Stigma Leads to Mental Health Information Avoidance: Behavioral Economics Approach

This version is draft. To be updated

Daehyun Kim

USC

August 2019

Abstract

This study presents experimental evidence of heightened "ego utility" in mental health dimension among individuals with mental health stigma. Theoretical expectations regarding the behaviors of decision makers with ego utility are that they may refuse to admit depression, and also avoid receiving diagnostic information that might force them to learn their true mental health states. In my randomized priming experiment, subjects read stigma-inducing messages or stigma-reducing messages, and then they are asked to report their self-assessment of their own depression severity and willingness to receive the diagnostic information. While subjects in some demographic categories do not show behavioral difference after reading different messages, others behave in a manner consistent with the theory, exhibiting more positively biased self-assessment and less willingness to receive the diagnostic information after reading stigma-inducing messages compared to reading stigma-reducing messages. The results indicate that belief bias and information avoidance might be driven by motivation of maintaining positive self-image. Also, in public health point of view, this study provides potential explanation regarding depression patients' simultaneous manifestation of denial of depression and treatment avoidance.

1 Introduction

It is reported that 30% of the American population is experiencing at least one diagnosable mental disorder [46]. Despite such prevalence of mental disorders, research shows that many

people with mental disorder delay or fail to seek help from mental healthcare system and do not receive timely care although effective treatments for a wide range of mental disorders are available [47,70]. For example, according to a study based on WHO world mental health survey data, 56% of people in the USA as of 2010 who have depression did not seek help from the formal mental healthcare system [47]. Also in the USA, even among those who eventually receive treatment for a major depressive disorder, it takes more than a decade for initial treatment contact to occur after the first onset of the disorder [71]. Although quality of life and the ability to achieve personal goals or social roles can be significantly impaired by symptoms that arise from mental disorders, those symptoms could be potentially eased by existing treatments [28]. Therefore, it is important to identify the discouraging factors for the service utilization and design effective strategies to facilitate the use of the service for those in need.

Research in public health and clinical psychology has shown that mental health stigma is one of the key factors that prevent people from seeking out timely treatment [4,23,33,62,69] [1]. Mental health stigma is a negative stereotype attached to those with mental health issue or those receive psychiatric treatment. According to research on mental health stigma, individuals who internalized so became agree with such view avoid seeking help in fear of negative consequence of visiting mental healthcare providers to occur such as discrimination or shame [23]. However, despite its importance, mental health stigma has been largely neglected in economics. Instead, following the long-held tradition that emphasizes the importance of belief on decision making, models of help-seeking for mental health issue in economics largely focus on the roles of self-perceived seriousness, in other words, subjective belief on the probability of having 'real' illness. For example, Oster et al. (2013) [58] build a model that optimistically biased belief about the probability of onset of disease in the future results in low likelihood of receiving screening test among individuals at risk for Huntington's disease.

Although above two factors, mental health stigma and the optimistic bias about having a disease, are separately explored in the area of psychology and economics, research on how those two factors influence each other and have a combined effect on the low help-seeking intention is rare. Thus far, to my knowledge, in economics a study by Bharadwaj et al. (2017) [12] is the only published study that investigates the effect of mental health stigma on patient's treatment decisions. The authors find that patients tend to under-report mental health-related symptoms while they do not under-report physical health issues, which may suggest the existence of mental health stigma. Also, the under-report of mental health symptoms is associated with reduced likelihood of seeking mental health care in the

[1] Delay or failure of help-seeking is also associated with some socio-demographic factors: males, racial/ethnic minorities, married people, and the less educated tend to delay or avoid contacting mental health services [55]. In addition, there are situational barriers to accessing services including long waiting period, lack of time or financial resources, and not knowing where to go for help [29].

future. These findings are consistent with discouraging effect of stigma on help-seeking for mental health issue which has been explored in clinical psychology research. However, this study relies on the indirect method to identify stigma, that is, whether or not mental health symptoms are under-reported, without using any direct measure of the stigma level of each subject. Therefore, it can be inconclusive whether the under-report of mental health symptoms is due to stigma and thus it is required to check if the under-report is associated with a measure of personal level stigma. Also, this study does not deal with the question of whether the under-report is solely attributable to the motivation of hiding the mental health symptoms or patients already have distorted belief about their mental health status even before the reporting is made.

In the current study, I aim to fill this gap by setting up and experimentally testing a comprehensive framework where stigma plays a significant role during the process of mental health belief formation. In my framework, mental health stigma leads individuals to believe themselves to be mentally healthier than the true state is. It is important to note that not only mental health stigma refers to the negative stereotypes each individual attaches to those with mental disorder, but also it represents the desire to keep social distance from them (See [5, 24]). Therefore, individuals with mental health stigma are motivated to believe that the probability that they currently have depression is low, which results in the optimistically biased belief. An important behavioral implication is that, out of the desire of maintaining this biased belief, they choose to ignore or even avoid available diagnostic information because the information may force their biased belief not to be feasible to maintain. This mechanism might partially contribute to treatment avoidance behavior of stigmatic individuals with mental health issues.

However, although such strategic information avoidance may explain the maintenance of biased belief, the question of how belief evolves into biased one in the first place is remained. To answer this question, I set up a possible mechanism, namely, cautious approach to information, according to which individuals are allowed to glance at a part of information and choose whether or not to proceed to receive the remaining set of information. For example, if the first page of diagnosis result sheet only contains the brief result summary (*e.g., 'you are diagnosed with depression'.*), and the detailed diagnosis results are on the next pages, then individuals might first read the brief result on the first page and make a decision whether to look at the next pages or just ignore them. Moreover, the likelihood that the individuals open the remaining pages depends on the emotional response they experienced by reading the brief result on the first page. That is, if the brief result generates relief or disappointment they are likely to continue to read or ignore remaining set of information respectively. This strategy contributes to developing optimistic bias in mental health belief because it allows individuals to be selectively more exposed to information that may bring

positive emotion. Importantly, such tendency is more prevalent among highly stigmatic individuals, which explains why they have more optimistically biased mental health belief than those with low level of stigma.

Next, I test mental health belief formation model using a series of survey experiment with subjects recruited from an Internet-based platform named Amazon Mechanical Turk (MTurk). The experiment consists of three sessions. In the first session, to perform an exploratory correlation analysis, I gathered individual level variables of depression level (CES-D score), degree of stigmatic attitude toward the depressed people, self-assessment of depressive symptom (self-expected CES-D score) and willingness to receive the CES-D test result. In the analysis, I find that, consistent with the theoretical expectation, subjects with high level of stigma tend to report their self-assessed depression level to be lower than the true states are, and also show low level of willingness to receive the CES-D test result.

In the second session of the experiment, I aim to find the existence of causal effects of stigma on under-assessment of depressive symptom, and on reduced willingness to receive the diagnosis result. To this end, I randomly divide subjects into two groups and provide them different messages, stigma inducing messages for one group and stigma reducing messages for another group. If the effects of stigma are causal, it is expected that the subjects who read stigma inducing messages show more under-assessment of depressive symptom and less willingness to receive the test result compared to those read stigma reducing messages. The result is partially consistent this expectation. That is, the effect of reading stigma inducing message compared to stigma reducing message is observed among the subjects who are married and employed, but not among other group of subjects. With my data, it is not possible to figure out the reason why the provision of different messages have expected effects only among subgroup of the subject, which might be an important future research question for designing an effective large scale anti-stigma campaign. Nevertheless, my result supports that the effects of stigma might be causal at least among some group of population.

In the last session of the experiment, I test the 'cautious approach to information' hypothesis, which is built to deal with the question on how the mental health belief evolves into the biased one. To this end, I divide diagnostic information into two pieces, the brief result (whether the CES-D score is above the cutoff for depression) and detailed result (percentile rank of CES-D score). Then the subjects are provided with brief result first and asked whether they want to receive the detailed result too. With the data, I analyze whether the content of brief result affects the demand for detailed information. The result is consistent with the theory. If subjects experience disappointment by receiving the brief result, that is, if the prior belief about the probability of being diagnosed with depression is low but brief result implies depression, they are more likely to avoid receiving the detailed result. In contrast, subjects who experience relief from the brief result tend to show higher

demand for the detailed result. Also, this tendency is more prominent among highly stigmatic subjects, which provides possible explanation on why stigmatic individuals develop optimistically biased mental health belief.

My framework is based on theoretical and experimental findings about information preference and subjective belief from economics literature. In neoclassical economics, having accurate belief is always assumed to be better than having biased belief because inaccurate belief could lead to mistaken decision. Therefore, if the source of information is credible and cost of acquiring information is negligible, individuals should not avoid receiving it. However, findings from recent economics and psychology suggest that individuals often prefer to hold biased belief and also forgo the chance of receiving information that can update the belief into accurate direction [41, 66]. The mechanisms posited to explain such information avoidance behavior include fear or anxiety of knowing reality [39, 48, 73], desire for optimism [7, 15, 45] and self-confidence management [8, 49]. Those mechanisms all assume that, not only the material outcomes, but also belief itself may affect utility. Thus, individuals are motivated to manage their belief in a way that can maximize their utility so that they develop bias in belief.

I argue that the current study have two contributions. First, I add the component of personal attitude, namely mental health stigma, as a parameter to the theory on information preference. Attitudinal traits have been largely neglected in economics perhaps because they have been assumed to be irrational part of human nature and also it is hard to obtain the variables measuring attitude. This study provides evidence that the addition of attitudinal variable might be relevant for economics research. Second, in public health perspective, the current study may provide a systematic and quantitatively tractable way to explore the complicated network of stigma, denial of having a mental illness and low demand for mental healthcare.

The rest of the paper is organized as follows. (**To be added**)

2 Background

In this section, the relevant theories and findings from economics and public health literature are reviewed.

