FACIAL EXPRESSION RECOGNITION IN PEOPLE WITH DIFFERING LEVELS OF EATING DISORDER SYMPTOMS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Ilya Nudnou

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Psychology

July 2022

Fargo, North Dakota

North Dakota State University
Graduate School

Title

FACIAL EXPRESSION RECOGNITION IN PEOPLE WITH DIFFERING LEVELS OF EATING DISORDER SYMPTOMS

By

Ilya Nudnou

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Benjamin Balas
Chair

Katherine Duggan

Lauren Schaefer

Simone Ludwig

Approved:

August 4, 2022
Date

Mark Nawrot
Department Chair

ABSTRACT

Previous studies of emotion categorization abilities of people with eating disorders used accuracy and reaction time to identify performance deficits for these individuals. The conclusions from this literature have been mixed, due in part to low sample sizes and inconsistent assessment of comorbid diagnoses. The current study re-examined eating disorder symptom severity as a function of emotion categorization abilities, using visual cognition paradigms that offer insights into how emotional faces may be categorized, as opposed to how well these faces are categorized. This relationship was examined while controlling for anxiety, depression, alexithymia, and emotion regulation. Visual information use, emotion representation fidelity, and categorization accuracy were unrelated to eating disorder symptom severity in a sub-clinical sample of undergraduate students. Future research may benefit from the visual cognition tasks validated in this study. More complex designs are needed to test mediational pathways through which recognition deficits may operate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

B .... Unstandardized regression weight for an individual predictor variable.

SE .... Standard Error.

Adj. R^2 .... Adjusted variance explained within the criterion variable by a regression model.

F .... ANOVA F-statistic.

p .... probability value for null hypothesis testing.

df .... degrees of freedom.

χ^2 .... likelihood ratio difference statistic.

1. INTRODUCTION

The term "eating disorder" is used to describe a number of similar yet varying pathologies, involving abnormal behaviors and conceptualizations related to the consumption of food (e.g., extreme dietary restriction, self-induced vomiting, excessive exercise). The yearly prevalence of eating disorders in U.S. adults is 3.6 million for binge eating, 0.9 million for bulimia nervosa, and 1.8 million for anorexia nervosa (Hudson et al, 2007; Merikangas et al., 2010). Even after recovery, the effects of eating disorders are likely to persist over the lifespan, such as a higher likelihood of developing obesity (de Oliveira et al., 2013). The estimated costs of eating disorders in the U.S. for the 2018-2019 fiscal year were upwards of $64.7 billion (Streatfeild et al., 2021). Recovery from eating disorders involves not only the re-initiation of healthy dietary behaviors, but also improvements on dimensions of psychological and social well-being (Bardone-Cone et al., 2010; de Vos et al., 2017).

In a recent review of the eating disorder literature, Mason et al. (2020) highlight emotion recognition and emotion regulation deficits as a primary component of problematic social cognition of individuals with an eating disorder. These deficits combine to worsen the effects of any one problematic component of eating disorders. For example, emotion recognition deficits prevent normal social functioning, which leads to a lack of practice at regulating emotions which may occur during everyday life (Lavender et al., 2015). The reduced social networks of these individuals preclude them from employing inter-personal emotion regulation strategies as an alternative to the self-regulation strategies, which they are already worse at using. This in turn leads to problematic dietary behaviors as a maladaptive but readily available emotion regulation strategy (Leehr et al., 2015).

Literature on emotion recognition deficits contains many accounts of group differences between eating disorder patients and neurotypical controls in emotion perception. Original work in this area primarily focused on female patients with anorexia nervosa (Kucharska-Pietura et al., 2004; Zonnevijlle-Bendek et al., 2002), and included emotion recognition tests as an additional measure of impaired emotional functioning. From this early work a connection between poor emotion recognition and eating disorders was established, but the mechanisms underlying this deficit were unclear.

Since then, a number of research studies replicated and expanded the relationship between eating disorders and emotion recognition (Castro et al., 2010; Dapelo et al., 2016; Fujiwara et al., 2017; Jänsch et al., 2009; Pollatos et al., 2008; Pringle et al., 2010). Jänsch et al. (2009) asked 28 patients with Anorexia Nervosa (AN) and 28 healthy controls to categorize morphed emotional faces. These stimuli were taken from Ekman and Friesen's (1976) Pictures of Affect Series and contained face images of six emotion categories (anger, disgust, fear, happiness, sadness, surprise). These images were blended in pairs with a neutral expression of the same actor in a continuous fashion to produce increasingly emotional faces. Images of different proportions of the target emotion were then shown to participants. AN patients were assessed for symptoms of anxiety and depression in addition to the primary eating disorder diagnosis. The authors found an overall worse categorization accuracy and slower reaction time for all emotions for AN patients as compared to healthy controls, with no emotion being miscategorized to a greater extent. This difference was found at all levels of the morph continuum, suggesting that the deficit was maintained across emotional intensity. Depression symptoms were also negatively related to overall accuracy within the AN patient group.

Castro et al. (2010) used happy and sad images, also taken from Pictures of Affect Series, which were blended with neutral according to methods in Surguladze et al. (2004). These were presented to 30 AN patients and 40 healthy controls. AN patients were also assessed for psychotic and substance abuse disorders (which were cause for exclusion), as well as anxiety, depression, and obsessive-compulsive disorder symptoms. The authors found that AN patients were less accurate at telling apart sad faces from neutral during categorization than healthy controls, but both groups had the same discrimination accuracy for happy faces. Furthermore, the accuracy difference was driven primarily by obsessive-compulsive symptoms within the AN patient group. This finding, among others, underscores the importance of taking into account comorbid conditions of patients with eating disorders.

Dapelo et al. (2016) also used blended emotions (anger, disgust, fear, happiness, sadness) which were taken from Pictures of Affect Series, and blended with neutral according to methods in Young et al. (1997). These stimuli were presented to 35 AN patients and 42 healthy controls. AN patients were also assessed for symptoms of anxiety, depression, obsessive-compulsive disorder, and alexithymia (the inability to recognize or describe one's own emotions). The authors found that AN patients were less accurate at categorizing expressions of disgust, with no other accuracy differences from healthy controls. Furthermore, there were no associations of this accuracy deficit with any of the other disorder symptoms which were assessed.

Fujiwara et al. (2017) also used blended emotions, morphing between specific pairs of six emotion categories (happiness-surprise-fear-sadness-disgust-anger) in different proportions, according to recommendations of Wilhelm et al. (2014), who developed their own database. In this study, a mixed group of 24 patients with both Anorexia and Bulimia Nervosa and 50 healthy controls was assessed. The participants were asked to estimate the proportions of the pairwise categories in the mixed emotional faces instead of just categorizing the emotion on these faces. AN patients were also assessed for symptoms of anxiety, depression, and alexithymia. As the authors were particularly interested in alexithymia, half of the healthy controls were recruited based on high alexithymia symptoms. Patients with an eating disorder were less accurate than healthy controls at judging the proportions in morphs which included anger and disgust, but not other emotions. Since alexithymia was closely confounded with eating disorder symptoms, it was not possible to make a definitive conclusion about its unique contribution.

While the above sampling of studies provides important avenues into eating disorder research, it is difficult to tie these results together. For example, Pollatos et al. (2008) found accuracy deficits for categorizing neutral, sad, and disgusted faces, Pringle et al. 2010 found deficits for neutral and angry faces, while Jänsch et al., (2009) found deficits for angry, fearful, and disgusted faces. The variation in outcomes across studies complicates the formation of a comprehensive conceptual understanding of what specific perceptual difficulties are involved in eating disorders, but there are still some common themes that have emerged. The recognition of positive emotions such as happiness and joy were not impaired in patients with eating disorders in the studies I examined.

Additionally, disgust was present in multiple studies, suggesting that this emotion in particular may be important for patients with eating disorders. A common misrepresentation of neutral and disgust as anger has been found in multiple studies (Dapelo et al., 2016, Fujiwara et al., 2017). It seems likely that patients are less able or willing to process negative expressions which are directed at them. However, it is still unclear whether this results from a low-level visual bias, or a higher-level cognitive process (such as attention or motivation).

Comorbid disorders explaining some or most of the variance in visual task performance over and above eating disorders is also a common finding in the literature. However, in this case, there is as little consistency in terms of which disorders are assessed. This is particularly problematic since anorexia nervosa, a common diagnosis of patients in previous studies, has been found to be very heterogenous in terms of comorbid conditions (Jordan et al., 2008).

Inconsistent results could also simply be due to a lack of statistical power to consistently detect an effect, primarily due to the small samples of patients tested. A sample size of a few dozen only allows for a detection of large group effects, which may be reduced in scope by sample heterogeneity, and may in fact be smaller than anticipated. For example, in order to detect an effect size of 0.52, which is the same as the relationship between increasing age and declining speed of information processing (Meyer et al., 2001; Verhaeghen & Salthouse, 1997), with 80% power would require 60 participants per group (Faul et al., 2009). A more reasonable effect size of around 0.15 would require *seven hundred* participants per group. Because of this, the sample sizes seen in the above studies cannot really be used to make definitive conclusions about uncovered emotion recognition accuracy effects. The average sample size of these studies (35 patients) would result in a power of 24% to detect a small-medium effect of 0.15.

In addition to the inconsistent positive effects found in the literature, other studies found no effect of eating disorder symptoms or diagnosis (Brewer et al., 2015; Phillipou et al., 2015; Sfärlea et al., 2016). This is particularly problematic for studies which are very similar in methodology and performed analyses. For example, both Jänsch et al. (2009) and Phillipou et al. (2015) tested 24 and 28 individuals with Anorexia Nervosa against a matched sample of neurotypical controls on a 7-alternative-forced choice emotion categorization task. The stimuli in both studies consisted of the six emotion categories from Pictures of Affect (Ekman and Friesen, 1976) – anger, disgust, fear, happiness, sadness, surprise, and neutral. However, only Jänsch et al. (2009) found an effect of eating disorder diagnosis on accuracy and reaction time, while Phillipou et al. (2015) did not find any differences. A recent large-scale study by Wyssen et al. (2019) which included 61 patients with AN and 58 patients with bulimia nervosa (BN) only found a mild disgust recognition impairment for the BN group as compared with healthy controls.

I attempt to build upon the relatively mixed literature with a set of contributions which attempt to expand and explain any potential emotion recognition deficits. Given the issue of statistical power, I adopt a paradigm and a target sample of participants which allow me to make a well-powered conclusion. To capture the variability present in comorbid diagnoses, I collected information about multiple symptoms that are typically comorbid with eating disorder symptoms and evaluated them along with eating disorders within the same statistical model. I employ a multiple linear regression approach in this study, focusing on symptom severity instead of diagnoses. By considering disordered eating and comorbid conditions on a continuous severity scale, I am able to talk about continuous change. That is, to what extent does an increase in symptom severity leads to worse emotion recognition ability. Considering symptoms in this way also allows me to look at whether the relationship between eating disorder symptoms and emotion recognition ability is stronger when a person also has severe symptoms in a comorbid disorder. Finally, the benefit of a multiple regression model over a group difference model is that I have more power to make conclusions about main effects (for example, eating disorder symptoms on emotion recognition while controlling for symptoms of other disorders). As an illustration, in order to have equivalent power to detect the same effect size for the main effect of accuracy on eating disorder symptoms while controlling for depression and anxiety, a regression model requires only 77 participants. Note that these would not even have to be patients – there just needs to be adequate spread in the ED symptom severity. On the other hand, a group difference account (such as a t-test with dichotomized symptoms or a diagnosis), would require the seven hundred participants I mentioned in a previous example. Finally, a regression-based model can more easily incorporate the potentially multivariate results from the complex visual cognition tasks I plan to employ.

