Gender Differences in Recognizing Depression and Seeking Help*

Catalina Franco† Akshay Moorthy‡ Sara Abrahamsson§

June 30, 2025

Abstract

Substantial gender disparities in depression rates are observed worldwide, with women more frequently diagnosed than men. This study examines two potential behavioral drivers of these gaps: recognition of symptoms and willingness to seek help. To circumvent challenges that may lead to underrecognition and underreporting in real-life cases, we use hypothetically generated depression scenarios in a pre-registered experiment with a representative U.S. sample. Our main finding is that men face a "double whammy" in mental health: they are less likely than women to recognize depression, especially at milder severity levels, and less likely to seek help when evaluating scenarios about themselves, despite believing others should seek help in similar situations. Among our three pre-specified mechanisms—perceptions, psychic costs, and social norms—psychic costs play a key role in men's help-seeking behavior.

1 Introduction

Mental health disorders are a leading cause of disability, affecting over 1 billion people worldwide (Rehm and Shield, 2019). In the U.S. alone, these conditions cost approximately US$201 billion annually (Bütikofer et al., 2024). Research has shown that mental health issues negatively impact labor market outcomes, including lost work days (Ridley et al., 2020; Currie, 2024) and income (Smith and Smith, 2010; Goodman et al., 2011; Biasi et al., 2021).

Worldwide, there are substantial gender disparities in depression rates (Van de Velde et al., 2010). As early as the 1970s, research reported that women were twice as likely as men to be diagnosed with depression and exhibited twice as many depressive symptoms

*We acknowledge financial support from the Norwegian School of Economics (NHH). We received invaluable feedback from participants at the FAIR breakfast meeting, the Lausanne XX, and individually from Aline Bütikofer, Amanda Kowalski, Petra Persson, etc.

†Center for Applied Research (SNF) and FAIR at NHH – Norwegian School of Economics.

‡Faculty of Business and Economics, University of Lausanne.

§Division of Health Services at the Norwegian Institute of Public Health in Oslo, Norway.

(Weissman and Klerman, 1977). For women, higher rates of depression diagnosis can affect labor force participation and career progression, directly contributing to gender gaps in labor market outcomes. For men, unreported and untreated mental health issues may undermine productivity and well-being at work. Despite the magnitude and importance of gender gaps in mental health, the underlying causes of these disparities remain unclear.

This study examines whether behavioral barriers may contribute to gender gaps in depression. While previous work in economics has looked at institutional factors such as differential screening of girls vs. boys (Corredor-Waldron and Currie, 2024),[1] less is known about how individuals perceive their own symptoms before seeking medical attention. Once in contact with a doctor or health provider, screening can take place, but the decision to discuss symptoms—and even to seek help in the first place—is personal. Understanding these motivations is challenging, as they occur before individuals enter the healthcare system, are typically not captured in health data registers and are affected by potential stigma (Roth et al., 2024a). In addition, the decision to seek screening or help is shaped by external constraints, such as health insurance coverage, which is not randomly assigned. Moreover, since screening tools are not administered uniformly across the population, self-selection into screening introduces endogeneity concerns.[2]

To address these challenges, we conduct a pre-registered vignette experiment using a representative U.S. sample from the online platform Prolific. Leveraging experimentally generated depression scenarios, we avoid reliance on self-reported diagnoses and directly examine recognition and willingness to seek help. In practice, participants review hypothetical scenarios in two treatment blocks: *Self* (where they imagine experiencing the symptoms themselves) and *Other* (where the subject is a hypothetical male or female individual). Each scenario is defined by a randomly selected PHQ-9 score and a corresponding set of symptoms and frequencies randomly drawn to match that score. After reviewing each scenario, participants assess whether the symptoms indicate depression (*recognition*), its perceived

[1]The topic has nevertheless received much more attention in fields like psychiatry, where research has focused on documenting the evolution of the gaps (e.g., Parker and Brotchie, 2010) and the underlying environmental factors and stressors for why differences emerge (e.g., Girgus and Yang, 2015).

[2]For instance, if women are more likely to visit doctors, encounter these tools, or be screened more frequently, observed gender differences in depression diagnoses would partly reflect differences in access.

severity, the likelihood of *seeking help*, and their beliefs about mental health. This approach also enables us to examine underlying factors—such as perceptions, psychic costs, and social norms—that may drive gender differences in depression recognition and help-seeking behavior.

Three main findings emerge. First, we document that women have lower thresholds than men for recognizing depression and tend to overestimate its severity. In all our hypothetical scenarios, which depict varying degrees of depression, over 85% of women identify the symptoms as depression. However, men are significantly less likely to do so, with an overall recognition gap of 5.3 percentage points (pp). This gap is particularly pronounced for milder scenarios of depression, where the gender difference in recognition is 14.5 pp or 17%. In addition, women are more likely than men to overestimate the severity of depression. Few people (5% of women and 12.9% of men) have correct perceptions of the severity at mild levels. Their accuracy increases as the symptoms reflect more severe cases, but does not reach 100% even in severe depression cases. The gender gap in severity accuracy persists across the full range of the PHQ-9 distribution.

Second, we do not find gender differences in willingness to seek help–in general, most participants report that they would seek help from at least one of six possible sources of help given the symptoms depicted in the scenario. An interesting gender gap emerges, nevertheless, when focusing specifically on help from specialist sources, such as general practitioners and therapists: men are 6.1 pp less likely to report that they would seek help from these sources. This gap widens to 10.6 pp in mild depression scenarios and does not fully close even at moderately severe and severe depression levels.

Third, among the mechanisms we pre-specified, we find that gender differences in psychic costs, i.e., seeking help for a mental health issue may impose emotional or mental burdens, are more likely to drive help-seeking behavior than differences in perceptions or social norms. We leverage the randomization of depression scenarios depicting symptoms for oneself (*Self*) or for another person (*Other*) and find that the insignificant gender gap in seeking help from specialist and non-specialist sources combined masks important differences across conditions. In *Self* scenarios, men are 4.2 pp less likely than women to report

willingness to seek help, whereas in *Other* scenarios, there is no gender gap. In other words, men report lower willingness to seek help when the symptoms concern themselves but believe that others experiencing similar symptoms should seek help.

Our paper contributes to the long-standing psychiatric literature that documents patterns and investigates potential causes of gender differences in depression diagnoses. The review study by Girgus and Yang (2015) discusses deep-rooted gender differences in coping mechanisms (e.g., rumination and need for approval) and in stressors (e.g., sexual abuse and social image concerns). They also discuss "artefactual" reasons for the gender gap such as women being more willing than men to admit (even to themselves) that they feel depressed, to report depressed feelings when asked and to seek help when they feel depressed. However, they conclude that "researchers who have considered these possibilities have uniformly concluded that none of them contribute to the gender difference in depression." Our paper departs from this view by showing that, conditional on a given level of depression, men face the double-whammy of lower recognition and lower willingness to seek help. In doing so, we contribute to the understanding of gender gaps in mental health by highlighting the role of behavioral factors, an explanation that has largely been dismissed in previous work. Importantly, our findings point to new and actionable public health policy levers.

To the best of our knowledge, no prior work in economics has directly examined the behavioral drivers of *gender gaps* in mental health. Most of the focus in economics has been on evaluating the effectiveness of psychotherapy (e.g., Baranov et al., 2020; Ridley et al., 2020; Angelucci and Bennett, 2024). However, the field is increasingly well positioned for research looking at behavioral drivers of mental health issues. Recent studies explore how misinformation about mental health conditions (Acampora et al., 2022), misperceptions about treatment effectiveness (Roth et al., 2024b), and the prevalence of stigma (Roth et al., 2024a) affect decisions to seek treatment. Our paper complements this work by directly identifying two behavioral barriers that affect treatment take-up and by showing that they are more prevalent among men. Our evidence suggests that gender gaps in how individuals perceive and respond to mental health symptoms have important implications for the design of both psychotherapy interventions and light-touch approaches targeting mental health

literacy (Acampora et al., 2022) or misperceptions (Roth et al., 2024b,a). In particular, our findings highlight the value of tailoring intervention content by gender to address distinct behavioral barriers.

2 Design

We implement a within-subjects survey experiment in which all participants are exposed to two randomized treatment blocks: *Self* and Other. In each block, participants evaluate one or more hypothetical scenarios based on the PHQ-9 instrument (Kroenke et al., 2001), a widely-used diagnostic tool for depression. Each scenario presents a set of symptoms and their reported frequency over the past two weeks. The symptoms are constructed to match a randomly selected PHQ-9 score between 4 (minimal depression) and 21 (severe depression), using a randomization procedure that first draws a target score and then selects symptom items and severities consistent with that score (See Appendix Section C for the complete symptom list and severity classification and Appendix Section F for the randomization procedure). Appendix Figure A.1 shows the distribution of the severity of scenarios seen by participants.