2.1 relevant studies in economics

2.1.1 empirical

In economics, study on the effects of mental health stigma on economic impacts experienced by mental health patients is rare. To my knowledge, only two published studies directly investigates the effects of stigma. Among them one finds an evidence of mediating effect

of public's stigmatizing attitudes on job market discrimination towards those with mental illness, with more likelihood for those with mental illness to become unemployed, in countries with high level of public stigma toward mental illness [36].

The more relevant to the current study is a study by Bharadwaj et al. (2017) [12] which empirically tests the hypothesis that, out of the fear of being stigmatized, those experiencing depression are likely to hide their mental health condition when answering survey question on mental health state, and the probability of the under-report is negatively associated with the likelihood of receiving treatment in the future. This result is in line with the current study where individuals who have internalized stigmatizing attitude against those with mental disorder tend to have positively biased belief about one's own mental health condition, and also, avoid receiving diagnosis result.

However, there are considerable differences between hypotheses in the two studies. First, in the hypothesis of Bharadwaj et al., individuals know their depression status but they want to hide this to others so under-report it when answering survey questionnaire, whereas, in my hypothesis, individuals already have biased belief about one's own depression status regardless of the possibility of being discovered by others about their depression status. Second, individuals choose not to seek treatment in order to avoid stigmatization in the hypothesis of Bharadwaj et al. However, the broad term 'stigmatization avoidance' can include different motivations such as avoidance of anticipated discrimination (social dimension) or anticipated internal shame (intrapersonal dimension). For the current study, I propose a specific mechanism that can lead to treatment avoidance, which may be summarized as the motivation of not being informed of one's own depression status (i.e., diagnosis avoidance) in order to maintain preferred self-image and this will be detailed in the later sections. Furthermore, by taking advantage of experimental research design, I can collect variables of stigmatizing attitude from each subjects, which lack in the study of Bharadwaj et al., and this variables allow me to directly test my hypothesis.

2.1.2 theory

In neoclassical utility framework, with given preferences, individuals make decisions based on their beliefs about uncertainty about state of the world, thus less biased or accurate belief will lead to better decision. For example, by being well-informed of one's own health state, one can make an appropriate treatment decision or choose suitable lifestyle such as paying more attention to the diet.

Therefore, valuableness of information is an important feature in the neoclassical framework. That is, because of the benefit of having accurate belief in making decision, information always has positive or at least non-negative instrumental value as long as the source of

information is trustworthy[2]. However, in various contexts, individuals are often observed to actively avoid information, which should be considered as anomalous behavior from the perspective of neoclassical economics [41]. Moreover, not surprisingly, information avoidance is likely to be more prominent when unwelcome news is expected[3]. For example, individuals tend to avoid monitoring their financial portfolios when markets are falling [45]. Also, many researches empirically demonstrate or theorize medical testing avoidance especially when one may expect that the test results imply serious illness [39, 48, 58].

To explain information avoidance and relevant observations such as demanding information of no instrumental value, various mechanisms have been proposed[4]. Under the non-strategic decision making context[5], disappointment or regret aversion, anxiety with regard to the contents of information and optimism maintenance are among the hypothesized mechanisms (see [41, 66] for review). Even though the deeper investigation of the psychological mechanisms is beyond the aim of this study, it is important to note that many of the mechanisms posed assume the significant roles of emotion in information acquisition decision. Given that neoclassical framework is silent about the roles of emotion on decision making, some revisions to the neoclassical utility theory are required in a way to incorporate emotion in preference.

In their seminal study, Caplin and Leahy (2001) [17] extend classical utility theory and come up with a framework that can allow for derivation of utility from emotions as well as consumption outcome. In their 'anticipatory utility' framework, individuals form emotion based on their beliefs regarding uncertainty involved with unrealized outcomes of interest, and the emotion at the moment directly affects instantaneous utility experienced[6]. This framework has methodological importance because it builds a foundation of the idea that individuals derive utility from their belief, through mediating roles of emotion.

Direct implication of deriving utility from belief is that individuals may have preference for biased belief over accuracy. For example, individuals would enjoy having optimistically biased belief about one's own ability, future wealth or future health state because optimism might be more emotionally pleasurable, or even materially beneficial than objective

[2]The instrumental value of information is equivalent to the extent it can help decision making. Zero instrumental value of information happens when the information does not affect decision, for example, when decision is already made or only one available option is presented to the decision maker.

[3]When avoiding information is not viable option, individuals are also often observed to choose pay diminished attention or deny the credibility of information, thus minimally update their belief in response to the unwelcome information. For example, individuals tend not to fully update their beliefs about their intelligence when obtaining IQ test results that fall short of their *a priori* expectations, compared to the opposite case [31, 56].

[4]Information seeking is the direct counterpart of information avoidance. Individuals often seek information even when the information does not benefit for decision making (see [34, 35]).

[5]Information avoidance could be strategically driven decision in some contexts. For example, individuals might strategically choose not to know the health risk of cigarette smoking because overestimating the risk of smoking is even helpful to control their temptation to start smoking [19].

[6]It is important to note that the anticipatory utility is conceptually distinguished from expected utility. That is, anticipatory utility is experienced at the very moment based on current emotional state, whereas expected utility is simply an expectation about utility that will be experienced in the future. (see [17] for detail. Also, one of simple examples can be found from the study by Oster et al. (2013) [58])

or pessimistic perspective by enhancing performance[7]. Consequently, a majority of people are observed to answer they are better than average when evaluating themselves on positive traits such as competence [14], likelihood of business success [16] or one's life expectancy [60]. This view is consistent with dissonance avoidance theory where individuals derive utility from their identity belief, that is, the self-assessment of one's position among distribution of peers with regard to identity-related traits (e.g., ability, health, political ideology, religion) while, in order to avoid cognitive dissonance, managing their beliefs to be aligned with preferred identity by selective exposure to information [1,2,9,38]. In accordance with these ideas, a growing literature within economic theory has explored the descriptive features of people's preference over beliefs and formulated the idea that optimistic belief is optimally chosen one [8, 10, 18, 49–51].

Importantly, having some control over their belief is the underlying assumption of the idea that the biased belief is in fact motivated. Following the well-established framework that belief is formed and updated by receiving information, selective acquisition of information is regarded as a mechanism by which manipulation of their own beliefs can be possible. Namely, individuals may select the sources of information that could confirm or shift their current beliefs toward the preferred beliefs [37, 43]. The tactics that people may use to be selectively exposed to information include avoidance of potentially threatening information source [39], dismissing the credibility of unwelcome information [63], biased interpretation of information [31, 56, 61], motivated inattention [45] or motivated forgetting [11]. Taken together, when a person has a preference toward a specific belief, he might end up having belief with self-serving, often optimistic, bias, and simultaneously, is likely to show selective information exposure behavior in order to maintain or achieve the preferred belief. The consequence (biased belief) and underlying motivations (maintaining preferred self-image) of each tactics may be overlapped and I will focus on the avoidance of information in this study.

However, individuals rarely become unrealistically optimistic. Individuals also take into consideration the cost of inaccurate belief when forming belief because belief bias may lead to mistaken action. For example, overly optimistic belief about one's health state may result in unhealthy behaviors or missing timely treatment. Economics literature has built theoretical arguments on the course of 'optimal belief' formation in which individuals choose the optimal level of belief bias by weighing the immediate anticipatory utility benefits of having optimistic belief against the costs of inaccuracy [8, 10, 15, 49, 64].

Among this strand of studies, 'optimal expectation' model by Brunnermeier and Parker (2005) [15] is notable in that it explicitly distinguishes between objective prior belief and

[7]There are a lot of evidence and theoretical arguments about the benefits of optimism. For example, while pessimism may cause anxiety, optimism can enhance confidence and willpower, thus result in better performance [8, 22]. Also, there is a research evidence that optimism bias is associated with better mental health [67].

subjective (optimal) prior belief, while allowing for the possibility that individuals restrain their belief bias by exerting cognitive resources when the stakes are large. It is consistent with cognitive dual-process theory in which human decision making relies on impulsive, automatic process of limbic brain system when doing habitual tasks or stakes are low, however starts activating rational, analytic process of prefrontal cortex (PFC) in brain when stakes are large enough to require some conscious reasoning [13, 44]. This model provides a parsimonious framework on prior belief formation, in which individuals first start as if they have objective belief then choose the degree of optimal bias, finally, the optimal prior is determined as the sum of objective prior and optimal bias. Experimental study that directly demonstrates the validity of this model is rare, probably because of the difficulty in measuring bias in prior beliefs[8]. Among empirical literature, Oster et al. (2013) [58] demonstrate that optimal expectation model is consistent with the behaviors of individuals at risk of Huntingon's diease such as optimistic bias and low screening test rate. In the next section, I describe this theory with application of mental health belief.

3 Conceptual Framework

In this section, following the optimal expectation model by Brunnermeier et al. (2005) [15], I build a model of mental health belief formation according to which individuals tend to motivatedly underestimate the probability of having clinical depressive disorder, which is comparable to under-evaluation of depressive symptoms. Also, I add some modification of the optimal expectation model in a way to allow the model to demonstrate the effect of the agent's own stigmatizing attitudes toward depressed people (internalized-stigma) on the biased belief, with more underestimation of the risk being displayed among people with higher level of internalized-stigma. Moreover, in order to maintain biased belief, individuals intentionally choose to forgo relevant information on their mental health states such as depression test result.

3.1 Choice of Optimal Belief

First, I present a model for how individuals form underestimation-biased beliefs about the probability of having clinical depression and how the stigmatizing attitudes affect the bias. The course of events and actions consist of two stages, $t \in \{0, 1\}$. Individuals choose optimal belief on the probability of being categorized as depression patient in the first stage and choose appropriate health behavior in second stage in accordance with the optimal belief chosen at the first stage.

[8]Only experimental study that I find to directly test the optimal expectation model by Coutts (2019) [27]. However, this study does not find support for optimal expectation model. According to the optimal expectation model, belief bias is expected to be reduced when the belief accuracy is associated with higher value of reward, but such anticipated effect is not observed in the study.