1.1. Emotion Recognition Tasks

To investigate potential mechanisms of impaired emotion recognition, I employed face processing tasks that make it possible to examine the perceptual mechanisms that support visual recognition in general. I selected the tasks in the current study because they offer additional insights regarding the image properties which contribute to emotion

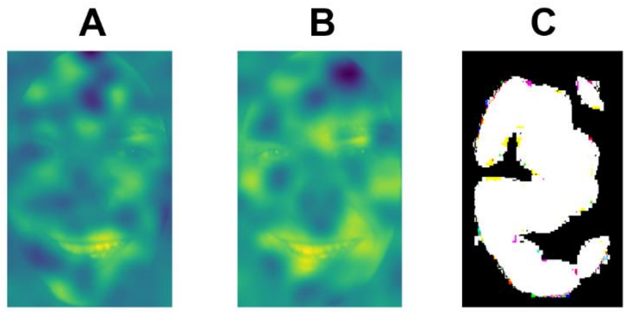

Figure 1. Bubbles Task.
(A) an information use map for an individual primarily using the mouth region. (B) an information use map for an individual using multiple regions. (C) representation of actual information that is available on a specific image (black areas are those covered by the bubbles noise).

recognition, and the internal representations maintained by observers for emotion recognition. Accuracy and reaction time-based tasks provide an important foundation for uncovering the nature of perceptual difficulties in eating disorder patients. The current study aims to build on this foundation to more specifically describe the nature of such perceptual difficulties. It is of course possible to vary visual information and obtain accuracy and reaction time metrics, and then compare these based on the information that was varied. For example, if we believe that the eye region is important in emotion recognition, we may obscure it on some trials and then examine how accuracy and reaction time is affected for those trials. However, besides the eye region, there are likely a large number of meaningful parts of a face which would affect emotion recognition accuracy or reaction time. Furthermore, defining the eyes as a discrete feature is in itself an assumption about how fragments of face images have perceptual relevance. Depending on available evidence, it may be very difficult to decide which facial features are important for all of the participants in a study. Instead, it is possible to examine and visualize individuals' visual information use without needing to make assumptions about the importance of discrete features. This is particularly relevant to situations in which an a-priori assumption is difficult to make. The *Bubbles task* (Gosselin & Schyns, 2001) involves randomly obscuring parts of an image while participants make categorization judgements. This is a method of evaluating the perceptual system's ability to make flexible categorizations of the same stimulus based on available information or task demands. This method allowed me to examine which parts of an image lead to successful recognition of emotions at the level of individual pixels. Observer information use can also be compared to a computational model (often referred to as the *ideal observer*) which optimally uses image pixels for categorization (Figure 1). The Bubbles task has been widely used to determine which parts of the face are important for recognizing different identities, genders, and emotions. A common result from that literature is that people focus primarily on the eye region for identity discrimination (Song et al., 2012), and the mouth region for emotion discrimination (Lee et al., 2011). Both the specific regions used by individuals and how much available information they use can inform a theory of underlying mechanisms involving successful emotion recognition. Observers may consistently avoid using information from parts of the face which would be useful to them. This may indicate inefficiencies in representation of the categories themselves, or higher-level cognitive mechanisms related to impaired attention or motivation.

A related behavioral task can assess individuals' representations of emotion categories. The *prototype estimation task* (Jack et al., 2012; Mangini & Biederman, 2004) involves randomly modifying pixels of an ambiguous image, while participants perform a categorization task. For example, participants may be presented with emotionally expressive faces, and asked to choose which of the images looks happier. Critically, the addition of image noise typically transforms the ambiguous image enough to bias

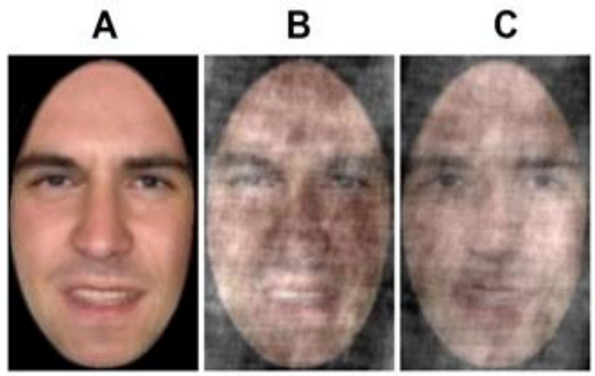

Figure 2. Prototype Estimation Task.
(A) an ambiguous image depicting blended anger and happiness. (B) the addition of noise makes this image look like an angry expression. (C) the subtraction of the same noise pattern makes this image look like a happy expression.

responses towards a particular category (Figure 2). Noise patterns based on participant responses are then combined and applied to the original image. This generates a visualization of the individual's internal representation of a category. These visualizations of target categories can then be presented to trained observers and image analysis algorithms to quantify appearance differences. For examples, new observers might rate emotion prototypes according to how closely they resemble target categories. Alternatively, observers can be asked to rate a property of the image which was not included in the original task, such as whether happy faces look more feminine or masculine. The prototype estimation task has been previously used in multiple applications, such as uncovering people's internal representations of other race faces and social traits (Todorov et al., 2011). For example, it was demonstrated that people with higher levels of prejudice towards Moroccans represent faces of that ethnicity as angrier than the neutral representations of the less prejudiced individuals. In the current study, I was primarily interested in the fidelity of emotion prototypes as a function of eating disorder severity. I expected that individuals with higher eating disorder severity would have prototypes which were less distinct from one another than those of individuals with less severe symptoms. Furthermore, prototypes of individuals with less severe symptoms are likely to possess more intense expressions of a specific category (such as a more defined smile of happiness).

1.2. Comorbid Conditions of Eating Disorders

In addition to uncovering the mechanisms underlying emotion recognition, I included a number of relevant comorbidities present in patients with eating disorders to assess shared mechanisms which underlie these pathologies. This allowed me to examine the extent to which emotion recognition performance may have contributed to eating disorder symptom severity over and above other comorbid conditions. When combined with the perceptual tasks I employed, these regression models provide a foundation for a theoretical account of emotion recognition in individuals with mental disorders.

Based on the previous literature on the topic, I identified a number of potential comorbidities which may affect facial emotion recognition: depression (Kucharska-Pietura et al., 2004; Rothschild-Yakar et al., 2019; Sfärlea et al., 2018), anxiety (Lulé et al., 2014; Xu et al., 2017), alexithymia (Brewer et al., 2015; Fujiwara et al., 2017; Zonnevijlle-Bendek et al., 2002). Emotion regulation ability was also assessed in multiple previous studies of eating disorders (Nalbant et al., 2019; Robinson et al., 2015; Seidel et al., 2018), which is why I also include it as a covariate of my effects. While there may be other disorders of interest which may affect emotion recognition, I had to limit myself to the above choice due to power concerns. Even while doing so, my power remains adequate only for tests of partial main effects. What is often found in previous literature about multiple disorders is that one or multiple disorders act as precursors, mediators, or moderators of a main effect of interest. I cannot definitively conclude that mediation is taking place both because of my cross-sectional design and due to the lack of power within my target sample size. However, finding that some variance is accounted for by the inclusion of a comorbid disorder would suggest that this relationship should be targeted in future studies about this topic. Essentially, I was powered to make substantial conclusions within the eating disorder literature, and was able to provide fruitful future directions for further research via my exploratory analyses. As can be seen from recent reviews (Cotter et al., 2018; Mason et al., 2020), emotion recognition deficits are shared by many disorders, and this study is a step in qualifying the nature of these deficits with a primary focus on eating disorders. I now turn to the discussion of individual covariates, and why they were included in the final model for this study.

1.2.1. Depression

I included depression primarily due to its prevalence in the population, as well as the large literature which focuses on this pathology specifically. Estimates for the 12-month and lifetime prevalence of MDD are approximately 10% and 20%, respectively (Hasin at al., 2018; Kessler et al., 2005). In a recent review of major depressive disorder, a widespread deficit in emotion recognition ability was found across all six categories, with happiness producing the largest effect size (Krause et al., 2021). This review assessed 25 studies which contained a facial emotion recognition task, validated stimuli, and an adequate assessment of major depression disorder or its symptoms. The link between depression and emotion recognition deficits is well established in the literature. What might be the mechanism by which these deficits manifest? Because of the high comorbidity between eating disorders and depression, any mechanism which underlies emotion recognition deficits in one disorder may also explain deficits stemming from the other disorder.

A number of relatively recent studies offer a two-part explanation for emotion recognition deficits in depression – people with depression find emotions to be more complex, and avoid thinking about them to a greater extent than controls. For example, Stroud et al. (2018) administered two separate doses of psilocybin to 17 patients, and compared their accuracy and reaction time to controls on a facial emotion recognition task before and after the treatment. The authors found that reaction time differences were reduced below statistical significance, such that patients with depression were not different from controls after treatment. On the other hand, overall emotion recognition accuracy was not different between the groups at either timepoint (though patients were significantly worse at recognizing anger). In combination, these results point to patients' increased ease of processing emotional stimuli after treatment. Torres et al. (2015) had 80 patients with an anorexia nervosa diagnosis complete questionnaires about depression and alexithymia. The authors then evaluated to what extent depressive symptoms mediated the relationship between alexithymia symptoms and an anorexia diagnosis. Their conclusion was that depression partially mediates the relationship, and that controlling for depression significantly reduces the frequency of responses related to the difficulty of recognizing own emotions. Finally, Zwick and Wolkenstein (2017) administered an emotion recognition task to 42 acutely depressed participants, and asked participants for confidence and difficulty ratings after their responses. The authors found that acutely depressed participants were less confident about recognizing angry and happy stimuli, and found the same stimuli (as well as fear) to be more difficult to recognize than controls. On the other hand, accuracy of the same participants was only significantly lower for happy stimuli.

As well as the increased complexity of emotional stimuli, depression involves people avoiding their emotions to a greater extent. Wildes et al. (2009) administered questionnaires about emotion avoidance, depression, anxiety, and eating behaviors to 81 participants with anorexia nervosa. Similarly to Torres et al. (2015), the authors conducted a mediational analysis, in which they concluded that the relationship between depressive symptoms and anorexia nervosa was mediated by emotion avoidance symptoms. Keating et al. (2018) followed participants with binge eating behaviors via ecological momentary assessment, and assessed depressive episodes as well as emotional avoidance. The authors found that binge eating was predicted by a prior depressive episode, as well as emotion avoidance (particularly in a subgroup of participants with high attachment anxiety). The current study is well-equipped to determine the degree to which depression contributes the emotion recognition deficits in people with eating disorder symptoms. In addition, I am also able to consider the mechanisms by which this contribution may occur, since I include measures of emotion regulation and alexithymia, and due to the increased explanatory scale of my behavioral tasks.