Treatment blocks. The two treatment blocks, *Self* and *Other*, differ only in the subject of the hypothetical scenario. In *Self*, the subject of the hypothetical scenario is the participant. In *Other*, the subject of the scenario is a hypothetical individual (who is either Male or Female). Participants evaluate two scenarios in each treatment block and the order of the blocks is randomized. In the *Other* block, the gender of the (hypothetical) individual is randomized between-subjects. The symptoms displayed in the second scenario of the *Other* block are the same for all participants, and is always shown at the end of the survey (i.e. after two *Self* and one *Other* scenario).

Main outcomes. After reviewing the hypothetical scenario, participants are asked to state their views on: (i) whether the symptoms indicate that the subject of the scenario suffers from depression (*recognition*), (ii) *accuracy* in depression severity, (iii) the likelihood of *seek-*

ing help from six different sources. We construct binary outcome variables for recognition, accuracy in severity classification, and willingness to seek help (overall, and from specialists and non-specialists).

2.1 Study procedures and sample

The experiment was conducted with a sample of 401 U.S. participants recruited through the online survey platform Prolific. Participants evaluated four different scenarios for a total sample size of 1,604 evaluations. The sample is representative of the U.S. population along gender, age, and ethnicity. The experiments were built using OTree (Chen et al., 2016). The survey was conducted in October 2024, and we pre-specified the design and hypotheses (AEARCTR-0014621) before data collection. The study was reviewed and approved by the IRB at the Norwegian School of Economics. The analysis follows the pre-analysis plan for the most part, and any deviations or exploratory analyses are clearly identified in the text.

We collected participants' background characteristics and their prior experience with mental health treatment. Of the 1,604 scenario evaluations we analyze,[3] 820 (51.1 percent) were completed by women. The mean age for women (men) in the sample is 46 (45) years old, 43% of women (34.7% of men) have less than a bachelor's degree, 59.5% of women (75% of men) are employed, 53.2% of women (62.8% of men) have an annual household income at or above US $50,000, and 21% of women (40% of men) were not familiar with depression screening tools at the moment they answered the survey.

Appendix Figure A.1 shows a histogram of the PHQ-9 scores of the 1604 scenarios displayed to participants. A quarter of the scenarios have a score of 4, which corresponds to the common scenario (shown fourth in the sequence of scenarios). For scores between 5 and 21, the distribution of scenarios is relatively uniform, as expected given the randomization of PHQ-9 scores.

[3]We exclude 44 evaluations from 11 individuals who reported non-binary gender.

2.2 Empirical strategy

We first analyze the raw data at the scenario level by plotting the three main outcome variables (depression recognition, accurate severity and help seeking) against the PHQ-9 score, and separately by the gender of the respondent. We use binned scatterplots, controlling for the treatment assignment (*Self* vs. Other), whether the subject in the *Other* scenario is a male or a female and the order in which the scenarios were presented.[4]

Second, we conduct a regression analysis of the three main outcomes on a gender indicator (*Male* = 1 if the participant is male) (see Equation 1), and *Male* along with indicators for severity levels of the scenario (Mild, Moderate and Moderately severe, with excluded indicator Minimal) and the interactions between Male and the severity levels (see Equation 2).[5] In the seeking help outcome, we do not have any scenario with Mild severity so the regressions exclude the Mild indicator instead. All regressions include demographic controls (participant age, education, employment, income level and familiarity with depression screening tools), treatment and order controls. Standard errors are heteroscedasticity robust and clustered at the participant level.

$$y_i = \alpha_0 + \alpha_1 \text{Male}_i + \gamma_i + \varepsilon_i \tag{1}$$

$$\begin{aligned} y_i = &\beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{Mild}_i + \beta_3 \text{Male}_i \times \text{Mild}_i + \beta_4 \text{Moderate}_i + \\ &\beta_5 \text{Male}_i \times \text{Moderate}_i + \beta_6 \text{Modsevere}_i + \beta_7 \text{Male}_i \times \text{Modsevere}_i + \gamma_i + \varepsilon_i \end{aligned} \tag{2}$$

2.3 Potential factors underlying gender differences in depression recognition and seeking help

We complement the main analysis with a pre-registered investigation of three plausible mechanisms related to the recognition and response to mental health symptoms: perceptions, beliefs, and psychic costs.

[4]The fourth scenario was fixed so instead of controlling for the numerical order of the scenario, we control for whether *Self* or *Other* was presented first.

[5]In the PAP, we intended to use indicators for Moderate, Moderately severe and Severe, with excluded indicator Mild. However, since we also have a scenario depicting minimal depression (PHQ-9 score equal to 4) that every participant evaluated, we decided to include it in the analysis and changed the omitted category to Mild. We also combined the categories Moderately severe and Severe because there did not seem to be any difference between these two by looking at the scatterplots.

Perceptions. Men and women may differ in how they perceive depression symptoms, what others would think of them, how effective treatments options are, or whether they have side effects. We evaluate whether there are important gender differences in these perceptions using survey question 4 (see Appendix E) which was presented to participants in each of the first three scenarios.

Psychic costs. Recognizing that one may be experiencing issues can be difficult for some people because it imposes a emotional and mental burden that can include feelings of shame, guilt, fear, or denial. These psychic costs would be lower when evaluation scenarios about someone else instead of oneself. We leverage the within-subject treatment assignment to *Self* or *Other* to examine whether there are gender differences in participants' ability to recognize depression. For example, if men are more likely to recognize depression or express willingness to seek help in *Other* scenarios than in *Self* scenarios, this would suggest that acknowledging mental health problems may be more psychologically costly when the symptoms are self-attributed.

Norms. Gender or masculinity norms may affect how men are able to recognize or seek help when they are experiencing mental health issues. Traditional norms often emphasize traits like toughness, self-reliance, and emotional restraint in men, which can create barriers to acknowledging mental health struggles. To test whether social norms contribute to gender differences in recognition, we include an incentivized second-order belief question (Q5 in Appendix E). This question was asked in the fourth scenario where all participants see the same (fixed) list of symptoms, and elicits participants' beliefs about the descriptive norm – the fraction of US participants in a previous study who would say that the scenario was a case of depression. If men, more than women, believe that others are unlikely to recognize depression or seek help, this would suggest that perceived social norms may help explain gender gaps in mental health recognition and help-seeking behavior.

We assess these patterns using both graphical evidence and regression analysis, modifying Equation 1 to include an indicator for the *Other* treatment and its interaction with the Male indicator when analyzing perceptions. To examine whether the gender of the scenario

subject influences responses, we restrict the sample to *Other* scenarios and estimate a similar specification using an indicator for whether the subject in the scenario is male (*Other* is male). This allows us to test whether participants respond differently based on the gender of the hypothetical individual experiencing symptoms.

3 Analysis Of Gender Differences

3.1 Depression recognition

Our first pre-specified outcome is whether individuals recognize that the symptoms presented in a hypothetical scenario correspond to depression. We hypothesize that men and women may differ in their propensity to identify depression, which could partly explain observed gender differences in depression prevalence. Our design allows us to isolate recognition differences conditional on a given PHQ-9 score by randomly assigning respondents a scenario constructed from a randomly drawn score (ranging from 4 to 21) and a corresponding set of symptoms and frequencies consistent with that score. This approach overcomes a key limitation of observational data, where men and women may report different symptoms or the distribution of experienced PHQ-9 scores may differ by gender, preventing recognition comparisons.

Our first main finding is that women are more likely than men to recognize depression. Figure 1 shows binned scatterplots of depression recognition by PHQ-9 score, separately by gender. Recognition is generally high, as expected given that all scenarios depict depression: over 90% of women and 70% of men correctly identify it. For both genders, recognition increases with depression severity, but the gradient is steeper for men. At moderate to severe levels (PHQ-9 scores of 15–21), nearly all respondents, regardless of gender, recognize depression. At minimal to mild levels (PHQ-9 scores of 4–10), however, a pronounced gender gap emerges, with women significantly more likely to recognize depression than men.