In the first stage (t=0), the agent forms an subjective belief q_0 on the probability of being categorized as a depression patient. This subjective belief is called 'optimal' belief because the choice of q_0 is determined by optimization process during which individuals consider both the loss of future utility resulting from biased belief and the current felicity of having optimistic belief rather than making purely random error in assessment of risk. The utility function used in optimization process consists of belief-based anticipatory utility as well as the standard expected utility. Inclusion of belief-based anticipatory utility in the decision problem comes from the idea that individuals derive contemporaneous utility directly from their beliefs as suggested by Caplin and Leahy (2001) [17].

There are two important assumptions in modeling this process. First, the degree of discomfort from the belief of possibly having clinical depression becomes higher as the agent has the greater degree of stigmatizing attitude. This assumption implies that the possibility of being defined as a depression patient might be deemed more threatening to maintaining their preferred self-identity among those have negative view against depressed people [9]. Second, following the framework of optimal expectation model, individuals choose their subjective probability of having clinical depression based on the objective probability. In other words, the optimal expectation model posits that individuals behave as if they at first know the objective probability but choosing optimal level of optimism bias, and then forget the objective probability and live with the subjective belief. This framework is consistent with the idea of dual process theory in psychology, according to which human thought is, in the most cases, dominated by automatic, unconscious process of the limbic system which is related with formation of biased belief, but when making important decision, by exerting some cognitive resources, controlled and conscious process of the prefrontal cortex become more prevalent so that individuals can recall the objective probability and make a better decision [13].

Let $s \in \{0, 1\}$ be an element of a binary state space where $s = 1$ represents the individual has clinical depression and $s = 0$ represents being mentally healthy. Also, the variable D represents the objective degree of depressive symptom which could be measured by established depression test. The objective probability p is a probability of being diagnosed with depression if objective assessment of mental health state is administered, which is monotonically increasing in the degree of depressive symptom D. Apart from the objective probability of having clinical depression, individuals hold subjective belief q_0 about the probability of having clinical depression, which might be formed through unconscious process of evaluation of various events experienced in daily life such as experience of low mood spell. In the course of forming subjective belief about the probability of having depression, the individual's attitudes about having depression might have effect as well as the current degree

[9] See [9, 10] for more detailed argument on how identity belief is incorporated in the anticipatory utility framework.

of depressive symptoms (D) does. The most relevant type of attitude towards depression is stigma, that is, the degree that the individual negatively views depressed people. The degree of stigmatizing attitudes is denoted as *stig*. This setting can be summarized as below equations. For individual i, the variables p_i and q_{i0} follow the relation :

$$\begin{aligned} p_i &= Pr(s_i = 1|D_i) \\ q_{i0} &= p_i + B_i \\ B_i &= f(stig_i) \end{aligned} \tag{1}$$

where the variable B_i (bias) represents the difference between objective probability of having clinical depression and subjective probability, namely, bias in belief. The individual is defined as optimistically biased if $B_i < 0$. Also, the function f summarizes the mechanism through which bias in belief arises.

The below formulation is optimization problem the individual faces in t=0. Denote the action chosen in stage t=2 is a. Utility is a function of chosen action $a \in [0, 1]$, the realized state s, given the degree of stigmatizing attitude *stig* and the depressive symptom D and is denoted as $u(a, s|stig, D)$. Utility loss is minimized when action is state-congruent but taking action suitable for depression patient generate shame and regarded as threatening to manage positive self-image.

Formally, individuals choose $q_0 \in [0, 1]$ to maximize

$$U(q_0|p(D)) = -\underbrace{E_{p(D)}[Loss(\hat{a}(q_0), s|D)]}_{\text{expected health loss}} - \underbrace{\gamma \cdot stig \cdot \hat{a}(q_0)}_{\substack{\text{emotional cost of} \\ \text{taking action}}} \tag{2}$$

where the function *p(D)* is $Pr(s = 1|D)$ and $Loss(a, s|D)$ is the health loss function and γ is the weight parameter for belief-based anticipatory utility. Also, $\hat{a} = \text{argmax}_a E_{q_0}[-Loss(a, s|D)]$, which implies individual choose action in t=1 based on their subjective belief.

The health loss function is defined as follows.

$$Loss(a, s) = \begin{cases} (a - s)^2 & \text{if} \quad s = 0 \\ W(a - s)^2 & \text{if} \quad s = 1 \end{cases} \tag{3}$$

where I assume that $W \geq 1$.

Then, solving for a yields $\hat{a}(q_0) = q_0 W / \{q_0(W - 1) + 1\}$ which is increasing in q_0 and W, that is, the individuals put more effort to manage their mental health risk if they think they have a higher probability of developing depression or the cost of under-treatment is higher.

Next, I derive the optimal belief chosen in the stage t=0 with backward induction approach.

PROPOSITION 1 (underestimation of risk): Individuals always optimally choose q_0 such that $q_0 \leq p$.

PROOF:

Because the action function $\hat{a}(q_0)$ is monotonically increasing in q_0, there exists its inverse function $q_0(\hat{a})$. Then utility is:

$$U(q_0(\hat{a})|p) \;\; = \;\; - \;\; [pW(1-\hat{a})^2 \;\; + \;\; (1-p)\hat{a}^2] \;\; - \;\; \gamma \cdot stig \cdot \hat{a}$$

Within the possible range of action choice ($a \in [0,1]$), The value of $\hat{a}$ which maximizes above equation is given by:

$$\hat{a} \;\; = \;\; \frac{max\{0, \;\; -\gamma \cdot stig + pW\}}{p(W-1)+1} \tag{4}$$

Because $\hat{a}(q_0) = q_0 W / \{q_0(W-1)+1\}$,

$$\frac{max\{0, \;\; -\gamma \cdot stig + pW\}}{p(W-1)+1} \;\; = \;\; \frac{\hat{q_0} W}{\hat{q_0}(W-1)+1} \tag{5}$$

Then,

1) if $\quad stig \geq pW/\gamma : \quad \hat{q_0} = 0 \leq p.$

2) if $\quad stig < pW/\gamma :$

Suppose $\hat{q_0} > p$, then $\frac{\hat{q_0} W}{\hat{q_0}(W-1)+1} > \frac{pW}{p(W-1)+1} \geq \frac{-\gamma \cdot stig + pW}{p(W-1)+1}$, which is contradictory to the setting $\frac{\hat{q_0} W}{\hat{q_0}(W-1)+1} = \frac{-\gamma \cdot stig + pW}{p(W-1)+1}$. Therefore, $\hat{q_0} \leq p$.

The proposition 1 implies that individuals optimistically distort their subjective belief about the probability of having depression from the objective probability in order to avoid applying stigma to themselves. This belief distortion might give them current felicity but result in sub-optimal action such as under-treatment.

Also, it is important to note that the degree of optimistic distortion of belief is prominent when *stig* is high, or W is low, which can be shown by solving the equation (5) for $\hat{q_0}$. Since the calculation is tedious and not an aim for this section, I skip the formula for comparative statics of $\hat{q_0}$ with respect to those two variables. Instead, the figure 2 displays these arguments graphically: when other variables are fixed, (1) high degree of stigma leads to more underestimation bias (panel a), (2) low self-assessed vulnerability to depression leads to more underestimation bias (panel b). This result is intuitively natural: (1) high level of stigma is associated with reluctance to accept depression (2) if optimism is costly, individuals become less optimistic.

3.2 Demand for mental health diagnosis information

One way to maintain biased belief is avoidance of belief update by not receiving relevant information. In this section, I describe a model according to which those with optimistically biased belief and high level of stigma avoid self-diagnostic information.

I start with defining a utility function of information according to which the decision of receiving diagnosis information determined. In building the utility function, I hypothesize that individuals care both about the instrumental value of information, that is, the degree that individuals perceive the information is helpful in future decision making, and emotional consequences of receiving information such as disappointment, relief or anxiety.

The decision problem that the individual faces is the choice of screening test for mental health state. There are three options available: 1) not taking the screening test, 2) receiving the screening test result only and 3) receiving the screening test result and additional information such as severity or curability of the recipient's mental health issue[10]. The degree of detail of each test option is denoted as λ_0, λ_1 and λ_2, respectively.

The first part, the perceived instrumental value of information is hypothesized as increasing in the degree of depressive symptom (D) and the degree of detail of the information (λ). Therefore, the instrumental value of information with degree of detail λ for those with depressive symptom D is denoted as $f(\lambda, D)$ where $f_D > 0$, $f_\lambda > 0$ and $f_{\lambda D} > 0$. The rationale for increasing instrumental value of information for the more highly depressed ($f_D > 0$) is that the cost of being uninformed of one's health state becomes riskier as the severity of symptom increases. In other words, the failure in choosing appropriate disease-specific lifestyle including receiving treatment is more destructive for more serious disease such as cancer than lighter disease such as flu. Therefore, as the individual experiences the depressive spell more often (high value of D), he might feel more need of knowing his true mental health state. Next, it is plausible to assume that information that has more detailed contents has more instrumental value ($f_\lambda > 0$). Last inequality ($f_{\lambda D} > 0$) comes from the idea that the additional detail of diagnostic information might be larger among those highly depressed.

Next, individuals also care for the emotional consequences of receiving information, as well as the instrumental value of the information. To model the emotional value of information, I employ two well-documented hypotheses from information preference literature. First, individuals experience disappointment or relief depending on the signal received [6, 51, 54]. Second, receiving information leads individuals to pay more attention and it generates specific, in this case, unpleasant emotion [39,42]. Additionally, I add a component that all of this emotional reactions are influenced by the degree of stigmatizing attitude of the individual

[10]By receiving only the test result, what the individual becomes to know is whether he is categorized as depressed person or not. Additional information is required to more rigorously assess the symptoms.

because stigma might reflect the subjective undesirability of having depression.