1.2.2. Anxiety

Anxiety was included because of its prevalence, and similarly to depression its potential to bias emotion recognition. In a review by Demenescu et al. (2010), five studies of adult anxiety were meta-analyzed, and the resulting effect size was -0.35, suggesting a sizable deficit in emotion recognition of individuals with anxiety. The mechanisms of this deficit are likely related to attentional biases common in anxiety, such as hyper-vigilance to threatening stimuli (Richards et al., 2014) and overall negative emotion bias (Joormann & Gotlib, 2006; Surcinelli et al., 2006). These biases often result in individuals with anxiety perceiving ambiguous or neutral emotions as negative, and needing less emotional intensity for negative emotions to be perceived. A more direct look at the underlying mechanisms reveals both biological and socio-cognitive differences in people with anxiety, such as increased reactivity of systems which lead to perceived arousal and the fear of social interactions where one is impaired (Tseng et al., 2014; White et al., 2014).

I am able to provide support for these mechanisms since tasks in the current study measure visual information use and internal representations of individuals. If these are biased towards negative emotions, these will be perceived more effectively and represented more clearly in participants with increased anxiety symptoms. Similarly to depression, my model for this study allows for a definitive test of whether anxiety contributes to emotion recognition deficits, and which mechanisms it shares with eating disorders.

1.2.3. Alexithymia

Alexithymia was included primarily because this common disorder (1 in 10 people, Poquérusse et al., 2018) is particularly problematic in conjunction with other mental disorders. Since alexithymia involves the inability to recognize or manage one's own emotions, this lack of ability will come into play when own emotions are affected by psychopathology. Alexithymia also involves impaired recognition of others' emotions, and seems to have a connection to eating disorder symptoms (Brewer et al., 2015; Fujiwara et al., 2017). However, because mechanisms underlying alexithymia-related deficits are unclear, previous researchers could not agree on whether alexithymia is the primary driver of emotion recognition deficits (Brewer et al., 2015), or is only relevant when seen in conjunction with other disorders, such as depression and anxiety (Montebarocci et al., 2006; Parling et al., 2010), autism (Bird et al., 2011), or eating disorders (Fujiwara et al., 2017). The inclusion of common comorbidities of eating disorders in the model which is used in the current study will provide a clearer account of how alexithymia relates to these disorders, and whether it alone is sufficient for deficits in emotion recognition.

1.2.4. Emotion Regulation

I also included a measure of emotion regulation ability, as deficits in emotion regulation are commonly seen in eating disorders and depression (Mason et al., 2020). For example, Kim et al. (2018) used the Difficulties in Emotion Regulation Scale with 31 patients who have bulimia nervosa. The authors found increased non-acceptance of emotional responses, increased difficulties with engaging in goal-directed behavior, and more limited access to emotion regulation strategies among the patient group. Tamiya et al. (2018) administered a number of cognitive assessments to patients with anorexia nervosa, which included Mayer-Salovey-Caruso Emotional Intelligence Test's Managing Emotions component (Mayer et al., 2003; DeTore et al., 2018), which measures emotional management and regulation in conflictual situations. The authors found that patients with anorexia nervosa performed significantly worse than healthy controls on this measure, suggesting that patients are unable to effectively regulate their emotions specifically within situations which would require it. In the current study, emotion regulation ability serves as yet another potential explanatory mechanism for poor emotion recognition. If emotion regulation ability is related to symptoms of other disorders and explains variance in visual information use and representation, it provides an additional mechanism by which emotion recognition is impaired in these disorders.

1.3. Overview of the Current Study

The current study aimed to combine relevant mental pathologies from previous work, and connect these to eating disorder severity within the context of visual information use and its representation. I hypothesized that visual information use and its representation during emotion recognition in individuals will explain significant variance in eating disorder symptoms. Performance in these tasks would remain a significant predictor of emotion recognition after other comorbid disorders are controlled for, though some or all of these may explain significant variance in emotion recognition. When considering the potential results of this study, I am able to say that even a completely null result will further advance the literature on eating disorders. If eating disorder symptoms are unrelated to emotion recognition ability, two broad conclusions can be reached. First, the relationship is best explained by one of the other disorders or abilities I consider. In this case, I will make a case for future research in the area to focus on that relationship, and will be able to provide concrete future directions for doing so. Second, the relationship is not significant for my sample of participants. Even though this is unlikely, such a result would still add to perceptual models of how people perceive and represent emotions. In addition, I would still be able to examine any relationships within the symptoms and abilities indexed by the questionnaires. Finally, because I have adequate power to make conclusions about main effects, I would be able to definitively say that future research in the area should focus on other potential mechanisms underlying emotion recognition deficits (such as hormonal changes, developmental trajectories, or stress reactivity).

2. METHODS

2.1. Participants

The participants for the current study were recruited from the NDSU SONA system. A total of 110 individuals participated in the study. Eight participants were excluded for incomplete data in the main visual tasks or the survey questionnaires. A further ten participants were excluded due to technical difficulties during data collection, which prevented me from performing planned analyses on these individuals. Finally, five more individuals were excluded due to malingering (abnormally fast responses - faster than 250ms - for more than 5% of task trials). The final sample which was analyzed for this study consisted of 87 individuals.

Power analysis calculations to determine sample size assumed a linear regression analysis framework. A total sample size of 92 participants would provide a power of ~.80 to detect a small/moderate effect (i.e., $r = .15$, assuming an alpha of 0.05), that is the contribution of a single predictor while controlling for the rest. Therefore, the current study achieved a power of 0.76.

2.2. Materials

2.2.1. Questionnaires

All questionnaires were administered at the end of the study. Participants were encouraged to respond truthfully, and told that there was no right or wrong answer. They were given unlimited time to enter their responses. The individual questionnaire order was randomized for each participant, with demographic information always occurring last.

2.2.1.1. Eating Disorder Diagnostic Scale.

The Eating Disorder Diagnostic Scale (Stice et al., 2000; 2004) is a brief 22-item self-report scale for diagnosing anorexia nervosa, bulimia nervosa, and binge eating disorder. The scale included questions such as *"During the past 6 months have there been times when you felt you have eaten what other people would regard as an unusually large amount of food (e.g., a quart of ice cream) given the circumstances?"* (response options are *YES* and *NO*) and *"How many times per week on average over the past 3 months have you used laxatives or diuretics to prevent weight gain or counteract the effects of eating?"* (response is a number from 0 to 14). The scale responses were coded for the presence of different eating disorders (anorexia nervosa, bulimia nervosa, binge eating). From the point of view of severity, ten different abnormal dietary behaviors are possible, which is the range of responses that I used for the final model. A score of greater than one point implies clinically significant symptoms.

2.2.1.2. State-Trait Anxiety Inventory.

The State-Trait Anxiety Inventory (Spielberger et al., 1983) is a 40-item self-report scale for diagnosing state and trait anxiety. The scale included questions such as "*I worry too much about something that really doesn't matter",* with response options ranging from 1 to 4 (*"Almost never"* to *"Almost always"*). The scale responses were coded for the presence of state and trait anxiety, which are separate constructs. In the current study, I combined all responses into a single measure of anxiety, which was entered into all relevant models. Scores in this measure range from 40-160, with higher values indicating increased anxiety severity.

2.2.1.3. Toronto Alexithymia Scale.

The TAS-20 (Bagby et al., 1994) is a 20-item scale which assesses the presence and severity of alexithymia. The scale includes questions such as *“It is difficult for me to find the right words for my feelings”*, with responses ranging from 1 - 5 (*“Strongly disagree* to *“Strongly agree”*). The questionnaire consists of three subscales: “difficulty identifying feelings”, “difficulty describing feelings” and “externally oriented thinking style”, which were combined into an overall measure of alexithymia for the current study. Scores range from 20 to 100 with higher scores reflecting higher alexithymia, with scores above 52 indicating further increased likelihood of alexithymia.

2.2.1.4. Beck Depression Inventory Second Edition.

Beck Depression Inventory Second Edition (Beck et al., 1996) is a 21-item self-report instrument intended to assess the existence and severity of symptoms of depression as listed in the Statistical Manual of Mental Disorders. The second edition involves a number of label and wording improvements, as well as having responders consider each statement as it relates to the way they have felt for the past two weeks. The scale includes four descriptive statements differing in severity for each of the 21 sub-categories of depression described by Beck, such as *“0 - I am not discouraged about my future”* to *“3 - I feel my future is hopeless and will only get worse”* for Pessimism. Scores range from 0 to 63, with higher values indicating increased likelihood of depression.

2.2.1.5. Difficulties in Emotion Regulation - Short Form.

Difficulties in Emotion Regulation - Short Form (Kaufman et al., 2016) is a shortened version of the Difficulties in Emotion Regulation Scale (Gratz & Roemer, 2004) which involves half the items of the original (from 36 to 18), while retaining over 80% of the variance relative to the full measure. The short form measure was developed via confirmatory factor analysis of the full measure, using data from five separate samples totaling over 500 individuals (while the original measure was developed with a single sample of over 300 individuals). The shortened measure involves 18 responses such as *“When I'm upset, I lose control over my behavior”* to the prompt of *“Please indicate how often the following apply to you.”*, with severity measured on a 5-point Likert scale, ranging from *“Almost never”* to *“Almost always”*. Scores range from 18 – 90, with higher values indicating problems with regulating emotions.

2.2.1.6. Demographic information.

In addition to the above questionnaires, demographic information was collected from participants, covering standard areas such as age, gender, ethnicity, and education. Participants were also asked a yes/no question about their previous diagnoses of an eating disorder, anxiety, depression, and alexithymia.

2.2.2. Visual Recognition Tasks

2.2.2.1. Bubbles Task.

Participants were asked to sit 80cm away from the computer screen. The on-screen instructions outlined the response keys and the nature of the task. Participants were required to make a choice between six emotion categories (anger, disgust, fear, happiness, sadness, surprise) within 2500ms. On each trial, the participant saw an emotional face image, partially obscured by the bubbles mask, as can be seen in Figure 3. I employed the QUEST staircasing procedure (Watson & Pelli, 1983) in order to maintain task difficulty at 75%. This was achieved by increasing the number of bubbles after each correct response, and decreasing the number after each incorrect response in

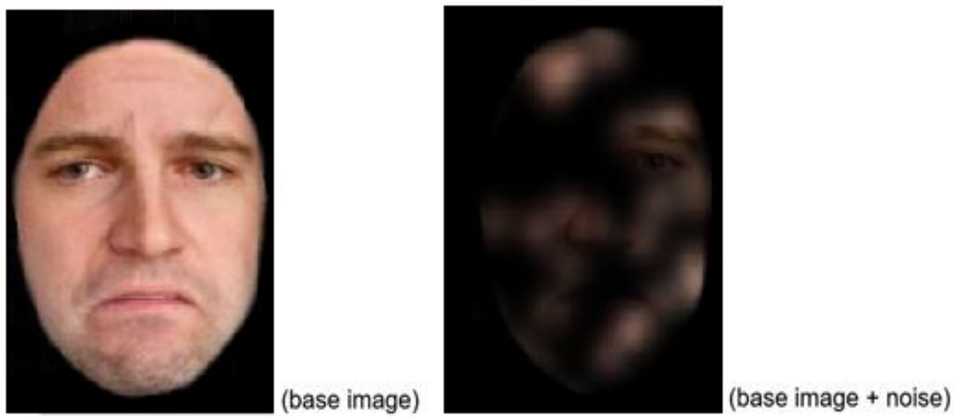

Figure 3. Bubbles Task Stimuli.
A sad face image obscured by a bubbles noise mask as seen during a single trial, with the original image shown on the left.

a manner which provides optimal information as to the number necessary to achieve 75% accuracy. A participant saw 150 trials for each image of an emotion category, for a total of 900 trials, with breaks evenly spaced in-between. A single image for each category was used, with the specific images chosen based on emotion ratings from previous work within the lab.