We quantify gender differences in depression recognition in Table 1. Column (1) shows an overall gender gap of 5.3 pp, from a base of 95% of women recognizing depression. The gap is substantially larger (14.5 pp, or 17%) at minimal depression levels, where 87% of

women identify the scenario as depression (Col. (2)). Recognition rates increase sharply with severity for both genders, reaching near-universal recognition at moderate, and moderately severe levels. Men show a larger increase in recognition across these categories, as reflected in the positive interaction terms between male and depression severity. Exploratory analyses show that the gender gap in recognition exhibits a pronounced age gradient, with men below 30 years old having similar recognition rates as women, but much lower rates at older ages.

In sum, we find a sizable gender gap in depression recognition, with women having a lower threshold to recognize depression symptoms than men. This pattern is driven by men being less likely than women to recognize depression at milder severity levels, but not at the more severe levels.

3.2 Accuracy in recognizing depression severity

Our second pre-registered outcome captures whether individuals accurately classify the severity of the depression indicated by the symptoms in a given scenario. Misclassification adds an additional layer to recognition failures, as systematically underestimating symptom severity may reduce the likelihood of seeking help. Conversely, overestimating severity could lead to overdiagnosis or excessive concern.

Figure 2 presents binned scatterplots of severity classification accuracy by PHQ-9 score. Panel 2a shows that accurate classification is generally low, particularly at lower severity levels: fewer than 20% of participant assessments correctly classify severity for scenarios with PHQ-9 scores up to 10 (minimal or mild depression). Accuracy improves with severity but remains low with only about 50% of assessments being correct even at the highest scores. Panels 2b and 2c decompose these inaccuracies into over- and underestimation. Most errors stem from overestimating severity, especially at the lower end of the PHQ-9 scale, though this pattern persists across the full support of the PHQ-9 score.

A second key pattern is that men are statistically significantly more accurate than women in classifying severity across all levels. Overall, men are 6.8 pp more accurate in their depression severity assessments than women, a 40% difference off a base of 18% of accurate

assessments among women (Table 1, Col. (3)). This gender gap in severity classification persists across the full range of PHQ-9 scores (Table 1, Col. (4)).

3.3 Willingness to seek help

Our third pre-specified outcome captures participants' willingness to seek help after viewing the list of symptoms. This question was asked regardless of whether respondents correctly recognized the symptoms as depression, since individuals may be inclined to seek help even without labeling the condition. Participants considered six possible sources of help, which we classify as either specialist or non-specialist. Help-seeking behavior is central to understanding gender gaps in diagnosed depression rates, as these rates often reflect contact with healthcare providers, whether through routine screenings or active help-seeking. Our measure focuses on the latter.

Figure 3 presents binned scatterplots of willingness to seek help as a function of PHQ-9 scores combining both specialist and non-specialist sources in Panel 3a, and shown separately in Panels 3b and 3c.[6] Overall, participants report being likely or very likely to seek help in over 90% of the scenarios, with little variation across the severity spectrum. Men and women exhibit similar patterns, though men report slightly lower average willingness to seek help. However, confidence intervals for the two groups overlap at several points across the PHQ-9 distribution (Panel 3a).

Two additional patterns emerge in help-seeking behavior. First, as shown in Panels 3b and 3c, reported willingness to seek help differs substantially between specialist and non-specialist sources, particularly for women. Both men and women express high willingness to seek help from non-specialist sources—such as counselors, friends, or AI chatbots—in approximately 80–85% of scenarios. However, willingness to seek help from specialist sources—such as general practitioners or therapists—is higher overall for both genders. Second, striking gender differences appear in willingness to seek help from specialist sources. Women report being willing to seek specialist help in 95% of scenarios, regardless

[6]Each participant evaluates four scenarios in which the subject is either themselves ("Self") or someone else ("Other"). In these plots, we pool across both conditions; the *Self* vs. *Other* treatment assignment is analyzed in Section 4.

of depression severity. In contrast, men's willingness starts at around 80% for mild depression and increases with severity, reaching parity with women only at the most severe levels. Among non-specialist sources, the only source where men would be more eager than women to seek help from is AI-enabled mental health chatbots (see Figure A.3).

Our findings suggest that while overall willingness to seek help is high, important gender differences emerge in the type of help sought. In particular, men are less likely than women to seek help from specialist providers, especially at lower levels of depression severity. This gap may contribute to underdiagnosis among men, as specialist contact is often required for formal diagnosis and treatment.

4 Factors underlying gender differences in depression recognition and seeking help

4.1 Perceptions

Using Q4 in the survey,[7] we generate binary outcomes for each of the answer options, which take the value of 1 when the respondent somewhat or strongly agrees with the statement.

We do not find any consistent patterns where one gender is more likely than the other to agree with the statements using visual evidence from scatterplots in Figure A.4 or estimates from Equation 1 in Figure A.5. However, some interesting descriptive evidence emerges. The overall agreement with the statements regardless of gender or scenario severity is 46.1% for "the issues would go away by themselves in some time," 32.7% for "the issues are not very serious and do not require treatment," 71.1% for "I/[NAME] worry/worries about what others would think of me/her/him if they became aware that [I/she,he] had these issues," 59.4% for " I/[NAME] would rather deal with these issues myself/herself,himself] than rely on help from others," 48.1% for "I/[NAME] do/does not think that the available treatments for these issues are effective," and 79.1% for "I/[NAME] am/is worried about the side effects

[7] Q4 reads: Based on the hypothetical issues and their frequencies experienced by [you/NAME], please indicate the extent to which you agree or disagree with the following statements. *[Strongly disagree, Somewhat disagree, Somewhat agree, Strongly agree]*.

of medication for depression." These overall averages are indicative of the pattern across severity levels, as the scatterplots remain relatively flat across the full range of PHQ-9 scores (see Figure A.4).

The only perceptions where men and women appear to slightly differ are in the beliefs that "the issues would go away on their own over time" and "the issues are not very serious and do not require treatment." For the latter, the confidence bands do not overlap starting at moderate depression levels, suggesting that men are more likely than women to dismiss the symptoms and believe that treatment is unnecessary, particularly as severity increases.

Although gender differences in perceptions are largely absent, the overall levels of agreement with the statements reveal striking patterns. In over 70% of scenario evaluations, participants report concerns about what others would think and about the side effects of depression medication. We interpret these findings in light of recent work in economics on perceptions of stigma and depression. Roth et al. (2024a) document widespread misperceptions about stigma: individuals believe that 38% of Americans hold stigmatizing beliefs, while the actual rate is only 16%. The high share of respondents expressing concern about others' opinions may reflect anticipated stigma, but could also reflect general discomfort with disclosure or fears of burdening others. Concerns about medication side effects build on evidence of misperceptions about the effectiveness of online therapy (Roth et al., 2024b), suggesting that reluctance to seek treatment may be even greater when pharmacological options are involved.

4.2 Psychic costs

To study potential psychic costs, we exploit the random assignment to *Self* and *Other* scenarios and define an indicator equal to 1 when the scenario corresponds to the *Other* condition. This allows us to examine whether the observed gender differences in depression recognition are driven by men being more likely to recognize symptoms as depression in others than in themselves. Such a pattern would suggest that psychological costs may affect self-attribution of depressive symptoms.

Table 2 presents estimates from regressions of our three main outcomes on indicators

for Male and Other, as well as their interaction, reported in Columns (1), (3), and (5). Consistent with our main results, men are on average 4 pp less likely than women to recognize depression (Column (1)). The interaction between Male and Other is small and statistically insignificant, suggesting that men are less likely than women to recognize depression regardless of whether the symptoms are about themselves or about others. A similar pattern is observed in Column (3), where men are more accurate in recognizing the severity of depression than women regardless of whether the scenarios is about the Self or Other. The evidence on recognition suggests that the gender gap is driven not by psychological costs of self-attribution.

In the case of help-seeking, the main results did not reveal a large gender difference. However, a clear pattern emerges when separating help-seeking responses in the *Self* and *Other* scenarios. When evaluating *Self* scenarios, men are 4 pp less likely than women to report that they would seek help. In contrast, when evaluating *Other* scenarios, the interaction coefficient fully offsets this gap, indicating no gender difference in the belief that someone else should seek help. Moreover, the difference between *Self* and *Other* is entirely driven by willingness to seek help from specialist sources (see Table B.2). The gender gap in this domain is 8 pp (Column 3) and remains of similar magnitude and statistically significant even after controlling for whether the participant recognized the scenario as depression (Column 4). In other words, reluctance to seek help for oneself appears to be distinct from the ability to recognize depression among men.

Our key takeaway from the *Self* vs. *Other* analysis is that men may be facing a "double whammy" when it concerns mental health: not only are they less likely than women to recognize depression in themselves and in others, but are also less likely to seek help when the symptoms concern themselves.

