., & Scherer, K. R. (2016). Introducing a short version of the Geneva Emotion Recognition Test (GERT-S): Psychometric properties and construct validation. *Behavior Research Methods,*

48(4), 1383–1392. https://doi.org/10.3758/s13428-015-0646-4

Summers, K. (2013). Adult Reading Habits and Preferences in Relation to Gender Differences. *Reference & User Services Quarterly*, *52*(3), 243-249.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion, 28*(7), 1164–1195. https://doi.org/10.1080/02699931.2013.875889

Trauffer, N. M., Widen, S. C., & Russell, J. A. (2013). Education and the Attribution of Emotion to Facial Expressions. *Psychological Topics*, *22*(2), 237-247.