The procedure for the bubbles task is available on the OSF page for this project. The 900-trial collection of stimuli seen by every participant is available upon request from the first author.

2.2.2.2. Prototype estimation.

Participants were asked to sit 80cm away from the computer screen. The on-screen instructions outlined the response keys and the nature of the task. Participants were required to select one of two images on the screen, according to which one better resembled an emotion category ("which image looks happier?"). Participants were further verbally instructed to "respond with their first impression" and "go with their gut", but there was no time limit for making a response. The experimenter provided feedback if participants were responding exceedingly slow (>5 sec per trial), or exceedingly fast (responding right after a stimulus appeared). The stimulus generation procedure and the stimuli are available on this study's OSF page. Pairs of emotional expressions from the six categories were blended together by linearly interpolating between auto-generated fiducial points on the face via WebMorph (DeBruine, 2018). The most ambiguous images (50% blend of each category pair) were selected. Fractal noise with similar amplitude and phase ranges as the original images was overlaid on top of the images in both an additive and a subtractive matter over 250 random iterations. This image manipulation produced the two images per trial which were seen by a participant; one if which the noise was added to the image, and one in which the noise was subtracted from the image. Due to practical constraints, I selected only a subset of all possible category pairs. Specifically, blends of anger-disgust, anger-happiness, and fear-sadness were used in the experiment. For each blend, a participant saw two blocks of 250 trials each, for a total of 1500 trials. Participants responded to one emotion in the blend in one block, and the other emotion in the blend in the other block. For example, in one block, anger-disgust blends were evaluated according to which looked angrier, and in another block the same images were evaluated according to which looked more disgusted. Each of the two block pairs consisted of the same 250 stimuli, presented in random order. The order of blocks was also randomly determined for each participant. Due to the pilot feedback about boredom, the bubbles task was completed after one or two block pairs (two or four blocks) of the prototype estimation task. Participants were also encouraged to take a break after each block of trials during this task.

I generated prototype visualizations by averaging the noise masks based on participant responses during the prototype estimation task. I then additively overlaid the average noise masks on top of the original ambiguous image, using the same method as for stimulus generation for the original task. The prototype visualizations produced by this method were rated by 108 new individuals. These participants were recruited online from the NDSU SONA system, and the Prolific (www.prolific.co) testing platform. In this task, participants were shown each observer's prototype visualizations, and asked to rate it on a scale from 0-7 according to how well it depicted the emotion it was supposed to represent. For example, for a happiness prototype, participants were asked how happy the image looked on a scale from 0-7, 0 being not at all happy, and 7 being extremely happy. The new participants were instructed to use their first impression when rating the prototype images, and that there was no right or wrong answer. They were given an unlimited amount of time to complete the task.

2.2.2.3. Facial Emotion Recognition.

Participants were asked to sit 80cm away from the computer screen. The on-screen instructions outlined the response keys and the nature of the task. Participants were required to make a choice between six emotion categories (anger, disgust, fear, happiness, sadness, surprise), without a time limit. On each trial, the participant saw a single face depicting an emotion from one of the six categories. For each category, there were two unique identities, as well as a third identity which was the blended average of the other two. A participant saw 10 repetitions of each unique identity, with 30 repetitions per emotion category, for a total of 180 trials. This task was always completed after the other visual tasks in order to avoid familiarizing participants with the stimuli which were used in the other tasks.

All questionnaires and behavioral tasks, stimulus sets, and other supplementary information is included on the OSF page for this study, accessible at the following link: https://osf.io/fh9vd/?view_only=bd5533cd19d946499e4b17337694bbd0.

Other information about this study is available upon request from the first author.

3. RESULTS

3.1. Visual Information Use and Representation

Prior to testing my main hypotheses, I first attempted to validate whether the bubbles and prototype estimation tasks worked as intended. One way to do this for the bubbles task is to examine the bubbles visualization images in terms of which facial features people use when making a correct response. If the task worked as intended, people should have used information around the eyes, mouth, and other features that change across emotions to a greater extent than other regions of the face. Such a pattern of results within the bubbles task has been found in previous studies using facial emotion stimuli (Blais et al., 2012, Blais et al., 2017). These visualizations are depicted in Figure 4. The regions highlighted in Figure 4 also map onto action units derived from common expressions of emotion (Ekman & Friesen, 1978; Mehu & Scherer, 2015). For example, the anger visualization maps onto the *AU4 – Brow Lowerer* in Table 4 of Mehu & Scherer's manuscript, and the surprise visualization neatly maps onto AU5 in the same table. Therefore, the bubbles task captures valid patterns of information use within the current sample of individuals.

The prototype estimation task also produced visualization images, depicting people's internal prototypes of a particular emotion. I presented these images to new participants and asked them to rate these images according to how well they depicted the emotion they were meant to represent. Individual category images were rated with the following ranges of responses: anger (disgust blend) = 2.47 - 5.02, anger (happiness blend) = 1.17 - 3.86, disgust = 3.85 - 5.5, fear = 2.26 - 5.52, happiness = 1.82 - 5.09, sadness = 3.77 - 5.01. All rating ranges either include or exceed the scale midpoint,

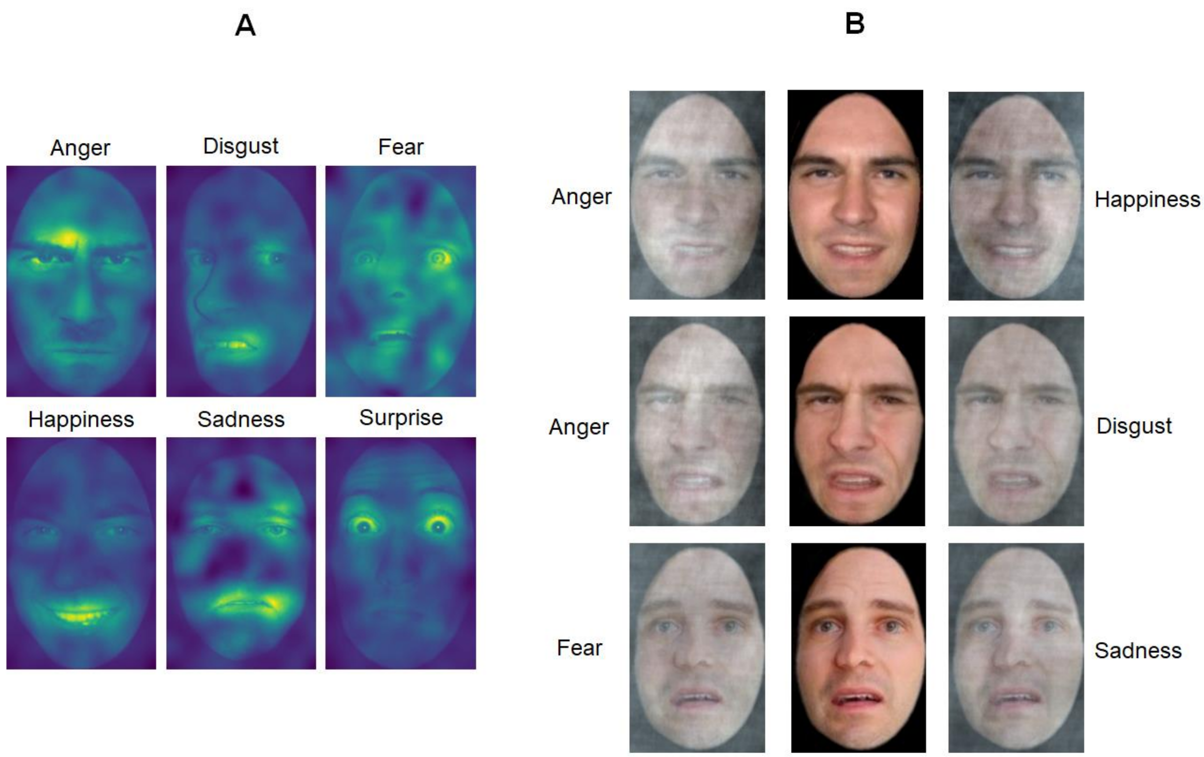

Figure 4. Visualizations of Visual Information Use and Representation
(A) between-participant average information use maps, overlaid on top of the six emotion category stimuli used in the bubbles task. Brighter color reflects greater importance of the pixels for emotion categorization.
(B) emotion prototypes of the three top-rated participants, averaged into a single image. The original ambiguous image (before the addition of noise) is included for comparison.

suggesting that most images were judged to be at least moderately expressive. Participant agreement, computed via Cronbach's alpha, was high for all categories (α > 0.8) except for sadness (α = 0.65). Visual inspection of the averaged top three prototypes depicted in Figure 4 reveals a marked difference across emotion categories. The combination of the above evidence is enough for me to conclude that the prototype estimation task does indeed capture valid internal representations of emotion categories.

3.2. Data Analytic Approach

Multiple linear regression was used in order to examine the influence of visual task performance on eating disorder symptom severity. This analysis was conducted in structured steps, and all models are nested with the same *N*. In conducting model comparisons, I looked at overall model fit and also interpreted individual parameters. These steps were performed with variables from a single task at a time. In step 1, only variables representing overall visual task performance averaged across emotion conditions were entered into the model. In step 2, variables representing symptoms of comorbid mental disorders were added to the model from step 1. Steps 1 and 2 were repeated for each visual task (bubbles, prototype, and six-alternative-forced-choice categorization). These steps constitute the main tests of my hypotheses.

Additional analyses were then conducted to explore additional relationships of interest. While the current study is underpowered to make strong conclusions about exploratory results, these may provide insight into potential future directions of this research. Specifically, one additional analysis involved the separation of visual task performance into individual emotion categories within each task. A regression was then conducted with the six category-specific variables replacing the average task performance. I decided not to add symptoms to category-specific variables as power for such an analysis generally fell below 60%. Therefore, the category specific models do not consider the contribution of comorbid disorder symptoms. A final exploratory analysis involved the separation of eating disorder symptom scores into three categories, representing anorexia nervosa, binge eating, and bulimia. I then re-ran the models from steps 1 and 2, using an eating disorder category as the criterion instead of overall symptom severity.