4.3 Norms

We conduct two tests to assess whether norms may help explain the observed gender differences in depression recognition. The underlying idea is that gender or masculinity norms could discourage men from recognizing depression or seeking help as normative expecta-

tions would shape individual responses. We examine the role of norms in two ways: (1) by comparing how participants evaluate *Other* scenarios depending on whether the subject is male or female. (2) by analyzing participants' second-order beliefs about whether others in the study would recognize a given scenario as depression.

In Table 2, we report regression results on our three main outcomes using the *Other* scenarios only in Columns (2), (4), and (6). The main regressors are indicators for Male and Other is male, as well as the interaction between the two. We have already shown that men are less likely than women to recognize depression, regardless of whether the symptoms are attributed to themselves or to others. If norms drive men's lower recognition, we should see especially low recognition when the subject in Other is also male, suggesting that normative expectations about how men should feel or behave influence how male respondents interpret symptoms in others of the same gender.

Column (2) of Table 2 shows that in *Other* scenarios where the subject is a woman, the gender gap in recognition is 11.7 pp. However, the interaction coefficient is 11.3 pp, implying that when the subject is a man, the gender gap effectively disappears. In sum, men are not less likely to recognize depression in others when the symptoms are attributed to another man.

The second test uses an incentivized second-order belief question, which asks participants to guess what fraction of Americans (specifically, participants in a previous study we conducted) believed that the symptoms described in a given scenario corresponded to depression.[8] The scenario was held constant across all participants; only the gender of the subject in the vignette was randomized. If norms are an important driver of men's lower recognition of depression, we would expect male participants to estimate a lower fraction of Americans recognizing the symptoms as depression compared to female participants.

Figure A.2 presents the empirical cumulative distribution functions (CDFs) of male and female participants' guesses about the share of Americans who recognized the scenario as depression. The two CDFs nearly fully overlap, indicating no meaningful difference in beliefs

[8]The scenario corresponds to a PHQ-9 score of 4 (mild depression) and includes questions 1 and 2. Details on the second-order belief question and incentivization are provided in Appendix E. We deviated from our pre-registration where we stated that this scenario would have a PHQ-9 score of 11, based on pilots which showed larger differences at milder severities.

across gender. This suggests that differences in perceived social norms are unlikely to explain the gender gap in depression recognition.

To summarize, both pieces of evidence suggest that gender or masculinity norms are unlikely to be the primary mechanism driving the gender gap in depression recognition. At the same time, the evidence suggests that men's recognition of depression in others varies by the gender of the subject.

5 Conclusion

Gender differences in depression diagnoses are well documented, but little is known about whether behavioral mechanisms may contribute to these gaps. Behavioral barriers may be more amenable to change than deeper-rooted causes of gender differences in depression, such as differential exposure to stressors or variations in coping mechanisms. As such, they represent a promising target for public health interventions.

In this paper, we showed that men face a "double whammy" in mental health. Using hypothetically generated depression scenarios to circumvent the very issue we aim to study—underrecognition and underreporting of mental health symptoms among men—we first document that men are less likely than women to recognize depression, especially when symptoms are not severe. Second, while we did not find substantial gender differences in overall willingness to seek help (though men were less likely to report that they would seek help from specialists), a key difference emerges when participants evaluate symptoms about themselves: men are less likely than women to seek help when the scenario concerns their own symptoms, yet believe that others should seek help in equivalent situations. These findings are important because early recognition is critical for timely intervention, potentially preventing symptoms from escalating into more severe forms of depression. Moreover, they offer guidance on the types of content that could be prioritized in public health policies, such as messaging focused on reducing self-stigma and promoting early engagement with mental health services.

References

Acampora, Michelle, Francesco Capozza, and Vahid Moghani, "Mental Health Literacy, Beliefs and Demand for Mental Health Support among University Students," Technical Report, Tinbergen Institute Discussion Paper 2022.

Angelucci, Manuela and Daniel Bennett, "The economic impact of depression treatment in india: Evidence from community-based provision of pharmacotherapy," *American economic review*, 2024, *114* (1), 169–198.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko, "Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial," *American economic review*, 2020, *110* (3), 824–859.

Biasi, Barbara, Michael S Dahl, and Petra Moser, "Career effects of mental health," Technical Report, National Bureau of Economic Research 2021.

Bütikofer, Aline, Rita Ginja, Krzysztof Karbownik, and Fanny Landaud, "(Breaking) intergenerational transmission of mental health," *Journal of Human Resources*, 2024, *59* (S), S108–S151.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Corredor-Waldron, Adriana and Janet Currie, "To what extent are trends in teen mental health driven by changes in reporting?: The example of suicide-related hospital visits," *Journal of Human Resources*, 2024, *59* (S), S14–S40.

Currie, Janet, "The Economics of Child Mental Health: Introducing the Causes and Consequences of Child Mental Health Special Issue," *Journal of Human Resources*, 2024, *59* (S), S1–S13.

de Velde, Sarah Van, Piet Bracke, and Katia Levecque, "Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression," *Social science & medicine*, 2010, *71* (2), 305–313.

Girgus, Joan S and Kaite Yang, "Gender and depression," *Current Opinion in Psychology*, 2015, *4*, 53–60.

Goodman, Alissa, Robert Joyce, and James P Smith, "The long shadow cast by childhood physical and mental problems on adult life," *Proceedings of the National Academy of Sciences*, 2011, *108* (15), 6032–6037.

Kroenke, Kurt, Robert L Spitzer, and Janet BW Williams, "The PHQ-9: validity of a brief depression severity measure," *Journal of general internal medicine*, 2001, *16* (9), 606–613.

Parker, Gordon and Heather Brotchie, "Gender differences in depression," *International review of psychiatry*, 2010, *22* (5), 429–436.

Rehm, Jürgen and Kevin D Shield, "Global burden of disease and the impact of mental and addictive disorders," *Current psychiatry reports*, 2019, *21*, 1–7.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel, "Poverty, depression, and anxiety: Causal evidence and mechanisms," *Science*, 2020, *370* (6522), eaay0214.

Roth, Christopher, Peter Schwardmann, and Egon Tripodi, "Depression stigma," 2024.

_ , _ , and _ , "Misperceived effectiveness and the demand for psychotherapy," *Journal of Public Economics*, 2024, *240*, 105254.

Smith, James Patrick and Gillian C Smith, "Long-term economic costs of psychological problems during childhood," *Social science & medicine*, 2010, *71* (1), 110–115.

Weissman, Myrna M and Gerald L Klerman, "Sex differences and the epidemiology of depression," *Archives of general psychiatry*, 1977, *34* (1), 98–111.

6 Figures

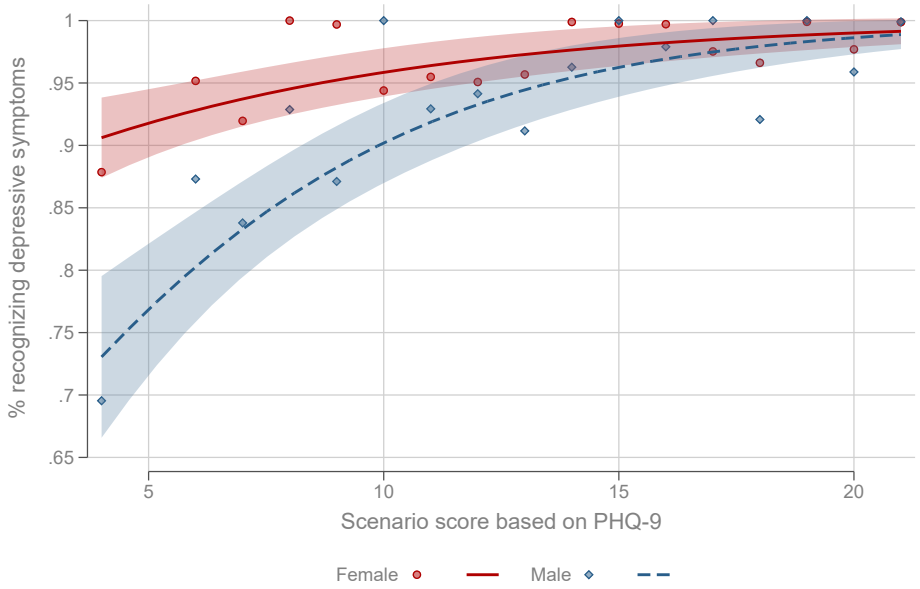
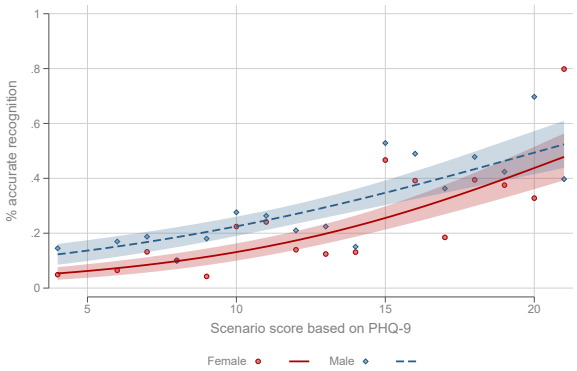