Formally, the utility of receiving information with the degree of detail λ is given below:

$$V(\lambda|q_0, D) = \underbrace{f(\lambda, D)}_{\substack{\text{1) instrumental} \\ \text{value of information}}} \underbrace{- \beta_\lambda^1 \cdot stig \cdot \mu(s - q_0)}_{\substack{\text{2) emotional reaction:} \\ \text{disappoint or relief}}} \underbrace{- \beta_\lambda^2 \cdot stig}_{\substack{\text{3) emotional reaction:} \\ \text{attention-based emotional cost}}} \tag{6}$$

where $s \in \{0, 1\}$ is the test result such that s=0 and s=1 indicates healthy and depression, respectively. Also, q_0 is a prior belief which has been optimally chosen according to the model in proposition 1. In this equation, 1) $f(\lambda, D)$ is instrumental value of information. 2) $\mu(y)$ is a psychological disappointment-relief function which activates when receiving diagnosis result and is weighted by constant α_λ and the degree of stigma. Also, this function is assumed to be $\mu(0) = 0$, strictly increasing in y and has disappointment aversion property, that is, $\mu(y) > -\mu(-y)$ for any $y \in (0, 1]$. The constant β_λ^1 is assumed to be non-negative and increasing in λ, which captures the idea that more detailed information gives rise in more attention paid toward the information so that the degree of emotional reaction becomes larger. 3) Next, β_λ^2 is utility loss from paying attention to uncomfortable information which is also weighted by stigma. The value of β_λ^2 is also increasing in the value of λ. Overall, the emotional reaction to information becomes prevalent in the utility if the degree of stigma is high which is consistent with the situation that stigma generates shame or anxiety if the individual applies the negative attitude toward oneself.

The important premise from optimal expectation is that individuals can recognize himself to have biased belief when exerting cognitive resources. Consistent with this view, I assume that when facing the decision problem about choosing the degree of detail of information, individuals form expected utility based on the objective probability of having depression, not based on the optimal prior belief. In other words, individuals behave as if they know they are optimistically biased when making the decision. They are willing to exert the cognitive resources because the contents of information might negatively affect their self-image and generate shame in the form of disappointment.

With this assumption, the expected value of receiving information λ is:

$$E_p[V(\lambda|q_0, D)] = f(\lambda, D) - \beta_\lambda^1 \cdot stig \cdot [p \cdot \mu(1 - q_0) + (1 - p) \cdot \mu(0 - q_0)] - \beta_\lambda^2 \cdot stig \tag{7}$$

For simplicity, it is assumed that $f(\lambda, D) = \alpha_\lambda \cdot D$ with α_λ being increasing in λ and $\mu(y) = (1 + \gamma) \cdot y$ if $y > 0$ (disappointment) where $\gamma > 0$ is the disappointment constant, and $\mu(y) = y$ if $y < 0$ (relief). Then, the above equation becomes:

$$\begin{aligned} E_p[V(\lambda|q_0,D)] &= \alpha_\lambda \cdot D - \beta_\lambda^1 \cdot stig \cdot [p \cdot \mu(1 - q_0) + (1 - p) \cdot \mu(0 - q_0)] - \beta_\lambda^2 \cdot stig \\ &= \alpha_\lambda \cdot D - \beta_\lambda^1 \cdot stig \cdot [\gamma \cdot p(1 - q_0) + (p - q_0)] - \beta_\lambda^2 \cdot stig \\ &= \alpha_\lambda \cdot D - \beta_\lambda^1 \cdot stig \cdot [G(\gamma, p, q_0) + (p - q_0)] - \beta_\lambda^2 \cdot stig \end{aligned} \tag{8}$$

where $G(\gamma, p, q_0)$ is disappointment aversion function defined as: $G(\gamma, p, q_0) = \gamma \cdot p(1 - q_0)$ [11]. Also, the expected utility of receiving no information (λ_0) is normalized to 0.

Then, individuals choose the degree of detail of information that will give the highest expected utility. If there are three types of information, that is, not receiving any information (λ_0), the diagnosis result only (λ_1) and the bundle of diagnosis result and other more specific information about the individual's mental health state (λ_2). The decision problem that individuals face is given as below:

$$\begin{aligned} &\lambda^* = \text{argmax}_\lambda E_p[V(\lambda|q_0, D)] \qquad\qquad \text{where} \\ &E_p[V(\lambda_0|q_0, D)] = 0 \\ &E_p[V(\lambda_1|q_0, D)] = \alpha_{\lambda_1} \cdot D - \beta^1_{\lambda_1} \cdot stig \cdot [G(\gamma, p, q_0) + (p - q_0)] - \beta^2_{\lambda_1} \cdot stig \\ &E_p[V(\lambda_2|q_0, D)] = \alpha_{\lambda_2} \cdot D - \beta^1_{\lambda_2} \cdot stig \cdot [G(\gamma, p, q_0) + (p - q_0)] - \beta^2_{\lambda_2} \cdot stig \end{aligned} \tag{9}$$

Since the constants α_λ, β^1_λ and β^2_λ are increasing in the value of λ, the expectation from this formulation is that individuals are likely to seek more detailed information as having more severe depressive symptom, and avoid detailed information when they have high level of stigmatizing attitude or when the optimistic bias is large. Also, this formulation allows for the use of ordered logit or multinomial logit method to estimate the model for choice of screening test.

3.3 Mechanism for developing underestimation-bias

Above theoretical models offer expectation about the effects of stigma on having optimistic bias and information avoidance, however, those models do not deal with specific mechanism for how individuals develop the bias in self-evaluation of one's mental health state. Indeed, the state of not being fully informed implies inaccuracy in self-evaluation rather than having systematic bias.

In this section, I describe a simple model in which which can also explain the systematic (also motivated) bias in belief and can be testable by the experimental framework used in this study. In sum, when information is presented as a bundle, that is, information contain more than single content (e.g., *depression diagnostic result can contain both test result regarding whether or not he has depression and also degree of severity*), even when individuals first have decided to receive the information bundle, they choose to opt out from receiving further information if the earlier revelation of part of the information bundle is disappointing. This mechanism is consistent with that individuals are selectively exposed to information in motivated way an thus developing optimistic bias in self-evaluation of one's

[11] The function $G(\gamma, p, q_0)$ is corresponding to the degree of fear about potential disappointment due to disappointment aversion property. Apart from the effect of disappointment constant γ, the value of this function becomes larger when p is high and q_0 is low, which is equivalent to the increased likelihood of disappointment.

depressive symptoms.

First, suppose an individual already has made a decision to receive a bundled information, which contains both depression test result (signal A) and more detailed assessment of the symptom (signal B) [12]. Timeline is as follows. First, the individual has expectation about the realization of signal A, which is equivalent to prior belief. Next, he receives the test result (signal A) and derive psychological utility W_A based on the result. Lastly, he forms psychological utility of receiving signal B which is denoted as W_B, and then make a decision whether to opt out from receiving further information. The value of W_B is a function of the individual's stigma level, the signal A, and W_A.

The signal A, that is, the test result which is denoted as σ_A and can have two value: $\sigma_A = 0$ represents not having clinical depression and $\sigma_A = 1$ implies depression. The emotional utility function the individual derives at the time of receiving test result is given by $W_A(\sigma_A|q_0) = -stig \cdot \mu(\sigma - q_0)$. The emotional reaction function $\mu(\cdot)$ is the identical one with the emotional reaction function appeared in the equation (7). Especially, the negative and positive value of W_A indicate disappointment and relief, respectively. Again, the emotional reaction is weighted by the degree of stigmatizing attitude.

Next, the individuals recompute the expected utility of receiving signal B. Even though he has already formed the expected utility of receiving the information bundle (signal A + signal B), the current computation of the utility of receiving signal B follows different mechanism because the signal A which is already realized affects the emotional value of receiving the additional part of information. Formally, the utility of receiving signal B is given by:

$$\begin{aligned} E[W_B|\sigma_A, p_0, stig] = & \underbrace{\theta \cdot \sigma_A}_{\text{instrumental value}} + \underbrace{W_A}_{\substack{\text{attention based psychological utility} \\ \text{: disappointment-relief expereienced} \\ \text{when receiving } \sigma_A}} + \underbrace{C}_{\substack{\text{utility of} \\ \text{curiosity fulfillment}}} \\ = & \quad \theta \cdot \sigma_A \quad - \quad stig \cdot \mu(\sigma_A - q_0) \quad + \quad C \end{aligned} \tag{10}$$

On the other hands, the expected utility of not receiving signal B is normalized to be 0. Therefore, the individual proceed to receive the remaining set of information when $E[W_B|\sigma_A, p_0, stig] > 0$.

The first term in the equation (10) is the instrumental value of receiving signal B. Inclusion of this term postulates that if the test result indicates depression ($\sigma_A = 1$), individuals might feel the need of better understanding of his symptom so that he would seek more detailed information. The constant θ represents the responsiveness of receiving the result of having depression on the perceived need of more detailed information. The second term is

[12] In my experimental design, the signal B is the percentile ranking of the depression test score. The depression test result (signal A) only informs the participant of whether or not his score meets depression criteria, without informing detail of how severe his depression state is compared to other people.

attention-based psychological utility which comes from the emotional state associated with paying attention to the signal B, which is simply assumed to be equivalent to W_A. If the individual experienced disappointment when receiving signal A, he might want to divert his attention from the depression test and escape from the unpleasant situation. On the other hands, if he felt relief, he may want to continue to pay attention to the remaining information [13]. Finally, last term $C > 0$ represents the benefit from fulfilling curiosity, or it can viewd as the disutility from forgoing the information that first intended to receive.