Finally, I did two robustness checks. The first evaluated regression assumptions and multivariate outliers. This check was performed for all final regression models, and the complete analysis pipeline for this study is available on its OSF page. After running

	1	2	3	4	5	6	7	8
average bubbles correlations (1)	-	-0.19	0.33 **	-0.23 *	-0.07	0.06	0.05	-0.02
average prototype ratings (2)	-0.19	-	-0.13	-0.04	-0.02	0.02	-0.09	-0.01
FER accuracy (3)	0.33 **	-0.13	-	-0.1	-0.07	-0.14	-0.11	-0.19
EDDS (4)	-0.23 *	-0.04	-0.1	-	0.45 **	0.33 **	0.39 **	0.47 **
STAI (5)	-0.07	-0.02	-0.07	0.45 **	-	0.61 **	0.71 **	0.78 **
TAS (6)	0.06	0.02	-0.14	0.33 **	0.61 **	-	0.73 **	0.61 **
DERS (7)	0.05	-0.09	-0.11	0.39 **	0.71 **	0.73 **	-	0.72 **
BECK (8)	-0.02	-0.01	-0.19	0.47 **	0.78 **	0.61 **	0.72 **	-
variable means	0.36	3.95	0.81	2.49	88.82	52.29	44.71	14.52
variable standard deviations	0.07	0.2	0.1	2.06	24.18	9.85	13.44	10.37

Figure 5. Zero-order correlations between study variables and descriptive statistics. 1: the average correlation between a bubbles visualization image of a participant and an ideal observer, 2: the average rating of a participant's emotion prototypes, 3: the average accuracy of a participant during the six-alternative-forced-choice emotion categorization task. EDDS: Eating Disorder Diagnostic Scale, STAI: State Trait Anxiety Inventory, TAS: Toronto Alexithymia Scale, DERS: Difficulties in Emotion Regulation Scale, BECK: Beck Depression Inventory. ** $p < 0.01$, * $p < 0.05$

models in the full sample, I computed values for leverage, Cook's distance, and studentized residuals in order to test for multivariate outliers in all final models. I excluded participants if they exceeded the cut-off on at least two of these metrics. I examined linearity with loess plots. Homoscedasticity was assessed with a fitted value vs standardized residual plot, as well as the non-constant variance score test and the Breusch-Pagan test. Normality in the residuals was assessed with a residual qq-plot and the Shapiro-Wilk test.

The second examined whether results are sensitive to the assessment of eating disorder symptom. Primary analyses use the EDDS continuously, but sensitivity analyses follow clinical coding, where a 0 means no eating disorder pathology, a 1 means possibly mild pathology, a 2 means definite mild pathology, and scores of 3 or more means definite severe pathology. I then re-ran the models from step 2 while using categorical EDDS as the criterion variable in multinomial logistic regression. RStudio (RStudio Team, 2022) was used for all analyses. All predictors, including questionnaires, were mean-centered. EDDS was not centered.

3.3. Bivariate Correlations and Distributions

Zero-order correlations between study variables are depicted in Figure 5. EDDS severity was below the threshold for subclinical symptom severity on average (M = 0.81 < 1). With a range of 11 severity levels on the questionnaire (0-10), the current sample spanned 9 levels of severity (0-8). Seventeen individuals had zero symptoms, 16 had one level of severity, and 54 had at least two severity levels (which constitutes clinically-significant symptoms).

3.4. Main Hypotheses

3.4.1. Bubbles and Overall Eating Disorder Symptoms

Visual information use was computed as the Spearman ranked correlation between bubbles visualization images of a participant and an ideal observer for the same image category. The contribution of overall visual information use to eating disorder symptom severity was assessed in step 1. This regression model was statistically significant (Adj. R^2 =.04, $F(1, 85) = 4.85$, $p = .03$), and overall visual information use significantly predicted eating disorders severity ($B = -0.07$, $p = .03$). In step 2, the addition of comorbid disorder symptoms also resulted in a statistically significant model, Adj. $R^2 = .24$, $F(5, 81) = 6.48$, $p < .0001$, which is included in Figure 6. This model was a significant improvement over the model from step 1 (Df = 85 vs 81, $F(4, 81) = 6.57$, p = .0001). In this model, overall visual information use remained a statistically significant predictor of overall eating disorders symptoms, even after adjustment for comorbid disorder symptoms (B = -0.06, p = .02).

I evaluated the robustness of these results using two follow-up sensitivity analyses. First, I examined regression assumptions and removed three multivariate outliers from step 2, resulting in a model N of 84. Results of this model were not substantially different from the results of step 2 in the full sample. A second robustness check evaluated whether associations were sensitive to using the EDDS continuously or categorically, as in clinical practice. Thus, I categorized overall EDDS symptoms and re-ran Models 1 and 2 in multinomial logistic regressions, with “0” symptoms (i.e., no psychopathology) as the reference group. Results were sensitive to the coding of EDDS. Step 2 fit significantly better than step 1 ($\chi^2(12) = 38.46$, p = .0001), however, none of the predictors significantly differentiated the groups of possible mild psychopathology (1), definite mild psychopathology (2), and definite severe psychopathology (3) from the reference group of no eating disorders pathology (0). Considering that the overall models detected effect sizes which we are underpowered for, coupled with possible differences in results due to the coding of EDDS, the results of these models suggest a possible, albeit weak or not robust association between bubbles task performance and overall eating disorders symptoms.

3.4.1.1. Category-Specific Bubbles Performance

I examined the contribution of visual information use for individual categories to eating disorder symptom severity. The model including performance for individual categories did not reach statistical significance (Adj. R^2 = -0.01, $F(6, 79)$ = .793, p = .579). One multivariate outlier was removed from analysis as part of a robustness check resulting in a model N of 86, but overall model fit or individual parameters were not

	Overall EDDS	Anorexia Nervosa	Bulimia Nervosa	Binge Eating
Model R^2	0.242 **	0.122 **	0.142 **	0.195 **
Predictor Variable	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
Intercept	2.49 (*0.19*) **	1.24 (*0.09*) **	0.93 (*0.09*) **	1.06 (*0.12*) **
Overall visual information use	-0.06 (*0.03*) *	-0.03 (*0.01*) *	-0.02 (*0.01*)	-0.02 (*0.02*)
STAI	0.01 (*0.01*)	0.00 (*0.01*)	0.00 (*0.01*)	0.01 (*0.01*)
TAS	0.01 (*0.03*)	0.00 (*0.01*)	0.00 (*0.01*)	0.00 (*0.02*)
DERS	0.01 (*0.03*)	0.01 (*0.01*)	0.01 (*0.01*)	-0.00 (*0.02*)
BECK	0.06 (*0.03*) t	0.01 (*0.02*)	0.02 (*0.02*)	0.03 (*0.02*) t

Figure 6. Regression models for visual information use and comorbid disorder symptoms predicting eating disorder symptoms and its sub-categories.
** p < 0.01, * p < 0.05, t p < 0.10

substantially affected. Visual information use for individual categories was not a significant predictor of eating disorder symptom severity.

3.4.1.2. Bubbles and Specific Eating Disorder Symptom Outcomes:

Anorexia

For anorexia nervosa, the contribution of overall visual information use alone resulted in a significant model (Adj. R^2 = .04, $F(1, 85) = 4.775$, p = .03) and overall visual information use significantly predicted anorexia severity (B = -0.03, p = .03). The addition of comorbid disorder symptoms also resulted in a statistically significant model (see Figure 6). This model was a significant improvement over the model from step 1 (*Df* = 85 vs 81, $F(4, 81) = 2.93$, p =.03), and overall visual information use remained a significant predictor of anorexia severity (B = -0.03, p = .03).

Two multivariate outliers were removed from analysis as part of a robustness check (new *N* = 85), but neither the overall model fit nor individual beta weights were not affected. Therefore, overall visual information use remained a significant predictor of anorexia symptom severity, even after controlling for comorbid disorder symptoms. However, this effect size is below the anticipated minimum effect size from the power analysis. The results of this model again suggest a possible, albeit weak or not robust association between bubbles task performance and anorexia symptoms.

I also examined the contribution of visual information use for individual categories to anorexia symptom severity. The model including performance for individual categories did not reach statistical significance. No multivariate outliers were removed from this model as part of a robustness check. Visual information use for individual categories was not a significant predictor of anorexia symptom severity.

3.4.1.3. Bubbles and Specific Eating Disorder Symptom Outcomes: Bulimia

For bulimia nervosa, the contribution of overall visual information alone did not reach statistical significance (Adj. R^2 = .01, $F(1, 85) = 1.73$, $p = .192$). The addition of comorbid disorder symptoms did result in a statistically significant model, depicted in Figure 6 (Adj. R^2 = .14, $F(5, 81) = 3.86$, $p = .003$). However, visual information remained a non-significant predictor of bulimia symptoms ($B = -0.02$, $p = .15$). This model fit significantly better than Step 1 (df = 85 vs 81, $F(4, 81) = 4.32$, $p = .003$) because of the inclusion of comorbid psychopathology and not the visual information.

Two multivariate outliers were removed from analysis as part of a robustness check (new model *N* = 85), but this did not substantially affect overall model fit or individual parameters. Multiple assumptions were violated for this model, which resulted in me transforming the bulimia variable via a square root transformation. However, this modification did not affect the significance of either of the above models. Visual information use in this model was still not a significant predictor of bulimia symptom severity.

I also examined the contribution of visual information use for individual categories to bulimia symptom severity. The model including performance for individual categories did not reach statistical significance. Two multivariate outliers were removed from analysis as part of a robustness check (new model $N = 85$), but this did not substantially affect overall model fit or individual parameters. Therefore, neither overall or category-specific visual information use significantly predicted bulimia symptom severity.

3.4.1.4. Bubbles and Specific Eating Disorder Symptom Outcomes: Binge Eating

For binge eating, the contribution of overall visual information did not reach statistical significance (Adj. $R^2 = .01$, $F(1, 85) = 2.29$, $p = .13$) . The addition of comorbid disorder symptoms did result in a statistically significant model, depicted in Figure 6 (Adj. $R^2 = .20$, $F(5, 81) = 5.18$, $p = .0004$). However, visual information was once again a non-significant predictor of binge eating disorder symptoms ($B = -0.02$, $p = .14$). This model fit significantly better than Step 1 ($df = 85$ vs 81, $F(4, 81) = 5.77$, $p = .0004$) because of the inclusion of comorbid psychopathology and not the visual information.

Two multivariate outliers were removed from analysis as part of a robustness check (new model $N = 85$). This robustness check revealed that the removal of outliers resulted in a significant model containing just visual information use (Adj. $R^2 = .04$, F(1,

85) = 4.775, p = .03), and overall visual information use significantly predicted binge eating severity (B = -0.04, p = .036) in the model containing comorbid disorder symptoms. Therefore, overall visual information use remained a significant predictor of binge symptom severity, even after controlling for comorbid disorder symptoms. However, as with other visual information use effects, the magnitude of this effect is below what I was adequately powered for. Therefore, the results of these models again only suggest a possible but weak association between bubbles task performance and binge eating symptoms.

I also examined the contribution of visual information use for individual categories to binge eating symptom severity. The model including performance for individual categories did not reach statistical significance. Two multivariate outliers were removed from analysis as part of a robustness check (new model N = 85), but this did not substantially affect overall model fit or individual parameters. Visual information use for individual categories was not a significant predictor of binge eating symptom severity.