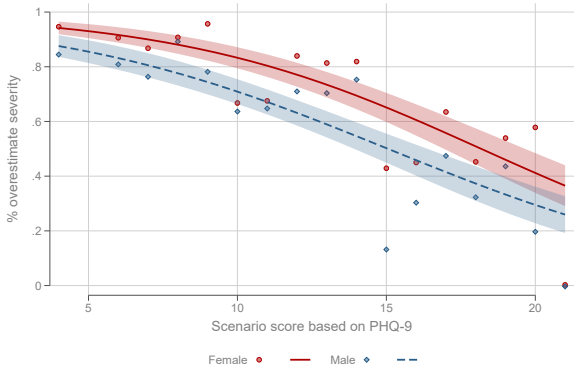
Figure 1: Gender differences in depression recognition

Notes. Binned scatter plot of depression recognition and the severity (PHQ-9) score of the hypothetical scenario, shown separately for male and female participants. The recognition of depressive symptoms binary variable is based on the question: Suppose that [you/NAME] were/was experiencing the hypothetical issues at the frequencies listed above, do you think [you/NAME] would have depression? *[Definitely yes, probably yes, probably no, definitely no]*. Each point in the plot represents the regression coefficient from a regression of recognition on the scenario severity, controlling for treatment assignment (self vs. other), whether the subject in the "other" scenario is a male or a female and the order in which the scenarios were presented. The bands show 95% confidence intervals, with standard errors clustered by participant.

(a) Accurate severity



(b) Overestimate severity



(c) Underestimate severity

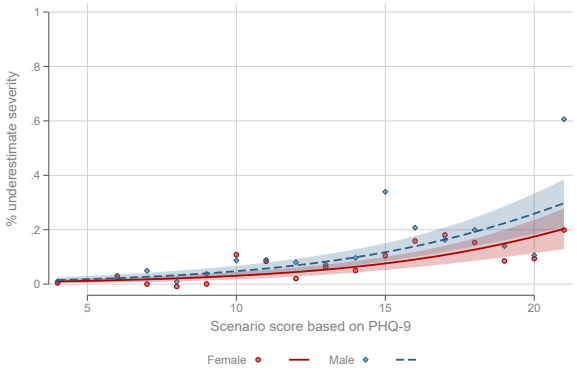
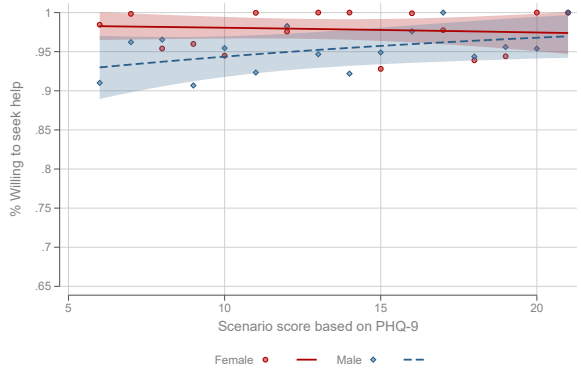


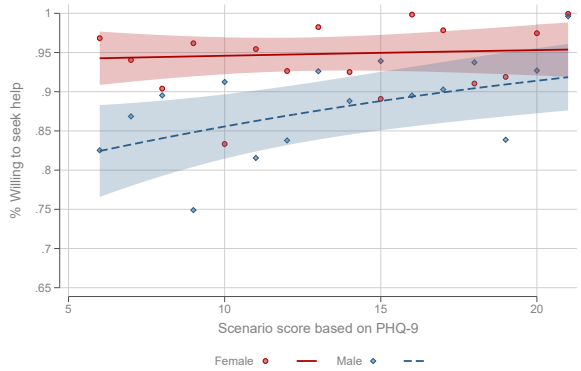
Figure 2: Gender differences in identifying the severity of depressive symptoms

Notes. Binned scatter plot of accuracy in recognizing the severity of depression and the severity (PHQ-9) score of the hypothetical scenario, shown separately for male and female participants. The binary variables are created based on the question: How severe do you think the depression would be if [you/NAME] were experiencing these issues in real life? *[None or minimal, Mild, Moderate, Moderately Severe, Severe]*. Over-(Under-)estimate is defined as classifying the symptoms in a more (less) severe category than they actually belong to. Each point in the plot represents the regression coefficient from a regression of recognition on the scenario severity, controlling for treatment assignment (self vs. other), whether the subject in the "other" scenario is a male or a female and the order in which the scenarios were presented. The bands show 95% confidence intervals, with standard errors clustered by participant.

(a) Would seek help



(b) Would seek help from a specialist



(c) Would seek help from a non-specialist

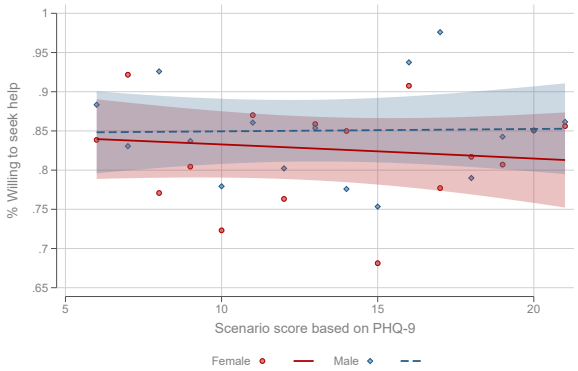


Figure 3: Gender differences in willingness to seek help

Notes. Binned scatter plot of willingness to seek help and the severity (PHQ-9) score of the hypothetical scenario, shown separately for male and female participants. The binary variables are created based on the question: If [you/NAME] were experiencing these hypothetical issues, how likely do you think it is that [you/she,he] will seek help from the following sources? For these questions, imagine that there are no constraints on the time or money that has to be spent, and no problems relating to health insurance coverage for these options. *[Very unlikely, Somewhat unlikely, Somewhat likely, Very likely]*. Seeking help from a specialist includes a answering somewhat likely or very likely to any of the following: General Practitioner solely for this purpose, a GP during a visit for another purpose or a psychologist or a therapist. Seeking help from a non-specialist includes a answering somewhat likely or very likely to any of the following: a counselor at your workplace or university, a close friend or relative or an AI-enabled mental health chatbot. Each point in the plot represents the regression coefficient from a regression of recognition on the scenario severity, controlling for treatment assignment (self vs. other), whether the subject in the "other" scenario is a male or a female and the order in which the scenarios were presented. The bands show 95% confidence intervals, with standard errors clustered by participant.

7 Tables

Table 1: Gender differences in depression recognition, accuracy and willingness to seek help

	Recognition		Accurate severity		Seek help	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.053***	-0.145***	0.068***	0.074**	-0.026*	-0.041*
	(0.018)	(0.040)	(0.022)	(0.030)	(0.014)	(0.022)
Mild	0.127***	0.095***	0.029	0.031	0.000	0.000
	(0.023)	(0.027)	(0.024)	(0.027)	(.)	(.)
Moderate	0.162***	0.095***	0.101***	0.119***	0.011	0.012
	(0.021)	(0.026)	(0.029)	(0.033)	(0.012)	(0.013)
Moderately Severe	0.191***	0.119***	0.329***	0.323***	0.010	-0.010
	(0.021)	(0.024)	(0.032)	(0.039)	(0.012)	(0.014)
Male × Mild		0.068		-0.004		0.000
		(0.047)		(0.039)		(.)
Male × Moderate		0.138***		-0.034		0.000
		(0.042)		(0.050)		(0.024)
Male × Moderately Severe		0.153***		0.014		0.042*
		(0.041)		(0.057)		(0.025)
Constant	0.774***	0.818***	-0.128	-0.129	0.974***	0.983***
	(0.065)	(0.064)	(0.085)	(0.085)	(0.047)	(0.048)
Demog. controls	Yes	Yes	Yes	Yes	Yes	Yes
Treat/Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean women	0.95	0.87	0.18	0.05	0.98	0.98
Observations	1604	1604	1604	1604	1203	1203