This formulation offers expectation about the behavior of individuals when facing decision situation whether to proceed to receive the remaining part of information bundle after receiving the first part of information. First, the individuals are likely not to change their decision to obtain the information bundle if the first revelation of partial information implies depression because it may enhance the perceived instrumental value of remaining information. Second, however, the first effect is likely to be offset if the first set of information bundle induces disappointment, the individual might change the first decision and opt out from receiving more detailed information. Also, this decision change is more likely to occur for those who have high degree of stigmatizing attitude toward depression.

It is important to note that this mechanism may contribute to the development of optimistic bias among those have high degree of stigma as in the following way. First, individuals face decision problem whether to attend to or receive information that consists of a 'set' of signals about the true state of the individual's mental health [14]. Even though those with stigmatizing attitude toward depression tend to avoid such information or choose less detailed information as discussed in the section **3.2**, it does not necessarily lead to systematic optimistic bias in self-evaluation of symptom. The systematic belief bias, if Bayesian update is assumed, arises only when the individuals are selectively exposed to information. The theoretical argument discussed in this section suggests one of the potential sources of why such selective exposure to information takes place. Having been disappointed by signal A, by not receiving signal B, the further update of the belief does not occur. Consider the opposite is likely to happen if relief is experienced by receiving signal A, that is, individuals seek more information from the same source if receiving additional information may allow them to update their belief in more optimistic way. Then the belief would be continually evolve into optimistically biased one by making similar choices in the course of life where

[13] The alternative explanation is based on the concept of disappointment aversion. Suppose that σ_A and σ_B are positively correlated given the true state s. It is a plausible assumption because both signals are released from the same information source. For example, in the experimental design I use for this study, after taking screening test, σ_A is the binary result on whether or not the subject's test score is above the cut-off point for clinical depression, and σ_B is the level of depressive symptom compared to others. Since both signals are based on a single test result, the assessment errors should be reflected in both signals, that is, they must be positively correlated given the true state. Therefore, if a subject experiences disappointment when receiving signal A, it is highly likely that signal B also generates disappointment so the subject might avoid receiving signal B.

[14] Diagnostic information can contain more than one type of signal because there are many ways to design the signal structure of information. For example, depression can be assessed both as a form of a binary assessment (healthy/depressed) or degree of severity.

individuals face the decision problems described in this section as various forms of activity participation decision as well as the decision of receiving treatment or diagnosis [15].

4 expectation with regard to behavior

to be edited

EFFECT of STIGMA

A. stigma causes optimistic (under-evaluation) bias regarding one's degree of depressive symptom.

from equation (5).

(A) Individuals form optimistic belief. The degree of such bias is greater for those have high level of stigma toward depressed people.

B. stigma causes avoidance of diagnostic information.

from equation (7).

(B) Individuals who have high level of stigma toward depressed people tend to avoid screening test, or chooses less detailed information source.

(B)' Individuals who have large degree of optimism bias tend to avoid screening test, or chooses less detailed information source. (Out of disappointment avoidance motive)

C. stigma causes individuals to stop receiving further detailed information if the first set of partial revelation of information disappointing.

5 Experiment design

The experiments consist of three parts. In the first set of experiments, I collected cross-sectional samples in order to perform motivational analysis to test whether the behaviors are consistent with theoretical predictions, which are (1) the positive association between stigmatizing attitudes and underestimation of one's own depressive symptom in self-evaluation and (2) the negative association between stigmatizing attitudes and demand for diagnostic information.

In the second experiment, in order to perform the causal analysis of the effects of stigma on the underestimation of symptoms and information demand, I employed priming ap-

[15] Similarly, Kőszegi (2006) [49] also assumes engaging in activities can generate signals about one's own ability so that individuals avoid an activity if it may force their optimistic belief or self-confidence to be broken down.

proach, where I randomly divided subjects into three groups and presented each of them different messages: stigma inducing messages (stigmatized group), stigma reducing messages (destigmatized group) and stigma-irrelevant messages (control group), respectively. Next, I collected data on self-evaluation of symptoms and demand for diagnostic information.

The subjects of last set of experiments is the same individuals who took part in the first experiment. This experiment is designed for testing the hypothesis about the mechanism by which stigmatized individuals develop optimistic bias (**Hypothesis 3**). To test this hypothesis, among those who had chosen to obtain information on both depression screening test result and depression percentile score among the participants, I presented the screening test (C-ESD) result first, and asked them again if they were still willing to receive information on the percentile score to see if the disappointing screening test result might deter them to seek out for further information.

5.1 Sample

The protocol was reviewed by an IRB from University of Southern California. Subjects were recruited through Amazon Mechanical Turk (MTurk). The subjects volunteered to participate in the surveys by clicking the link I posted on the website. Because this study does not specifically aim for exploring the behaviors of depression patients, I did not set any participation criteria about the level of depressive symptoms or depression treatment history. However, I limited the participation only to the people living in the USA. Also, because of the possibility that some individuals might be sensitive about reporting their depression related information, I displayed the message at the beginning that they could freely close the survey page and cancel the participation at any time. The payment for participation was between \$0.2 to \$0.5.

5.2 Design

5.2.1 Experiment 1: Cross-sectional Design

This experimental procedure is designed to collect data on the observational patterns of behavior with particular focus on the level of stigmatizing attitude and its effect on the underestimation of one's own depression symptom and the demand for diagnostic information.

Before collecting the data on depression and stigma related information, as it is possible that their socioeconomic status might affect the behaviors and those factors should be controlled for, I first asked the participants for their socio-demographic information which includes age, gender, marital status, race, level of education, level of income, employment status and the number of previous experience of treatment for mental health issues.

Next, I gathered the depression related information of the participants. This information

consists of three variables. First, right after providing the socio-demographic information, I asked the participants to report their prior belief about the probability of being diagnosed with depression if taking a depression screening test (*prior belief*) in the unit of percent probability. Next, I measured the degree of depressive symptoms (*objective depression level*) of each individual by asking them to answer the Center for Epidemiologic Studies Depression Scale (CES-D) which is commonly used questionnaire for screening for deprssion (**CESD**). When answering this questionnaire, the participants were not informed that the questions are from screening test for depression because under-reporting might happen if they know it, especially among those have high degree of stigmatizing attitude. Last, after finishing the C-ESD test, each participants were asked to report their self-estimation of their percentile score of depression level among the survey participants (*self-evaluation*). In order to prevent confusion, I displayed the instruction that the high value of percentile score implies more serious depression and low percentile score means less degree of depression. Because the self-evaluation is one of the important variables, in order to improve accuracy of the self-reporting, I incentivised the participants by offering bonus payment ($1) if their estimation was within the top 5% in terms of accuracy [16].

One of the aims of this study is to test whether depression stigma induces optimistic bias in self-evaluation of one's depressive symptoms. The *self-evaluation bias* is simply defined as the difference between the self-estimated percentile score and true percentile score (*true percentile score − self-estimated percentile score*). If the self-evaluation bias is negative, the participants is deemed as showing underestimation bias of depressive symptoms, and *vice versa*.

Next, I measured the stigmatizing attitudes of the participants by asking them to answer these two questions: (1) "*Some people think the cause of depression is lack of willpower. Do you agree with this statement?*" and (2) "*Some people think depressed people are not pleasant to be around. Do you agree with this statement?*" The participants could indicate their degree of agreements to each of the question by choosing one of 4 choices which are 'agree', 'somewhat agree', 'somewhat disagree' and 'disagree'. The first question is designed to measure the degree that each participant has internalized the public stereotype that depression is associated with weakness in personality (*stigma-will*), and the second question measures the degree of social undesirability each participant attaches to depression (*stigma-social*). The score of each type of stigma is computed by indexing the answers to number such as 'agree'=3, 'somewhat agree'=2, 'somewhat disagree'=1 and 'disagree'=0. Lastly, the composite score of stigma is defined as the sum of both types of stigma measures (*stigma = stigma-will + stigma-social*) .

In the next part of the survey, I measured the demand for diagnostic information (*de-*

[16]The accuracy of self-evaluation is measured by the absolute difference between the true percentile score based on the C-ESD test result and the self-estimated percentile score.

mand). Each participant could choose one of the three options: (1) "*I want to know both my depression test result and my depression percentile score among all the participants.*", (2) "*I want to know my depression test result only*" and (3) "*I do not want to any information on my depression state*". I clarified that the wanted information will be displayed on the last page of the survey depending on the choice. Because there are no monetary cost associated with receiving information and also the time spent to check out the contents of the information is minimal (less than 15 seconds), those who do not choose option (1) might be categorized as showing information avoidance [17]. Therefore, choosing option (1) and option (3) implies strong demand for information and information avoidance, respectively, while option (2) implies slight degree of information avoidance.

As such, I collected the relevant variables including objective depression measure, subjective evaluation, stigma measure and demand for information. Apart from those information, I collected some additional information from the participants which might be interesting to explore their relationships between above variables. First, I asked the participants their attitude toward mental health related topic which include the expected ability to cope with depression (*How do you assess your ability to manage your emotion on your own without any help from psychiatrists if you develop depression?*), the degree of trust toward mental healthcare system (*Do you believe that mental health care professionals can really help people with mental health issues?*). Answers for both variables (*expected self-efficacy, trust*) can be ranged from 0 to 100. Also, self reported loss aversion is obtained by asking the participant to report emotional response toward hypothetical events of losing $10 of money or winning $10: The variable *loss aversion* is defined as expected negative emotional consequence when losing the money minus expected positive emotional consequence when obtaining the money.

Lastly, at the last page of the survey, depending on the subject's degree of willingness to know the depression test result, the subject are presented with the wanted information [18].

5.2.2 Experiment 2: Causal Investigation - Effects of message provision

The next set of experiment is designed to explore whether stigma causally induces optimistic bias in self-evaluation of depressive symptoms and avoidance of diagnostic information. The samples are recruited through Amazon Mechanical Turk again. As a design for causal investigation of the effect of stigma, I divided the samples into three groups and applied different treatment (stigma-inducing, stigma-reducing and control) in order to explore whether stigma-inducing treatment contributes optimistic bias in self-evaluation of depressive symptoms and avoidance of diagnostic information compared to other treatments.