3.4.2. Prototype Estimation

Prototype fidelity was computed as the average rating of the prototype image but a sample of new participants on a scale of 0 – 7. The contribution of average prototype fidelity to eating disorder symptom severity was assessed in step 1. This regression model was not significant (Adj. R^2 = -0.01, F(1, 85) = .125, p = .724). In step 2, the addition of comorbid disorder symptoms resulted in a statistically significant model, Adj. R^2 = .19, F(5, 81) = 5.057, p = .0004, which is depicted in Figure 7. However, average prototype fidelity remained a non-significant predictor of overall eating disorders symptoms (B = -0.30, p = .77). This model fit significantly better than Step 1 (df = 85 vs 81,

	Overall EDDS	Anorexia Nervosa (transformed)	Bulimia Nervosa (transformed)	Binge Eating
Model R^2	0.325 **	0.067 t	0.171 **	0.209 **
Predictor Variable	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
Intercept	2.44 *(0.18)* **	0.94 *(0.06)* **	0.74 *(0.06)* **	1.03 *(0.11)* **
Average prototype fidelity	-0.52 *(0.90)*	0.02 *(0.33)*	0.15 *(0.30)*	-0.08 *(0.59)*
STAI	0.01 *(0.01)*	0.00 *(0.00)*	0.00 *(0.00)*	0.01 *(0.01)*
TAS	0.03 *(0.03)*	0.00 *(0.01)*	0.01 *(0.01)*	0.01 *(0.02)*
DERS	-0.01 *(0.02)*	0.00 *(0.01)*	0.01 *(0.01)*	-0.01 *(0.02)*
BECK	0.09 *(0.03)* **	0.01 *(0.01)*	0.01 *(0.01)*	0.04 *(0.02)* t

Figure 7. Regression models for emotion prototype fidelity and comorbid disorder symptoms predicting eating disorder symptoms and its sub-categories.
** p < 0.01, * p < 0.05, t p < 0.10

$F(4, 81) = 6.28$, $p = .0002$) because of the inclusion of comorbid psychopathology rather than prototype fidelity.

I once again evaluated the robustness of these results using two follow-up sensitivity analyses. First, I examined regression assumptions and removed three multivariate outliers from step 2, resulting in a model *N* of 84. Results of this model were not substantially different from the results of step 2 in the full sample. A second robustness check evaluated whether associations were sensitive to using the EDDS continuously or categorically, as in clinical practice. In this case, the results of the model containing categorical EDDS were not different from the model in step 2 which contained continuously coded EDDS. Therefore, average prototype fidelity did not significantly contribute to eating disorder symptom severity.

3.4.2.1. Category-Specific Prototype Fidelity

I examined the contribution of prototype fidelity of individual categories to eating disorder symptom severity. The model including fidelity for individual categories was not statistically significant (Adj. R^2 = -0.01, F(6, 80) = .877, p = .516). Three multivariate outliers were removed from analysis as part of a robustness check, but the overall model and individual beta weights were not affected. Prototype fidelity of individual categories was not a significant predictor of eating disorder symptom severity.

3.4.2.2. Prototypes and Specific Eating Disorder Symptom Outcomes: Anorexia

For anorexia nervosa, the contribution of average prototype fidelity alone did not result in a significant model (Adj. R^2 = 0, F(1, 85) = 0.12, p = .73).. The model including comorbid disorder symptoms also failed to reach statistical significance (see Figure 7). In this model, prototype fidelity remained a nonsignificant predictor of anorexia symptoms (B = -0.11, p = .82). Although this model fit significantly better than step 1 (df = 85 vs 81, F(4, 81) = 2.76, p = .03), this is because of the inclusion of psychopathology rather than prototype fidelity.

Considering sensitivity analyses, only multivariate normality was moderately violated. For this reason, I performed a square root transform of anorexia severity. This modification did not affect the significance of either model. Average prototype fidelity was not a significant predictor of anorexia symptom severity.

I also examined the contribution of prototype fidelity of individual categories to anorexia symptom severity. The model including fidelity of individual categories did not reach statistical significance. One multivariate outlier was removed from analysis as part of a robustness check (new model *N* = 86), but this did not substantially affect overall model fit or individual parameters. Prototype fidelity of individual categories was not a significant predictor of anorexia symptom severity.

3.4.2.3. Prototypes and Specific Eating Disorder Symptom Outcomes: Bulimia

For bulimia nervosa, the contribution of average prototype fidelity alone did not result in a significant model (Adj. R^2 = 0, $F(1, 85) = 0.02$, $p = .88$). The addition of comorbid disorder symptoms resulted in a statistically significant model, depicted in Figure 7. In this model, prototype fidelity remained a non-significant predictor of bulimia symptoms (B = 0.17, $p = .72$). Although this model fit significantly better than Step 1 (df = 85 vs 81, $F(4, 81) = 4.20$, $p = .003$), this is once again because of the inclusion of psychopathology rather than prototype information.

Considering sensitivity analyses, two multivariate outliers were removed from analysis as part of a robustness check, resulting in a model *N* of 85. Multiple assumptions were violated for this model, which again resulted in me transforming the bulimia variable via a square root transformation. However, neither modification affected the significance of either of the above models. Average prototype fidelity was not a significant predictor of bulimia symptom severity.

I also examined the contribution of prototype fidelity for individual categories to bulimia symptom severity. The model including performance for individual categories was statistically significant (Adj. R^2 = .08, $F(6, 80) = 2.30$, p =.04). One multivariate outlier was removed from analysis as part of a robustness check, resulting in a model *N* of 86. I also transformed the bulimia variable via a square root transformation due to the violation of multiple assumptions. The removal of an outlier did not substantially affect model fit or individual parameter estimates. However, while the transformation was effective in meeting assumptions, this resulted in a statistically non-significant model (Adj. R^2 = .06, F(6, 79) = 1.96, p = .08). Due to the already low power and this being an exploratory analysis, a model which no longer violates assumptions but becomes non-significant must be interpreted as containing an inconsistent effect. Therefore, I conclude that the contribution of protype fidelity of individual categories is unlikely to be significant for bulimia nervosa.

3.4.2.4. Prototypes and Specific Eating Disorder Symptom Outcomes: Binge Eating

For binge eating, the contribution of average prototype fidelity alone did not result in a significant model (Adj. R^2 = 0, F(1, 85) = 0.002, p = .96). The addition of comorbid disorder symptoms resulted in a statistically significant model, depicted in Figure 7. In this model, prototype fidelity remained a non-significant predictor of binge eating symptoms (B = -0.01, p = .99). Once again, although this model was a significant improvement over Step 1 (df = 85 vs 81, F(4, 81) = 5.77, p = .0004), this is because of the inclusion of comorbid psychopathology rather than prototype fidelity.

Considering sensitivity analyses, one multivariate outlier was removed from analysis as part of a robustness check, resulting in a model N of 86. The removal of an outlier did not substantially affect model fit or individual parameter estimates.

I also examined the contribution of prototype fidelity of individual categories to binge eating symptom severity. The model including performance for individual categories did not reach statistical significance. One multivariate outlier was removed from analysis as part of a robustness check, resulting in a model *N* of 86. The removal of an outlier did not substantially affect model fit or individual parameter estimates. Prototype fidelity of individual categories was not a significant predictor of binge eating symptom severity.

3.4.3. Facial Emotion Recognition

Facial emotion categorization accuracy was computed as the average accuracy for each participant on a six-alternative-forced-choice categorization test. The contribution of average categorization accuracy to eating disorder symptom severity was assessed in step 1. This regression model was not statistically significant (Adj. R^2 = -0.00, $F(1, 85)$ = .858, p = .357). In step 2, the addition of comorbid disorder symptoms resulted in a statistically significant model, Adj. R^2 = .19, $F(5, 81)$ = 8.87, p = .0004 (Figure 8). This model fit significantly better than step 1 (df = 85 vs 81, $F(4, 81)$ = 6.05, p = .0003) because of the inclusion of comorbid psychopathology. In this model, categorization accuracy remained a non-significant predictor of eating disorder severity (B = -0.01, p = .78).

I evaluated the robustness of these results using two follow-up sensitivity analyses. First, I examined regression assumptions and removed four multivariate outliers from step 2, resulting in a model *N* of 83. Results of this model were not substantially different from the results of step 2 in the full sample. A second robustness check evaluated whether associations were sensitive to using the EDDS continuously or categorically, as in clinical practice. Results were mildly sensitive to the coding of EDDS. Step 2 fit significantly better than step 1 ($\chi^2(12)$ = 36.18, p = .0003), with average categorization accuracy significantly differentiating the group of definite mild psychopathology (2) from the reference group of no eating disorders pathology (0; B = .08, p = .047), while controlling for the contribution

	Overall EDDS	Anorexia Nervosa	Bulimia Nervosa (transformed)	Binge Eating
Model R^2	0.324 **	0.080 *	0.199 **	0.174 **
Predictor Variable	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
Intercept	2.39 (*0.17*) **	1.24 (*0.09*) **	0.71 (*0.06*) **	1.04 (*0.12*) **
Average categorization accuracy	0.01 (*0.02*)	0.01 (*0.01*)	0.01 (*0.01*) t	-0.00 (*0.01*)
STAI	0.01 (*0.01*)	0.00 (*0.01*)	0.00 (*0.00*)	0.01 (*0.01*)
TAS	0.02 (*0.03*)	0.00 (*0.01*)	0.01 (*0.01*)	0.00 (*0.02*)
DERS	-0.01 (*0.02*)	0.01 (*0.01*)	-0.00 (*0.01*)	0.01 (*0.02*)
BECK	0.10 (*0.03*) **	0.01 (*0.02*)	0.02 (*0.01*)	0.03 (*0.02*)

Figure 8. Regression models for emotion categorization accuracy and comorbid disorder symptoms predicting eating disorder symptoms and its sub-categories.
** p < 0.01, * p < 0.05, t p < 0.10

of other comorbid disorders. No other groups were differentiated from the reference group. Considering that the continuous EDDS model was not able to detect a contribution of average categorization accuracy to EDDS, our relatively low power, and difficulty differentiating all EDDS severity groups from the no symptom group, these results suggest little-to-no contribution of categorization accuracy to EDDS.

3.4.3.1. Category-Specific Categorization Accuracy

I examined the contribution of categorization accuracy for individual categories to eating disorder symptom severity. The model including accuracy for individual categories did not reach statistical significance (Adj. R^2 = -0.01, $F(6, 80)$ = .805, p = .569). Five multivariate outliers were removed from analysis as part a robustness check, but this did not affect overall model fit or individual beta weights. Categorization accuracy for individual categories was not a significant predictor of eating disorder symptom severity.

3.4.3.2. Facial Emotion Recognition and Specific Eating Disorder Symptom Outcomes: Anorexia

For anorexia nervosa, the contribution of average categorization accuracy alone did not reach statistical significance, Adj. $R^2 = 0$, $F(1, 85) = 0.33$, $p = .57$. The addition of comorbid disorder symptoms resulted in a statistically significant model (see Figure 8). This model fit significantly better than step 1 ($df = 85$ vs 81, $F(4, 81) = 3.02$, $p = .02$) because of the inclusion of comorbid psychopathology symptoms. Average categorization accuracy remained a non-significant predictor of anorexia symptoms in this model ($B = 0.01$, $p = .29$).

Considering sensitivity analyses, no multivariate outliers were detected in this model. Average categorization accuracy was not a significant predictor of anorexia symptom severity.

I also examined the contribution of categorization accuracy for individual categories to anorexia symptom severity. This model did not reach statistical significance. Three multivariate outliers were removed from analysis as part of a robustness check, resulting in a model *N* of 84. This model also involved a deviation from multivariate normality, which resulted in me performing a square root transform on the anorexia variable. However, neither outlier removal or the transformation substantially affected overall model fit or individual estimates. Categorization accuracy for individual categories was not a significant predictor of anorexia symptom severity.