*Notes.*OLS regressions of depression recognition (Col. (1)–(2)), recognition accuracy (Col. (3)–(4)) and willingness to seek help (Col. (5)–(6)) on a male indicator, and interactions with the severity category of the hypothetical scenario. Demographic controls are participant age, education, employment, income level and familiarity with depression screening tools. Order controls include which of the treatment blocks was presented first. Standard errors are heteroscedasticity robust and clustered at the participant level. The mean of women at the bottom of the table corresponds to the mean outcome for women across all severity levels in Col. (1), (3) and (5) and the mean for women in the minimal severity level in Col. (2), (4) and (6). The seeking help question was not asked in the minimal depression scenarios so the constant and Male coefficient correspond to the Mild severity scenario. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Gender differences in depression recognition, accuracy and willingness to seek help on Self vs. Other

	Recognition		Accurate severity		Seek help	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.035*	-0.103***	0.071**	0.114***	-0.042**	-0.020
	(0.019)	(0.034)	(0.031)	(0.038)	(0.018)	(0.019)
Other	0.024		0.010		-0.005	
	(0.015)		(0.030)		(0.014)	
Male × Other	-0.038		-0.005		0.049**	
	(0.027)		(0.039)		(0.021)	
Other is male		0.002		0.029		-0.042
		(0.031)		(0.034)		(0.027)
Male × Other is male		0.111**		-0.130**		0.065*
		(0.050)		(0.053)		(0.035)
Constant	0.754***	0.809***	-0.125	-0.213***	0.959***	1.018***
	(0.065)	(0.104)	(0.082)	(0.073)	(0.046)	(0.051)
Demog. controls	Yes	Yes	Yes	Yes	Yes	Yes
Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean	0.98	0.96	0.21	0.19	0.98	0.99
Observations	1604	802	1604	802	1203	401

*Notes.*OLS regressions of depression recognition (Col. (1)–(2)), recognition accuracy (Col. (3)–(4)) and willingness to seek help (Col. (5)–(6)) on a male indicator, and interactions with the "Other" treatment in hypothetical scenario in Col. (1), (3) and (5) and with the "Other is male" treatment within the "Other" scenarios in Col. (2), (4) and (6). Demographic controls are participant age, education, employment, income level and familiarity with depression screening tools. Order controls include which of the treatment blocks was presented first. Standard errors are heteroscedasticity robust and clustered at the participant level. The mean of women at the bottom of the table corresponds to the mean outcome for women evaluating the "Self" treatment in Col. (1), (3) and (5) and women evaluation the "Other is female" treatment in Col. (2), (4) and (6). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ONLINE APPENDIX

A Additional Figures

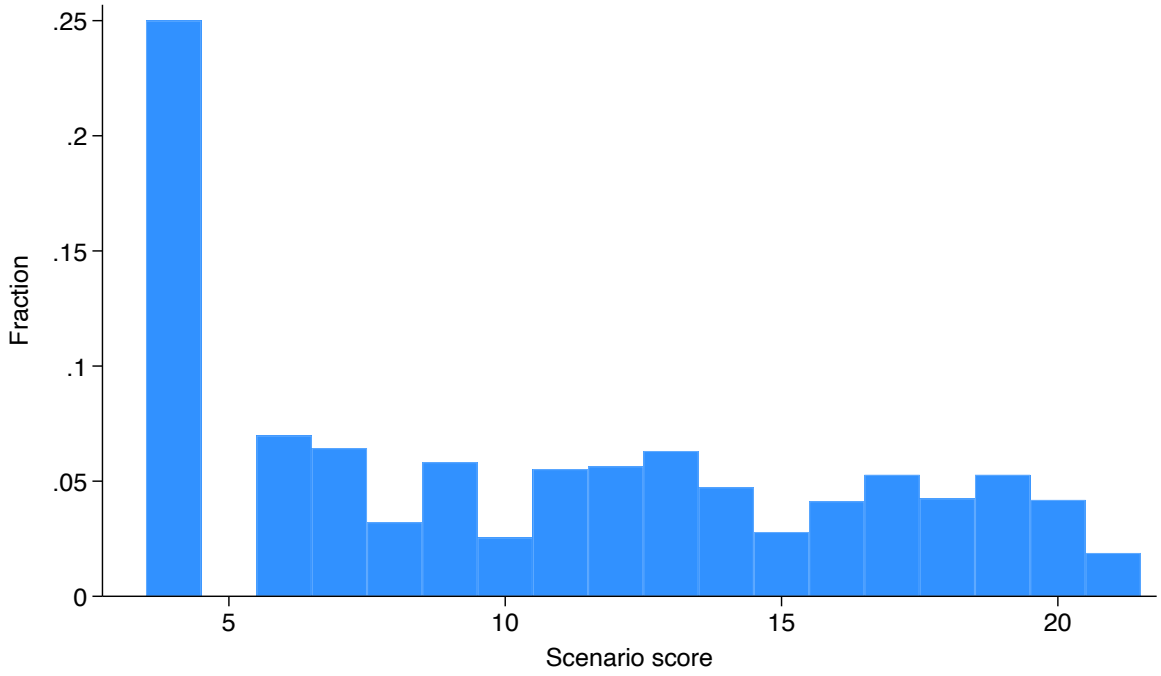


Figure A.1: Distribution of scenarios seen by participants

Notes. Fractions of scenarios presented to participants at every point of the support of the PHQ-9 score that we randomly generated.

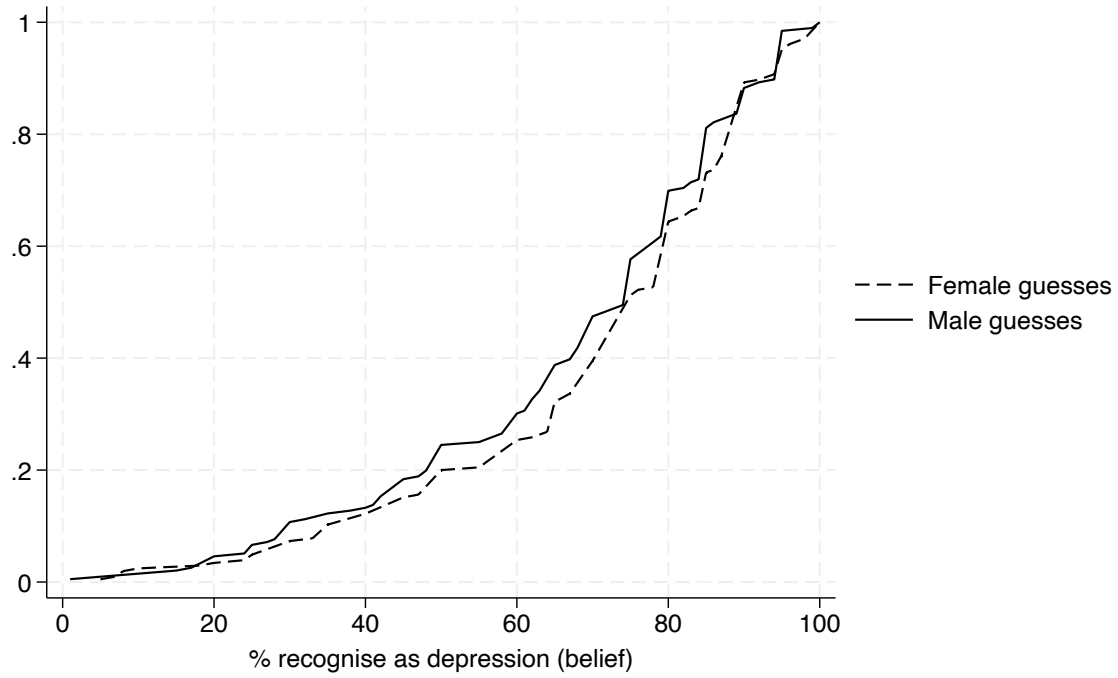


Figure A.2: Distribution of scenarios seen by participants

Notes. Empirical CDF of the guesses on the fraction of Americans recognizing depression in the scenario presented. CDFs of guesses are plotted separately for female and male participants.