[17] Information avoidance is defined as not wanting information that might have instrumental value even when no material cost is associated with receiving information [41]

[18] The C-ESD test used in this survey experiment is one of the commonly utilized depression screening test. However, the accurate diagnosis only can be done through in depth interview procedure by certified psychologist or psychiatrist. I clarified this to the participants when displaying the test result.

Every subject in three groups first reported socio-demographic information and two of depression related information, that is, *prior belief* and *objective depression level* which has the same meaning and measurement procedure as in experiment 1 [19].

Next, I provided different messages to subjects depending on the group they were in. The first group is stigmatized group, and was presented with three stigma-generating messages:

(1) "*People bullied as kids are less mentally healthy as adults. Kids don't easily outgrow the pain of bullying*".
(2) "*Low self-esteem can cause depression. Because those with low self-esteem are prone to replay and focus on negative thoughts far more than those who have high self-esteem, putting themselves at higher risk for low moods*".
(3) "*Being around depressed people may cause you to be depressed as well because their negative thinking style can influence your own such that over time, you too become more vulnerable to depression*".

All of messages are taken from the internet articles written by psychologist or psychiatrist. Also, the name of the authors and their affiliations were also displayed in order to enhance the priming effects of the message provision. Also, while reading messages, in order to increase the attention to the contents of the messages, participants were asked to indicate the degree of agreement to each of the message.

Similarly, to the second group (destigmatized group), I provided messages which have positive or non-stigmatizing connotations. Those are:

(1) "*Researches found that objective physical attractiveness and depression has no significant association. It's just that depressed people tend to underestimate their own attractiveness*".
(2) "*Depression can affect anyone regardless of physical or mental strength. Some of America's most well-known citizens – including Abraham Lincoln, Terry Bradshaw, and Judy Collins - have experienced depression*".
(3) ""*Creative people are more likely to have depression compared to non-creative people. There was a clear relationship between being creative and having a diagnosis of depression*".

The last group is control group. They were presented with three messages which are irrelevant of depression or mental health. Those messages are:

[19] Provision of mental health stigma-related message might affect the subjects' C-ESD test score so I choose to display the messages after finishing the test.

(1) "*Climate change is expected to have negative effects on human health although it is not clear what causes current climate change*".

(2) "*Drinking warm water is healthier than drinking cold water*".

(3) "*Fukushima nuclear disaster raised concerns as to whether eating contaminated seafood might impair human health—not just locally but across the Pacific*".

After reading messages, all the remaining procedures are the same as experiment 1. First, the participants report their self-estimation of their own depressive symptom in an unit of percentile score. Next, participants reported stigmatizing attitudes toward depression. This variable will be used to verify whether or not the randomized message provision treatments influence the participants' attitudes as intended way. After reporting stigmatizing attitude, the participants chose the amount of contents that they are willing to know: (1) both depression test result and percentile score, (2) depression test result only and (3) no information. Finally, the survey ended with displaying the diagnostic information that each participant were willing to know.

5.2.3 Experiment 3: Moderating effect of stigma on the response to partial revelation of information

The third experiment is designed to test **hypothesis 3**. This hypothesis posits that, when information is bundled with a set of partial information, even when individuals first chooses to receive the information bundle, if the first revelation of partial information induces disappointment, he opts out from receiving further information. Also, the likelihood of opting out becomes larger if he has a high degree of stigmatizing attitude toward depression. This mechanism may explain why those with stigmatizing attitudes develop optimistic bias.

This experiment was performed in the course of the first experiment. The subjects are limited to those had chosen to receive both depression test result and percentile score. The procedure is as following. First, I randomly divided the participants who had chosen to receive both information into two groups (treatment group vs. control group). The procedure applied to both group is summarized in figure **2**.

For the treatment group, at the end of the survey, they were first presented with depression test score and the criteria for depression (C-ESD score $\geq$ 16). At the moment they receive the depression test result, they might feel disappointment or relief depending on the result. The emotion of disappointment arises when the depression test score implies depression. On the other hands, the participants feel the relief if the depression test score is below the cut-off point. The emotional response is defined as $[\textit{prior belief} - I_{CESD \geq 16}]$ where the variable *prior belief* is the self-reported belief about the probability of being diagnosed with depression and $I_{CESD \geq 16}$ is the indicator function whose value is 1 if the test result implies

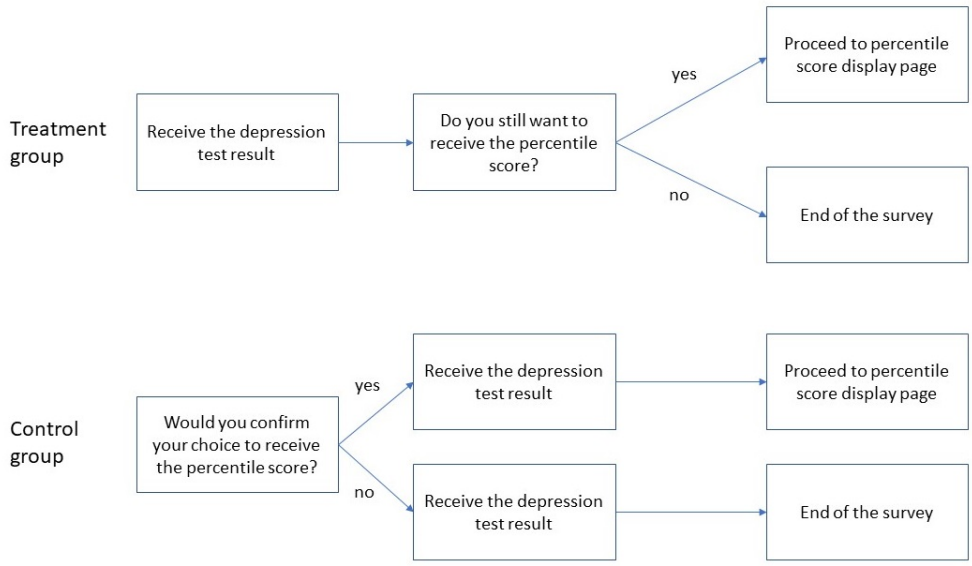


Figure 1: Procedure of Experiment 3

depression and 0 otherwise. The negative value and positive value of emotional response indicates disappointment and relief, respectively. Next, they were again asked if they still wanted to receive their percentile score. If they chose to opt out, the percentile score was not displayed. Whereas, for the control group, before being presented with depression test result, they were first asked whether to confirm to receive both test result and percentile score or receive the depression test only.

6 Result

Summary statistics for the two experiments are available in **Appendix**.

6.1 Cross-sectional Observation

6.1.1 underestimation bias

The results of experiment 1 which is based on cross-sectional design confirms the theory. First thing to note is that subjects' self-estimation of their percentile depression score among all the participants is generally accurate. Figure 3 represents the relationship between true percentile score and self-estimated percentile score, where those two variables follow linearly increasing relationship. Such linear trend implies that participants are generally aware of their depression state compared to others. However, in the graph, there is also individual variation in the accuracy of self-estimation. Also, the self-estimation biases are observed to be skewed toward negative values, that is, the majority of samples (63%) show

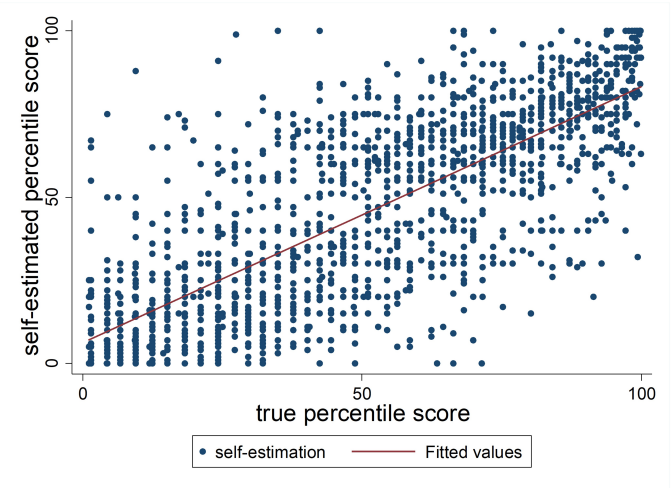
underestimation bias in evaluation of their degree of depression compared to others.

The equation (5) poses the possibility that such underestimation of symptom degree might be motivated. The theory expects the degree of underestimation bias would be greater for those have the high level of stigmatizing attitudes or the low level of subjective expected loss of health when developing depression. I measured the stigmatizing attitudes of each participants (*stigma*). Also I measured the subjective expected ability to cope with depression (*self-efficacy*) which I will use as a proxy for expected loss of health when developing depression, that is, the high value of *self-efficacy* implies low value of expected health loss if depression develops. The equation (5) offers expectation that high value of *stima* or *self-efficacy* result in underestimation bias. Consistent with this expectation, the participants who have high value of *stigma* or high value of *self-efficacy* are observed to underestimate their percentile depression ranking. This observation is represented in the figure 5 which compares the marginal mean values of the bias (with CES-D score as the control factor) in self-estimation of one's own percentile depression ranking between lower and upper half in terms of the value of stigma (panel a), also between lower and upper half in terms of the value of self-efficacy (panel b) [20]. The high stigma group estimates their percentile depression ranking about 2 rank healthier than low stigma group (-6.46 vs. -4.19). Also, the comparison between high and low self-efficacy group shows the similar pattern (-6.46 vs. -4.22).