3.4.3.3. Facial Emotion Recognition and Specific Eating Disorder Symptom Outcomes: Bulimia

For bulimia nervosa, the contribution of average categorization accuracy alone did not reach statistical significance, Adj. R^2 = 0, F(1, 85) = 0.93, p = .34. The addition of comorbid disorder symptoms resulted in a statistically significant model (see Figure 8). This model fit significantly better than Step 1 (df = 85 vs 81, F(4, 81) = 3.95, p = .006) because of the inclusion of comorbid psychopathology. In this model, average categorization accuracy remained a non-significant predictor of bulimia symptoms (B = -0.003, p = .70).

Four multivariate outliers were removed from analysis as part of a robustness check, resulting in a model N of 83. Due to violation of multiple assumptions, I once again transformed the bulimia variable via a square root transform. However, neither outlier removal or the transformation substantially affected overall model fit or individual estimates. Average categorization accuracy was not a significant predictor of bulimia symptom severity.

I also examined the contribution of categorization accuracy for individual categories to bulimia symptom severity. This model did not reach statistical significance. Seven multivariate outliers were removed from analysis as part of a robustness check, resulting in a model N of 80. This model also involved a deviation from multivariate normality, which resulted in me performing a square root transform on the bulimia variable. However, neither outlier removal or the transformation substantially affected overall model fit or individual estimates. Categorization accuracy for individual categories was not a significant predictor of bulimia symptom severity.

3.4.3.4. Facial Emotion Recognition and Specific Eating Disorder Symptom Outcomes: Binge Eating

For binge eating, the contribution of average categorization accuracy alone was not statistically significant, Adj. $R^2 = 0$, $F(1, 85) = 1.16$, $p = .29$). The addition of comorbid psychopathology symptoms resulted in a statistically significant model (see Figure 8). This model fit significantly better than Step 1 ($df = 85$ vs 81, $F(4, 81) = 5.40$, $p = .0007$) because of the inclusion of comorbid psychopathology symptoms. In this model, average categorization accuracy remained a non-significant predictor of binge eating disorder symptoms ($B = -0.01$, $p = .37$).

One multivariate outlier was removed from analysis as part of a robustness check, resulting in a model *N* of 88. The removal of an outlier did not substantially affect model fit or individual parameter estimates. Average categorization accuracy was not a significant predictor of binge eating symptom severity.

I also examined the contribution of categorization accuracy for individual categories to binge eating symptom severity. This model did not reach statistical significance. Three multivariate outliers were removed from analysis as part of a robustness check, resulting in a model *N* of 84. This model also involved a deviation from multiple assumptions, which resulted in me performing a square root transform on the binge eating variable. However, neither outlier removal or the transformation substantially affected overall model fit or individual estimates. Categorization accuracy for individual categories was not a significant predictor of binge eating symptom severity.

4. DISCUSSION

I was interested in examining the contribution of visual emotion recognition ability to the severity of eating disorder symptoms in a sub-clinical sample of individuals. I examined how well individuals used the visual information which was available on an emotional image, how well individuals represented emotional categories, and how accurate individuals were at categorizing facial emotions. Almost no single metric of emotion recognition as indexed using visual tasks was a direct predictor of eating disorder symptom severity. Only overall visual information use was a significant predictor, and remained so when comorbid disorder symptoms were added to the model. However, the effect size of this estimate fell below the planned minimum detectable effect size, when considering the statistical power of my sample. Overall, the results of this study suggest there is a small, albeit possibly not robust association between visual information use and eating disorder symptomology, over and above other psychopathology symptoms.

Category-specific task performance metrics were not significant predictors of eating disorder severity of its sub-categories. However, it is possible that the effect of task performance in individual categories may occur via comorbid disorder symptoms, as has been found previously. However, the current study does not possess sufficient power to test this assumption via structural equation modeling, and was a cross-sectional experimental design. Future research in this area should consider sample sizes which allow for sufficient power to test potential mediational pathways. Research designs which track visual task and symptom metrics over time would provide an even stronger test of an assumption of mediation.

Metrics based on individual task performance were primarily uncorrelated, suggesting that these tasks assessed unique perceptual and representational abilities of individuals. However, even when considering individual metrics at the category level, no significant direct contributions were found to overall eating disorder symptom severity. Therefore, I conclude that visual emotion recognition ability is unrelated to eating disorder symptom severity in a sub-clinical sample of individuals.

What might be a reason for my null findings, despite previous literature uncovering significant differences? In two previous studies with sub-clinical samples which focused on eating disorder severity, only very modest differences in some categories (Ridout et al., 2012) or no differences (Vander Wal et al., 2020) were found. The bulk of the remaining literature has focused on patients with eating disorders, primarily women with anorexia nervosa compared to healthy age-matched controls. While many of these studies did find differences in facial emotion recognition between patients and controls, a substantial number of studies (Brewer et al., 2015; Phillipou et al., 2015; Sfärlea et al., 2016), including large scale recent work (Wyssen et al., 2019), also found null results. Wyssen et al. (2019) suggest that the difficulty of patients may be with interpreting what emotions mean within a social context, rather than the recognition of the emotion itself. My study confirms that recognition, representation, and focus on specific features of emotions do not differ across eating disorder symptom severity. Rather, the difficulty in social interactions exhibited by the more disordered individuals is likely to be related to a latter process within social emotional perception.

One candidate process which may affect social perception after emotion recognition is an individual's emotion regulation abilities. In the current study, beliefs about emotion regulation and the individual's own abilities as indexed by the DERS questionnaire was related to symptom severity of all eating disorders which were examined. Additionally, emotion regulation abilities may be related to an individual's visual information use, as can be seen from the zero-order correlations in Figure 5; and here, we adjusted for this possible pathway in Step 2, possibly reducing associations. According to Mason et al., (2020), emotion regulation abilities are a trans-diagnostic marker of psychopathology. Beyond relying on trait-like constructs which are most easily derived from questionnaires, future work may benefit from directly evaluating the effectiveness of different emotion regulation strategies. More complex experimental paradigms may combine emotion recognition and regulation by directly inducing an emotional state to be regulated.

The current study builds on previous literature to provide a more extensive test of facial emotion recognition than seen in purely accuracy-based designs. Both the bubbles task and the prototype estimation task provide unique insights into a person's emotion recognition abilities. The bubbles task provided a valid description of an individual's visual information use, and the prototype estimation task provided a valid representation of an individual's internal emotion prototypes. While overall performance on these tasks was not related to eating disorder symptom severity, these tasks can continue to be used to uncover additional unique abilities of individuals when performing emotion categorization. Additionally, data gathered from these tasks can be further analyzed with additional hypotheses in mind, and will be made freely available online. For example, potential differences in prototype visualizations along dimensions of valence or intensity may be related to symptom severity of eating or other disorders.

The current study has an important limitation when considering these results within the framework of perceptual difficulties of individuals with eating disorders. The sample for the current study consists of a normal student population, which in a sense is only subclinical due to the large range of disorder symptoms among them. It is possible that a true subclinical sample (for example, individuals on a waitlist for receiving treatment or those at increased risk, as in Pringle et al., 2010) would respond to the questionnaires in a qualitatively different manner. Similarly, a sample of diagnosed eating disorder patients may have an even more experience with their disorder(s), which further increases the validity of their self-reported symptoms. Therefore, the results of this study should only be considered to apply to the general population of neurotypical healthy young adults, and may not directly apply to patient populations which are commonly seen in eating disorder literature. While I did ask individuals to provide information about a previous diagnosis of an eating disorder, there were not enough individuals (eight in total) to make a formal statistical test based on diagnosis status. Future research within this area may benefit from comparing responses between patients and individuals with severe symptomology in a qualitative manner.

Despite this limitation, my study provides a strong data point in favor against considering sub-clinical visual emotion recognition difficulties of people at risk for developing an eating disorder. I believe that it is more fruitful to look into early emotion regulation difficulties in healthy individuals, as these may be a metric of multiple disorders, eating disorder among them. The validation of the perceptual tasks used in this study should encourage further research into individual mechanisms of emotion recognition, and how these may be related to various individual traits and biases. Both the bubbles and the prototype estimation task contributed unique understanding into the emotion recognition abilities of individuals beyond categorization accuracy metrics. The distribution of my methods and analysis pipelines would allow for an efficient continuation of the same paradigms in future research.

5. REFERENCES

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and crossvalidation of the factor structure. Journal of Psychosomatic Research, 38(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1.

Bardone-Cone, A. M., Harney, M. B., Maldonado, C. R., Lawson, M. A., Robinson, D. P., Smith, R., & Tosh, A. (2010). Defining recovery from an eating disorder: Conceptualization, validation, and examination of psychosocial functioning and psychiatric comorbidity. *Behaviour research and therapy*, *48*(3), 194-202.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of personality assessment*, *67*(3), 588-597.

Bird, G., Press, C., & Richardson, D. C. (2011). The role of alexithymia in reduced eye-fixation in Autism Spectrum Conditions. Journal of Autism and Developmental Disorders, 41(11), 1556–1564. https://doi.org/10.1007/s10803-011-1183-3.

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*, *17*(7), 1107.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830-2838.

Brewer, R., Cook, R., Cardi, V., Treasure, J., & Bird, G. (2015). Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *Royal Society open science*, *2*(1), 140382.

Castro, L., Davies, H., Hale, L., Surguladze, S., & Tchanturia, K. (2010). Facial affect recognition in anorexia nervosa: is obsessionality a missing piece of the puzzle?. *Australian & New Zealand Journal of Psychiatry*, *44*(12), 1118-1125.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. 20th–.

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience & Biobehavioral Reviews*, *84*, 92-99.

Dapelo, M. M., Surguladze, S., Morris, R., & Tchanturia, K. (2016). Emotion recognition in blended facial expressions in women with anorexia nervosa. *European Eating Disorders Review*, *24*(1), 34-42.

de Oliveira, C., Colton, P., Cheng, J., Olmsted, M., & Kurdyak, P. (2017). The direct health care costs of eating disorders among hospitalized patients: A population-based study. *International Journal of Eating Disorders*, *50*(12), 1385-1393.

de Vos, J. A., LaMarre, A., Radstaak, M., Bijkerk, C. A., Bohlmeijer, E. T., & Westerhof, G. J. (2017). Identifying fundamental criteria for eating disorder recovery: a systematic review and qualitative meta-analysis. *Journal of eating disorders*, *5*(1), 1-14.

DeBruine, L. M. (2018). debruine/webmorph: Beta release 2 (Version v0. 0.0. 9001). Zenodo.

Demenescu, L. R., Kortekaas, R., den Boer, J. A., & Aleman, A. (2010). Impaired attribution of emotion to facial expressions in anxiety and major depression. *PloS one*, *5*(12), e15058.

DeTore, N. R., Mueser, K. T., & McGurk, S. R. (2018). What does the Managing Emotions branch of the MSCEIT add to the MATRICS consensus cognitive battery?. *Schizophrenia Research*, *197*, 414-420.

Ekman, P., & Friesen, W. V. (1976). Pictures of facial affect [Slides]. *San Francisco: Department of Psychology, San Francisco State University*.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*, *44*(5), 458-468.

Fujiwara, E., Kube, V. L., Rochman, D., Macrae-Korobkov, A. K., Peynenburg, V., & University of Alberta Hospital Eating Disorder Program. (2017). Visual attention to ambiguous emotional faces in eating disorders: role of alexithymia. *European Eating Disorders Review*, *25*(6), 451-460

Geisler, W. S. (2003). Ideal observer analysis. *The visual neurosciences*, *10*(7), 12-12.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research*, *41*(17), 2261-2271.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of psychopathology and behavioral assessment*, *26*(1), 41-54.