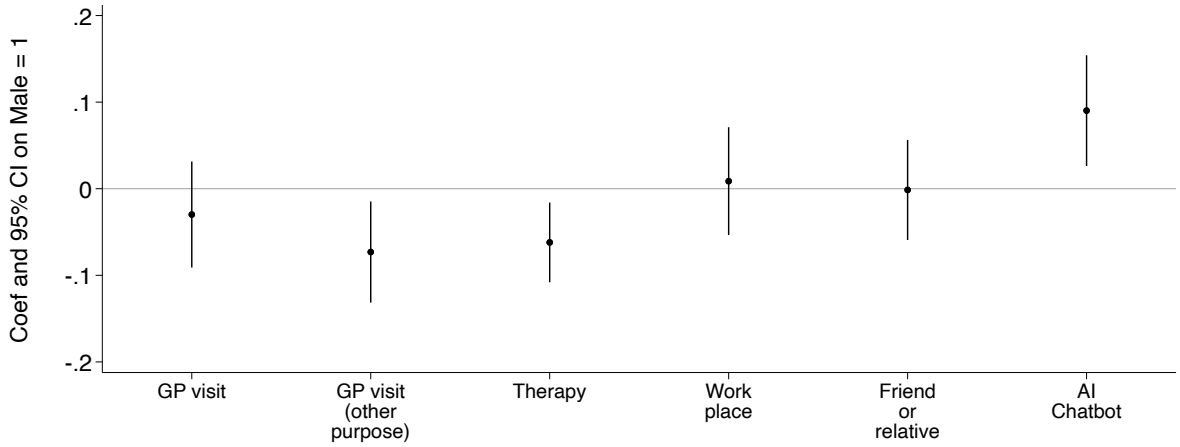
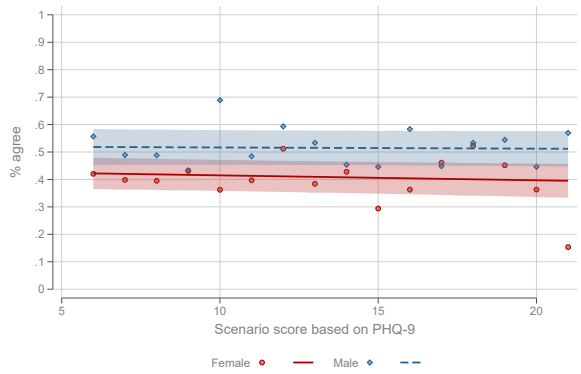


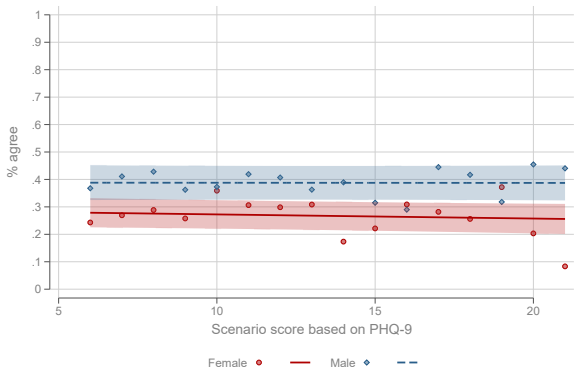
Figure A.3: Gender differences in sources of seeking help

Notes. Each point in the plot represents the coefficient from separate regressions based on regressing each of the sources on the male indicator following Equation 1. The binary outcomes are equal to 1 when the respondent responds somewhat or very likely to seek help from that source in the x-axis. The estimates show gender gaps in sources of seeking help, where a positive point estimate indicates that men are more likely to women to agree seek help from that source. We plot 95% confidence intervals along with the point estimates of the gender gaps, with standard errors clustered by participant.

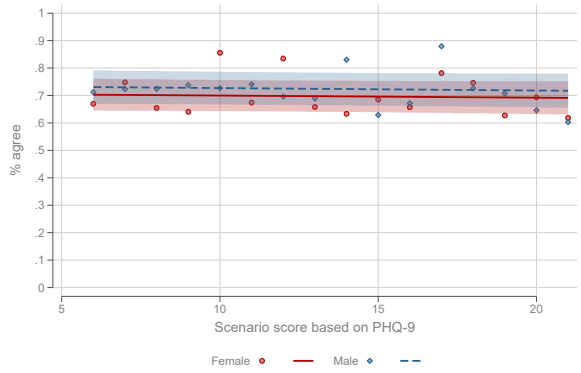
(a) The issues would go away by themselves in some time



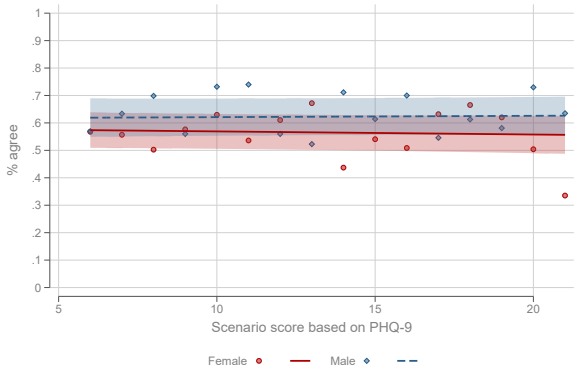
(b) The issues are not very serious and do not require treatment



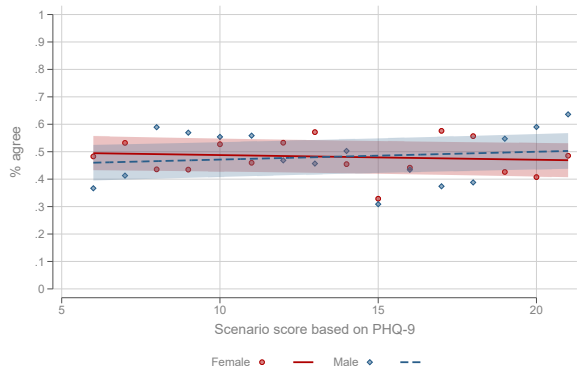
(c) Worry about what others would think



(d) Would rather deal with these issues oneself



(e) Do not think available treatments are effective



(f) Worry about the side effects of medication for depression

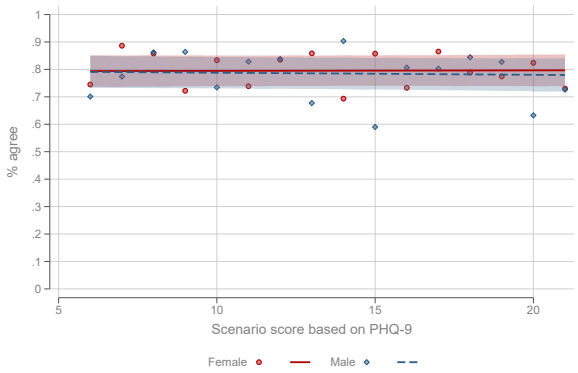


Figure A.4: Fraction agreeing with the perceptions statements in each panel heading

Notes. Binned scatterplot of the perceptions variables and the severity (PHQ-9) score of the hypothetical scenario, shown separately for male and female participants. The binary variables are equal to 1 when the respondent somewhat ore strongly agrees with the statement. Each point in the plot represents the regression coefficient from a regression of recognition on the scenario severity, controlling for treatment assignment (self vs. other), whether the subject in the "other" scenario is a male or a female and the order in which the scenarios were presented. The bands show 95% confidence intervals, with standard errors clustered by participant.

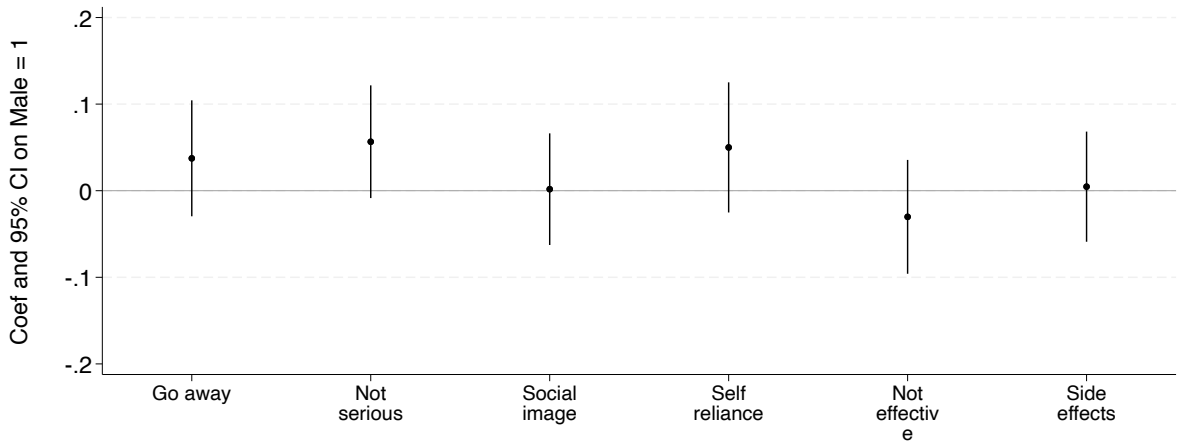


Figure A.5: Gender differences in perceptions about depression symptoms

Notes. Each point in the plot represents the coefficient from separate regressions based on regressing each of the binary perceptions outcomes on the male indicator following Equation 1. The binary outcomes are equal to 1 when the respondent somewhat or strongly agrees with the statement in the x-axis. The estimates show gender gaps in perceptions, where a positive point estimate indicates that men are more likely to women to agree with the perceptions statement. The baseline levels that each outcome takes are in Figure A.4. We plot 95% confidence intervals along with the point estimates of the gender gaps, with standard errors clustered by participant.