To begin the econometric analysis, I regress the subjects' self-estimated percentile ranking of their own CES-D test score on the true ranking, stigma, self-efficacy and socio-demographic information as control variables using OLS. The result of analysis is presented in the column (1) to (3) in Table 1. Consistent with this theoretical expectation and graphical observation in figure 5, the degree of underestimation bias is higher among those have higher value of the variable *stigma* and *self-efficacy*. First, in column (1), the self-estimated percentile score is highly associated with true percentile score which implies that individuals are generally accurately aware of their mental health state. This high degree of association is not attenuated when adding psychological variables as in column (2), or adding socio-demographic control variables as in column (3). Also, the variable stigma has negative association with self-estimated percentile depression ranking. The standard deviation unit increase in stigma reduces the self-estimated depression score around 1 percentile rank lower and this pattern is observed both when socio-demographic variables are controlled for(column 3) or not (column 2). Similarly the negative association between self-estimated depression score and self-efficacy. Lastly, the column (3) displays that the socio-demographic variables are not associated with self-estimation of depression ranking except for being non-

[20]Marginal mean bias of each category (e.g., high stigma group) is the mean value of the bias averaged across all level of CES-D test score, which prevents the over-representation of samples within specific range of the CES-D score when computing the mean value.

Figure 2: Self-estimation of percentile depression score ranking



Note. The higher value of percentile score implies more severe depressive symptom

Figure 3: Self-estimation Bias

(a) Stigma vs. Bias

(b) Self-efficacy vs. Bias

Caucasian. Non-Caucasian subjects tend to estimate their depression percentiles score 3 rank lower than Caucasian subjects. This result is consistent with observed high level of self-stigma among ethnic minorities in the US [40, 72].

6.1.2 Avoidance of Diagnostic Information

In this section, I report the observed relationship between avoidance of mental health diagnostic information and relevant variables including stigma and self-estimation bias. In sum, participants are less likely to seek information if they have higher degree of stigma and/or underestimation bias with respect to their depression percentile ranking. Especially, the information avoidance is more prominent among those who display high value of stigma and underestimation bias simultaneously. Even though the results do not establish the causal relationship due to the cross-sectional design of the study, they are consistent with the theoretical expectation which is posed in equation (8).

The options that participants can choose are (1) "*I want to know both my depression test result and my depression percentile score among all the participants.*", (2) "*I want to know my depression test result only*" and (3) "*I do not want to any information on my depression state*". Among the whole sample ($N = 1372$), a majority (60%) chooses to receive both information (option 1). The portion of samples who choose option (2) and (3) is 32% and 8%, respectively.

First, the theory (equation 8) expects that individuals with more severe depressive symptoms derive more instrumental value from knowing one's true mental health state so they are likely to choose the information with more detail. The graph (a) in figure 6 shows that the observation from the data is consistent with this theoretical expectation. On this graph, 'nonavoider' refers to the participants who choose to receive both the test result and percentile ranking (option 1), while 'avoider' denotes the group of samples who choose option (2) or (3). It is apparent in the graph that the group of nonavoiders have the distribution of the variable *severity* (true percentile depression score) more skewed to right than avoiders, which implies there is a positive association between information seeking and symptom severity.

The key component of the theory is the negative effect of stigmatizing attitude toward depression on willingness to know one's diagnosis result. The graph (b) of figure 6 shows the sample distribution of variable *stigma* among both the group of avoiders and nonavoiders. Consistent with the theory, the sample distribution of the variable *stigma* is more skewed to the right among avoiders. Also, the equation (8) offers prediction that the deterring effect of stigma would be more prominent when the individual is displaying under-biased self-estimation of symptom severity. The figure 7 depicts the observed relationship between the willingness to know one's depression test result and the interaction variable of *high or*

low stigma (lower half or upper half) and *underestimation or overestimation*. The y-axis represents the average value of willingness to know among each group of participants [21]. As predicted by model, information avoidance behavior is observed among (and only among) the participants who both having high degree of stigmatizing attitude and underestimated self-evaluation of symptom.

For econometric analysis, I regress the subjects' willingness to know their own depression test result on depression severity (percentile score), variables about stigma and mental health related attitude and socio-demographic control variables. The dependent variable *willingness to know* is defined same as in footnote 22. Considering the ordinal nature of *willingness to know* variable, I apply the ordered logit regression method. The column (1) on the table 2 shows simple regression result where I use three key variables *true percentile score*, *stigma* and the degree of *optimistic self-evaluation bias* described in the previous section as independent variables. Those three variables are strongly associated with *willingness to know*. The more severe depressive symptom has positive association with demand for mental health information, whereas *stigma* and *optimistic bias* have negative association with the demand for information. On the column (2), instead of assuming the coefficient of *optimistic bias* is universal among all the participants, I perform sub-group analysis regarding the effect of optimistic bias among high stigma group and low stigma group. Consistent with the theory and the crude observation demonstrated in the figure 7, the participants who have both high level of stigma and optimistic bias are likely to avoid detailed information.

Columns (3)-(5) describe the regression results when other mental health related variables are additionally included as dependent variables. First, the column (3) shows that the variable perceived *self-efficacy* has negative association with willingness to receive information. There are two potential explanations of this negative association. First, as believing the cost of developing depression is low because of high degree of perceived self-efficacy, one may derive little instrumental value of knowing current mental health state. Second explanation is that high value of self-reported self-efficacy might be motivated from stigmatizing attitude, thus the variable *self-efficacy* might reflect an aspect of stigma. This second explanation is supported by the decreased coefficient of *stigma* in column (3) compared to column (2) which implies the positive correlation between *stigma* and *self-efficacy* [22].

In column (4), I add other variables *intention to utilize mental healthcare*, *intention to utilize physical healthcare*, *perceived discrimination toward depressed people* and *previous mental health service use*. First, the variable *intention to utilize mental healthcare* is positively associated with *willingness to receive information*. This observed association might be explained by the plausible presumption that one aspect of visiting mental health ser-

[21] I assigned the value 2 to the variable *willingness to know* if a participant chooses to receive both information (both depression test result and percentile score), value 1 if he/she chooses to receive test result only and value 0 if refuses to receive any information.

[22] the correlation between *stigma* and *self-efficacy* is 0.146

vices is the possibility of receiving unpleasant diagnostic information. That is, individuals who are vulnerable to anxiety about receiving disappointing diagnostic result would avoid receiving information in this experiment and also visiting mental health clinic. However, the column (5) shows that when including the variables *trust toward mental healthcare* and *trust toward depression screening test* used in this survey experiment in the regression, the positive association between *intention to utilize mental healthcare* and *willingness to receive information* is reduced and no longer statistically significant, which implies positive correlation between trust toward mental healthcare system (including treatment instrument such as screening test) and intention to utilize mental healthcare. Also, those two attitude variables have strongly positive association with *willingness to receive information*. Taken together, this observation implies positive attitude toward formal mental healthcare system may act as encouraging factor for individuals to receive mental health information in this survey experiment, as well as encouraging them to visit mental healthcare professionals if developing depression.

Next, interestingly, participants who think depressed people experience discrimination from general public (*perceived discrimination toward depressed people*) seek more detailed information. This seemingly odd observation might be because of the Internet-based thus anonymous nature of the survey experiment. Namely, individuals with high level of stigma may consider internet-based informal depression test as a substitute for actual visit to formal mental health services which can generate shame among people with stigmatizing attitude. This argument is supported by the negative association between *perceived discrimination toward depressed people* and *intention to utilize mental healthcare* on the column (7). That is, individuals who consider visiting mental health professionals as potentially causing discriminating response from others avoid the visit in person however seek information from the anonymity guaranteed source.

6.2 Causal investigation: Effect of priming

In this section, I describe the findings from the effect of priming experiment. Even though the results describe in the previous section offer supporting evidence of the theory, they have limitation in establishing the causality because the experimental design of the first experiment is cross-sectional. The priming experiment aims to investigate causal relationship about the patterns observed in cross-sectional data. Mainly, I test two hypotheses in this section: (1) stigmatizing attitude toward depression causes themselves to develop underestimation of symptom bias, (2) stigmatizing attitude toward depression causes themselves to avoid diagnostic mental health information.

To explore the causal relationship, I use priming approach. That is, after randomly assigning participants into two groups, I show different messages to each group (stigma-

reducing vs. stigma-inducing message). If the observed patterns (positive association between stigma and underestimation of symptom, negative association between stigma and willingness to receive information) are causal, it is expected that participants who read stigma-inducing messages would display more underestimation of symptom bias, also less willingness to receive information compared to participants who are provided with stigma-reducing messages.

To put the conclusion first, there are certain group of participants who display the differential behavioral patterns depending on the type of messages provided as expected. They are those who are married also employed as a regular worker. However, the expected results are observed only among this group of participant. The observed difference of self-estimation and willingness to know by receiving stigma-inducing and stigma-reducing messages among those married and employed as a regular worker are displayed on the the figure 8. First, among this group of participants, those read stigma-inducing messages make self-evaluation of their percentile depression score 8.11 point less than those read stigma-reducing messages. Also, those read stigma-reducing messages show slightly higher degree (0.11) of willingness to receive information. Interestingly, when excluding samples who answer themselves to have 0% chance of being diagnosed with depression at the first part of the survey, the effects of different message provision become more prominent. Figure 9 describes the behavioral difference of the same group of participant (married & employed as regular worker) but after excluding those deny the possibility of having depression. On panel (a) and (b) on figure 9 shows that the difference of self-evaluation bias and willingness to know between two groups become larger after excluding the denial samples, with the change from 8.11 to 10.15 for self-evaluation bias, from 0.11 to 0.19 for willingness to know, respectively. This implies denial of possibility of having depression has negative association with the likelihood that the effects of different message provision arise.