Hasin, D. S., Sarvet, A. L., Meyers, J. L., Saha, T. D., Ruan, W. J., Stohl, M., & Grant, B. F. (2018). Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA psychiatry*, *75*(4), 336-346.

Hudson, J. I., Hiripi, E., Pope Jr, H. G., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological psychiatry*, *61*(3), 348-358.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*(1), 19.

Jänsch, C., Harmer, C., & Cooper, M. J. (2009). Emotional processing in women with anorexia nervosa and in healthy volunteers. *Eating behaviors*, *10*(3), 184-191.

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of abnormal psychology*, *115*(4), 705.

Jordan, J., Joyce, P. R., Carter, F. A., Horn, J., McIntosh, V. V., Luty, S. E., ... & Bulik, C. M. (2008). Specific and nonspecific comorbidity in anorexia nervosa. *International Journal of Eating Disorders*, *41*(1), 47-56.

Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and replication in adolescent and adult samples. *Journal of Psychopathology and Behavioral Assessment*, *38*(3), 443-455.

Keating, L., Mills, J. S., & Rawana, J. S. (2019). Momentary predictors of binge eating: an attachment perspective. *Eating behaviors*, *32*, 44-52.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, *62*(6), 593-602.

Kim, Y. R., Eom, J. S., Leppanen, J., Leslie, M., & Treasure, J. (2018). Effects of intranasal oxytocin on the attentional bias to emotional stimuli in patients with bulimia nervosa. *Psychoneuroendocrinology*, *91*, 75-78.

Krause, F. C., Linardatos, E., Fresco, D. M., & Moore, M. T. (2021). Facial emotion recognition in major depressive disorder: A meta-analytic review. *Journal of Affective Disorders*, *293*, 320-328.

Kucharska-Pietura, K., Nikolaou, V., Masiak, M., & Treasure, J. (2004). The recognition of emotion in the faces and voice of anorexia nervosa. *International Journal of Eating Disorders*, *35*(1), 42-47.

Lavender, J. M., Wonderlich, S. A., Engel, S. G., Gordon, K. H., Kaye, W. H., & Mitchell, J. E. (2015). Dimensions of emotion dysregulation in anorexia nervosa and bulimia nervosa: A conceptual review of the empirical literature. *Clinical psychology review*, *40*, 111-122.

Lee, J., Gosselin, F., Wynn, J. K., & Green, M. F. (2011). How do schizophrenia patients use visual information to decode facial emotion?. *Schizophrenia bulletin*, *37*(5), 1001-1008.

Leehr, E. J., Krohmer, K., Schag, K., Dresler, T., Zipfel, S., & Giel, K. E. (2015). Emotion regulation model in binge eating disorder and obesity-a systematic review. *Neuroscience & Biobehavioral Reviews*, *49*, 125-134.

Li, S., & Deng, W. (2019). Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. *International Journal of Computer Vision*, *127*(6), 884-906.

Lulé, D., Schulze, U. M., Bauer, K., Schöll, F., Müller, S., Fladung, A. K., & Uttner, I. (2014). Anorexia nervosa and its relation to depression, anxiety, alexithymia and emotional processing deficits. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *19*(2), 209-216.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209-226.

Mason, T. B., Lesser, E. L., Dolgon-Krutolow, A. R., Wonderlich, S. A., & Smith, K. E. (2020). An updated transdiagnostic review of social cognition and eating disorder psychopathology. *Journal of Psychiatric Research*.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2. 0. *Emotion*, *3*(1), 97.

Mehu, M., & Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: An integrated approach. *Emotion*, *15*(6), 798.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., ... & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement

(NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980-989.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., ... & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, *56*(2), 128.

Montebarocci, O., Codispoti, M., Surcinelli, P., Franzoni, E., Baldaro, B., & Rossi, N. (2006). Alexithymia in female patients with eating disorders. *Eating and weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *11*(1), 14-21.

Nalbant, K., Kalaycı, B. M., Akdemir, D., Akgül, S., & Kanbur, N. (2019). Emotion regulation, emotion recognition, and empathy in adolescents with anorexia nervosa. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *24*(5), 825-834.

Palser, E. R., Palmer, C. E., Galvez-Pol, A., Hannah, R., Fotopoulou, A., & Kilner, J. M. (2018). Alexithymia mediates the relationship between interoceptive sensibility and anxiety. *PloS one*, *13*(9), e0203212.

Parling, T., Mortazavi, M., & Ghaderi, A. (2010). Alexithymia and emotional awareness in anorexia nervosa: time for a shift in the measurement of the concept?. *Eating behaviors*, *11*(4), 205-210.

Phillipou, A., Abel, L. A., Castle, D. J., Hughes, M. E., Gurvich, C., Nibbs, R. G., & Rossell, S. L. (2015). Self perception and facial emotion perception of others in anorexia nervosa. *Frontiers in Psychology*, *6*, 1181.

Pollatos, O., Herbert, B. M., Schandry, R., & Gramann, K. (2008). Impaired central processing of emotional faces in anorexia nervosa. *Psychosomatic Medicine*, *70*(6), 701-708.

Poquérusse, J., Pastore, L., Dellantonio, S., & Esposito, G. (2018). Alexithymia and autism spectrum disorder: a complex relationship. *Frontiers in psychology*, *9*, 1196.

Pringle, A., Harmer, C. J., & Cooper, M. J. (2010). Investigating vulnerability to eating disorders: biases in emotional processing. *Psychological Medicine*, *40*(4), 645

Richards, H. J., Benson, V., Donnelly, N., & Hadwin, J. A. (2014). Exploring the function of selective attention and hypervigilance for threat in anxiety. *Clinical psychology review*, *34*(1), 1-13.

Ridout, N., Wallis, D. J., Autwal, Y., & Sellis, J. (2012). The influence of emotional intensity on facial emotion recognition in disordered eating. *Appetite*, *59*(1), 181-186.

Robinson, A. L., Dolhanty, J., & Greenberg, L. (2015). Emotion-focused family therapy for eating disorders in children and adolescents. *Clinical psychology & psychotherapy*, *22*(1), 75-82.

RStudio Team. (2022). *RStudio: Integrated Development Environment for R*. Boston, MA. Retrieved from http://www.rstudio.com/

Seidel, M., King, J. A., Ritschel, F., Boehm, I., Geisler, D., Bernardoni, F., ... & Ehrlich, S. (2018). The real-life costs of emotion regulation in anorexia nervosa: a combined ecological momentary assessment and fMRI study. *Translational psychiatry*, *8*(1), 1-11.

Sfärlea, A., Greimel, E., Platt, B., Bartling, J., Schulte-Körne, G., & Dieler, A. C. (2016). Alterations in neural processing of emotional faces in adolescent anorexia nervosa patients–an event-related potential study. *Biological psychology*, *119*, 141-155.

Song, Y., Kawabe, T., Hakoda, Y., & Du, X. (2012). Do the eyes have it? Extraction of identity and positive expression from another's eyes in autism, probed using "Bubbles". *Brain and Development*, *34*(7), 584-590.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the state-trait anxiety scale. *Consulting Psychologists*.

Stice, E., Fisher, M., & Martinez, E. (2004). Eating disorder diagnostic scale: additional evidence of reliability and validity. *Psychological assessment*, *16*(1), 60.

Stice, E., Telch, C. F., & Rizvi, S. L. (2000). Development and validation of the Eating Disorder Diagnostic Scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychological assessment*, *12*(2), 123.

Streatfeild, J., Hickson, J., Austin, S. B., Hutcheson, R., Kandel, J. S., Lampert, J. G., ... & Pezzullo, L. (2021). Social and economic cost of eating disorders in the United States: Evidence to inform policy action. *International Journal of Eating Disorders*, *54*(5), 851-868.

Stroud, J. B., Freeman, T. P., Leech, R., Hindocha, C., Lawn, W., Nutt, D. J., ... & Carhart-Harris, R. L. (2018). Psilocybin with psychological support improves emotional face recognition in treatment-resistant depression. *Psychopharmacology*, *235*(2), 459-466.

Surcinelli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of anxiety disorders*, *20*(1), 110-117.

Surguladze, S. A., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology*, *18*(2), 212.

Tamiya, H., Ouchi, A., Chen, R., Miyazawa, S., Akimoto, Y., Kaneda, Y., & Sora, I. (2018). Neurocognitive impairments are more severe in the binge-eating/purging anorexia nervosa subtype than in the restricting subtype. *Frontiers in psychiatry*, *9*, 138.

Todorov, A., Dotsch, R., Wigboldus, D. H., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*, *5*(10), 775-791.

Torres, S., Guerra, M. P., Lencastre, L., Miller, K., Vieira, F. M., Roma-Torres, A., ... & Costa, P. (2015). Alexithymia in anorexia nervosa: the mediating role of depression. *Psychiatry research*, *225*(1-2), 99-107.

Tseng, A., Bansal, R., Liu, J., Gerber, A. J., Goh, S., Posner, J., ... & Peterson, B. S. (2014). Using the circumplex model of affect to study valence and arousal ratings of emotional faces by children and adults with autism spectrum disorders. *Journal of autism and developmental disorders*, *44*(6), 1332-1346.

Vander Wal, J. S., Kauffman, A. A., & Soulliard, Z. A. (2020). Differences in alexithymia, emotional awareness, and facial emotion recognition under conditions of self-focused attention among women with high and low eating disorder symptoms: a 2 x 2 experimental study. *Journal of eating disorders*, *8*(1), 1-9.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological bulletin*, *122*(3), 231.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*, *33*(2), 113-120.

White, S. W., Mazefsky, C. A., Dichter, G. S., Chiu, P. H., Richey, J. A., & Ollendick, T. H. (2014). Social-cognitive, physiological, and neural mechanisms underlying emotion regulation impairments: Understanding anxiety in autism spectrum disorder. *International Journal of Developmental Neuroscience*, *39*, 22-36.

Wildes, J. E., Ringham, R. M., & Marcus, M. D. (2010). Emotion avoidance in patients with anorexia nervosa: Initial test of a functional model. *International Journal of Eating Disorders*, *43*(5), 398-404.

Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in psychology*, *5*, 404.

Wyssen, A., Lao, J., Rodger, H., Humbel, N., Lennertz, J., Schuck, K., ... & Munsch, S. (2019). Facial emotion recognition abilities in women experiencing eating disorders. *Psychosomatic Medicine*, *81*(2), 155-164.

Xu, J., Harper, J. A., Van Enkevort, E. A., Latimer, K., Kelley, U., & McAdams, C. J. (2017). Neural activations are related to body-shape, anxiety, and outcomes in adolescent anorexia nervosa. *Journal of psychiatric research*, *87*, 1-7.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, *63*(3), 271-313.

Zonnevijlle-Bendek, M. J. S., Van Goozen, S. H. M., Cohen-Kettenis, P. T., Van Elburg, A., & Van Engeland, H. (2002). Do adolescent anorexia nervosa patients have deficits in emotional functioning?. *European child & adolescent psychiatry*, *11*(1), 38-42.

Zwick, J. C., & Wolkenstein, L. (2017). Facial emotion recognition, theory of mind and the role of facial mimicry in depression. *Journal of affective disorders*, *210*, 90-99.