B Additional Tables

Table B.1: Seeking help from specialist vs. non-specialist sources

	Specialist		Non-specialist	
	(1)	(2)	(3)	(4)
Male	-0.061***	-0.106***	0.020	0.015
	(0.021)	(0.033)	(0.030)	(0.039)
Moderate	0.015	-0.010	-0.028	-0.014
	(0.018)	(0.022)	(0.023)	(0.033)
Moderately Severe	0.035*	-0.000	-0.015	-0.033
	(0.018)	(0.022)	(0.021)	(0.030)
Male × Moderate		0.052		-0.028
		(0.037)		(0.047)
Male × Moderately Severe		0.074**		0.038
		(0.037)		(0.042)
Constant	0.902***	0.923***	1.183***	1.187***
	(0.061)	(0.062)	(0.088)	(0.089)
Demog. controls	Yes	Yes	Yes	Yes
Treat/Order FE	Yes	Yes	Yes	Yes
Mean women	0.95	0.95	0.83	0.84
Observations	1203	1203	1203	1203

*Notes.*OLS regressions of seeking help from specialist sources (Columns (1)–(2)) and from non-specialist sources (Columns (3)–(4)) on a male indicator, and interactions with the severity category of the hypothetical scenario. Demographic controls are participant age, education, employment, income level and familiarity with depression screening tools. Order controls include which of the treatment blocks was presented first. Standard errors are heteroscedasticity robust and clustered at the participant level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2: Seeking help from specialist vs. non-specialist sources in Self vs. Other scenarios

	Overall		Specialist		Non-specialist	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.042**	-0.037**	-0.080***	-0.068***	0.003	0.008
	(0.018)	(0.018)	(0.025)	(0.024)	(0.036)	(0.036)
Other	-0.005	-0.004	-0.018	-0.015	0.039	0.041
	(0.014)	(0.014)	(0.021)	(0.021)	(0.027)	(0.027)
Male × Other	0.049**	0.045**	0.058*	0.048	0.054	0.050
	(0.021)	(0.021)	(0.031)	(0.031)	(0.038)	(0.038)
Recognized depression		0.119**		0.285***		0.114*
		(0.048)		(0.070)		(0.062)
Constant	0.959***	0.866***	0.909***	0.687***	1.139***	1.050***
	(0.046)	(0.063)	(0.058)	(0.085)	(0.086)	(0.105)
Demog. controls	Yes	Yes	Yes	Yes	Yes	Yes
Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean women	0.98	0.98	0.95	0.95	0.81	0.81
Observations	1203	1203	1203	1203	1203	1203

Notes. OLS regressions of seeking help from specialist and non-specialists sources combined (Columns (1)–(2)) and separate (Columns (3)–(6)) on a male indicator, and interactions with *Other* treatment assignment of the hypothetical scenario. Demographic controls are participant age, education, employment, income level and familiarity with depression screening tools. Order controls include which of the treatment blocks was presented first. Columns (2), (4) and (6) add a control for whether the participant recognized depression at an earlier stage right after evaluating the scenario. Standard errors are heteroscedasticity robust and clustered at the participant level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C PHQ-9 questionnaire and severity classification

PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use "✔" to indicate your answer)*	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

The PHQ-9 score is calculated by the simple addition of the frequencies for each symptom, with no weight for how "serious" the symptom is.

Severity classification:

- 0-4: None or minimal
- 5-9: Mild
- 10-14: Moderate
- 15-19: Moderately severe
- 20-27: Severe

D Vignette scenarios

D.1 Treatment *Self*

First, participants are introduced to the context of the vignette:

> Please imagine that you have been experiencing the following issues. Note that this is a *hypothetical scenario*. Review the list of hypothetical issues along with how often you would have experienced them **over the last two weeks**:

Three or more of the randomly selected symptoms were shown to the participant in a tabular form, along with a frequency which is one of *Several days, More than half the days, Nearly every day*.

D.2 Treatment *Other*

The scenario was introduced in the same way as in treatment *Self*. The key difference is that the subject of the scenario is a hypothetical male or female individual. We use first names to make gender identity salient. The chosen names – Michael and Jessica – are among the most popular names in the birth cohort in Texas.

> For these questions, imagine a hypothetical individual, [NAME]. [NAME] is 34 years old, lives in [TOWN in the US], and works as a marketing professional.
>
> Please imagine that [NAME] has been experiencing the following issues. Note that this is a *hypothetical scenario*. Review the list of hypothetical issues along with how often [NAME] would have experienced them **over the last two weeks**:

E Survey questionnaire

Q1. Suppose that [you/NAME] were/was experiencing the hypothetical issues at the frequencies listed above, do you think [you/NAME] would have depression? *[Definitely yes, probably yes, probably no, definitely no]*

Q2. How severe do you think the depression would be if [you/NAME] were experiencing these issues in real life? *[None or minimal, Mild, Moderate, Moderately Severe, Severe]*

Q3. If [you/NAME] were experiencing these hypothetical issues, how likely do you think it is that [you/she,he] will seek help from the following sources? For these questions, imagine that there are no constraints on the time or money that has to be spent, and no problems relating to health insurance coverage for these options. *[Very unlikely, Somewhat unlikely, Somewhat likely, Very likely]*

1. a General Practitioner solely for this purpose
2. a General Practitioner during a visit for another purpose
3. a Psychologist or a therapist
4. a counselor at your workplace or university
5. a close friend or relative
6. an AI-enabled mental health chatbot

Q4. Based on the hypothetical issues and their frequencies experienced by [you/NAME], please indicate the extent to which you agree or disagree with the following statements. *[Strongly disagree, Somewhat disagree, Somewhat agree, Strongly agree]*

1. The issues would go away by themselves in some time
2. The issues are not very serious and do not require treatment
3. I/[NAME] worry/worries about what others would think of me/her/him if they became aware that [I/she,he] had these issues
4. I/[NAME] would rather deal with these issues [myself/herself,himself] than rely on help from others
5. I/[NAME] do/does not think that the available treatments for these issues are effective
6. I/[NAME] am/is worried about the side effects of medication for depression

Q5. We conducted a similar survey with a sample of 100 Americans. The composition of respondents in this survey was broadly representative of the American population. Participants in that survey evaluated the following exact hypothetical scenario and answered whether or not they thought [NAME] would have depression if he/she were experiencing these symptoms.

List of symptoms

Of the 100 Americans who participated in that survey, how many do you think answered Definitely yes or Probably yes to the question: Suppose that [NAME] were experiencing the hypothetical issues at the frequencies listed above, do you think [NAME] would have depression? [*Number between 0 and 100*]

F Randomization procedure

Each vignette always includes the first two questions of the PHQ-9. Additional questions and symptom severities are randomly generated. The procedure to randomly generate a list of symptoms was as follows:

1. Randomly select one of the four categories. The likelihood with which categories are picked are: Mild - 1/6, Moderate - 1/3, Mod. Severe - 1/3, Severe - 1/6. Each category has an upper and a lower score bound.[9]

2. For the first two questions of the PHQ-9, select severities such that the score adds up to 4 (so one of (1,3),(2,2),(3,1)).

3. Next, generate a random sequence of numbers from 3 to 9, and pick them one by one. These correspond to questions in the scenario.

4. For each item in the sequence, pick a score from 1 to 3. This represents the severity of the symptom.

5. Check whether the total score exceeds the minimum threshold for the category. If not, repeat step (4). If so:

    - If all questions have been iterated through, stop
    - If score exceeds the max score for that category as well, stop
    - Else, flip a coin. If Heads, pick another question (step (4)). If Tails, stop.

[9]According to the PHQ-9 scoring, there are 5 severity levels: 0-4 None or minimal, 5-9 Mild, 10-14 Moderate, 15-19 Moderately severe, 20-27 Severe. We do not focus on the lowest severity level to ensure that we are presenting scenarios where some level of depression is expected.