The regression results are consistent with these observations. Table 3 shows that there are no difference in the behavioral patterns between the participants who read stigma-reducing messages and stigma-inducing messages when using whole samples. However, the panel (1) and (2) on the table 4 shows that when the analysis is limited to those married and employed, the participants who receive stigma-inducing messages significantly underestimate their percentile depression score compared to those receive stigma-reducing messages. Moreover, the effect of differential messages provision becomes more prominent when excluding samples who deny the possibility of having depression (panel (3) and (4) in the table 4). As for willingness to receive information, on panel (2) in the table 3, I could not find any difference between the stigma-induced group and stigma-reduced group. However, when I limit the analysis to the married and employed, as in panel (1) and (2) in the table 5, stigma induced group displays less demand for information even though the difference is

not statistically significant. Panel (3) and (4) in the table 5 shows the regression result when the samples showing denial are excluded from the samples in the panel (1) and (2). After excluding those samples from the analysis, the effect of stigma-inducing messages on information demand starts to be observed.

6.3 mechanism for developing biased belief: partial revelation of information

In this section, I present the experimental evidence of the theoretical expectation summarized in the equation (10). The expectation about the choice of subjects in treatment group based on my theory is as follows: (1) they are likely to proceed to receive the percentile score if the CES-D test result indicates depression and (2) they are likely to reverse the first decision and not to receive the percentile score if the CES-D test result generates disappointment. The counterpart expectation about those who receive relieving test result (not depressed) is that one might be more likely to proceed to receive the percentile score if the degree of relief is greater. The econometric specification directly follows equation (10) with some modification. A subject's decision whether or not to continue to receive the percentile CES-D score ($Stay = 1$: Receive it, $Stay = 0$: Not receive it) is described with the below equation.

$$Stay = \begin{cases} 1 & \text{if} \quad Stay' > 0 \\ 0 & \text{if} \quad Stay' \leq 0 \end{cases}$$

where

$$Stay' = \theta_1 \sigma_A + \theta_2(\sigma_A - q_0) + \theta_3 + \epsilon \tag{11}$$

Here, σ_A is the CES-D test result with the value of 0 and 1 representing 'not depressed' and 'depressed', respectively. Also, q_0 is the subject's prior belief about the probability of being diagnosed with depression. The disappointment happens when $\sigma_A = 1$ and the degree of disappointment is $1 - q_0$, whereas relief is experienced when $\sigma_A = 0$ and the degree of relief is q_0. According to the theory, the coefficient θ_1 and θ_2 is expected to be positive and negative respectively.

To investigate if the response to the realized signal of CES-D test result is different depending on whether the subject experienced disappointment or relief, and the degree of stigmatizing attitude of the subject is high or low, I perform the sub-group regression with respect to the coefficient θ_2 by subjects with the experience of disappointment and relief, and by subjects with high level (the value of variable *stig* is greater than or equal to 3) and low (the value of variable *stig* is less than 3) level of stigmatizing attitude. In this case, the

equation (11) is re-written as

$$Stay' = \theta_1 \sigma_A + \sum_{gr} \theta_{2,gr} I_{gr}(\sigma_A - q_0) + \theta_3 + \epsilon \tag{12}$$

where the indicator gr refers to the group the subject belongs to, which is one of among ld (low stigma/disappointment), lr (low stigma/relief), hd (high stigma/disappointment) and hr (high stigma/relief). The expectation about the value of the coefficient $\theta_{2,gr}$ based on the theory is as follows.

1) $\theta_{2,gr} < 0$ for all gr

Experiencing disappointment/relief at the first signal leads subjects to avoid/seek additional information.

2) $|\theta_{2,hd}| > |\theta_{2,ld}|$, and $|\theta_{2,hr}| > |\theta_{2,lr}|$

Responsiveness to the first signal with respect to the demand for second signal is greater among the subjects with higher degree of stigmatizing attitude.

To estimate the model, I use the logit regression. The result is on the table 6. The value of dependent variable is 1 if the subject proceeds to receive the percentile score and 0 otherwise. The standard errors of estimated coefficients are calculated by bootstrap resampling because the sample size in each of the four categories (disappointment/relief $\times$ high/low stigma) is low and asymptotic normality assumption of estimated coefficient may not be a good approximation. Also, even though the signs of the coefficient of both disappointment and relief are expected to be negative according to the theory (equation (12)), I change the sign of the estimated coefficient of the variable relief to positive because it is intuitively more straightforward to interpret.

The columns from (1) to (3) report the estimation results among those in the control group. The results support my theory. First, the first row shows that receiving the first set of information that implies depression leads subject to have stronger willingness to proceed to receive the remaining set of information, while subjects in control group do not show the same pattern as on column (4) to (6) ($\theta_1 > 0$ in equation (12)).

Second, the column (2) shows that if the first set of information generates disappoint or relief, the subjects tend to show decreased or increased willingness to receive the further information, respectively ($\theta_2 < 0$ in equation (12)). Also, the coefficient of 'first signal' on the column (2) is 3 times as high as the coefficient of same variable on the column (1), which implies the omission of disappointment/relief variable in regression results in significant underestimation of the positive impact of receiving a test result implying depression on the willingness to receive additional information. This happens because, according to the theory,

although the pure impact (when the prior belief about the probability of being diagnosed with depression holds constant) of receiving first signal implying depression can be high, this impact might be mitigated or even canceled out by the motivation of avoidance of further disappointment when receiving test result *depression* is unexpected so felt unpleasant surprise. Similarly, among those receive the test result *not depression*, those highly expected to be diagnosed with depression so felt pleasant surprise, or relief, are more likely to proceed to receive additional information.

Third, the column (3) reports how the highly ($stigm > 3$) and mildly ($stigm \leq 3$) stigmatic subjects differently respond to the contents of the first set of information. As described on the table the deterring effect of disappointment from the first signal is twice as great among those with high level of stigma as those with low level of stigma. Similarly, even though the difference is not high, the encouraging effect of experiencing relief is greater among high stigma group compared to low stigma group.

7 discussion and conclusion

to be added

8 Appendix

8.1 Review of theories on mental health stigma and help-seeking

8.1.1 stages of problem recognition and information

According to theories of help-seeking pathways for mental disorders [3, 20, 32], help-seeking can be divided into three identifiable stages which are problem recognition, the decision to seek help, and the selection of a help provider. Among them understanding the problem recognition stage is one of the objects of this study. In order for the decision to receive a treatment to occur, the individual first needs to recognize one's emotional disturbance to be attributable to some kind of disease rather than temporary mood swings. This phase is problem recognition stage in help-seeking theories, and it includes the sub-stages of self-recognition of symptom, self-identification of the type of disease and perception of need for treatment [20].

The proposed factors that may influence problem recognition and perception of need for treatment include the characteristics of the disorder such as the severity of symptoms, socio-demographic factors such as age, gender, race/ethnicity and attitudinal factors such as trust toward health care system, beliefs about the causes of mental illness [68].

Among the studies on the factors influencing problem recognition and help-seeking, the work by Cause et al. [20] is particularly of interest in that they propose the importance of

social network that the patient is in. Specifically, they suggest that problem recognition for a specific type of mental disorder may not easily occur in a society that accepting of that psychiatric symptoms because the psychiatric state might be interpreted as normal in that social network [59]. However, there is a complication when it comes to consideration of the relationship between social norm and problem recognition. This is because if social norm defines having mental disorder as inferior personality trait such as lack of willpower then individuals might want to hide their symptom in public and it may interfere with help-seeking intention. Therefore, to explore the effect of social network on problem recognition and help-seeking on a certain mental disorder, it is required to consider not only the prevalence of such symptom in the society but also the public attitudes on the disease. Those public attitudes are closely related with concept of stigma, which will be reviewed in the next section.

8.1.2 stigma: concept and its implication on help-seeking

This study aims to explore the relationship between stigmatizing attitude toward those with depression and two types of mental health behaviors, that is, under-assessment of symptoms and avoidance of self-relevant mental health information. In this section, the relevant concepts, stigma, and its implications from literature are reviewed.

A growing body of literature has studied the concept and health consequence of stigma which is mostly from the area of clinical and social psychology [21, 23, 26]. Mental health stigma is a negative stereotype attached to those with mental health issue, whether it is true or not, such as the statements '*mental health conditions are caused by lack of willpower*' or '*depressed people are socially unpleasant*'. According to Corrigan (2004) [23], mental health stigma is manifested twofold. The first type of stigma is public-stigma which refers to negative stereotypes or prejudices that naive public has towards those with mental health issue. The second type is self-stigma which means the self-blaming or self-discriminating attitudes about one's own mental health condition such as the statements '*I am incompetent*' or '*people will not like me*'. Self-stigma often leads to negative emotional consequences including shame, diminished self-esteem or loss of hopefulness. Also, those emotional consequences of self-stigma further impair the social or vocational abilities of oneself by resulting in self-isolation from the social network and loss of motivation to pursue a job or education [25].

Between two types of stigma, self-stigma is the focus of this study. The self-stigmatizing process follow a complicated socio-psychological mechanism. In short, in the course of self-stigmatizing process, an individual first internalizes the public stigma before becoming a patient (internalized stigma) and applies such negative views towards oneself when developing psychiatric symptoms (self-stigma) [53]. This framework implies that those who have highly internalized public stigma into their own attitudes are more likely to experience self-stigma when they develop mental health problem [25].

Stigma is picked as one of the crucial barriers for mental health care utilization. Specifically, desire to avoid being a target of public or self stigmatization is considered as a potential cause for reluctance or delay in help-seeking [21, 40, 62]. In a meta analysis by Clement et al. (2015) [21], several sub-themes of mental health stigma are emerged to have negative association with help-seeking, which include (1) desire to keep one's mental health state secret, and (2) desire to avoid dissonance between one's preferred self-identity and reality. The first theme is interpreted as fear of anticipated discrimination by others such as employer or peer group, and might be termed as public-stigma avoidance intention. Whereas, the second theme implies individuals avoid mental health care service not only from the desire of non-disclosure but also from the desire to avoid internal shame which is purely psychological motives, which can be conceptualized as self-stigma avoidance intention [69]. These meta analysis results are consistent with the theoretical argument of Corrigan (2004) [23], which hypothesizes the two pathways how stigma can deter help-seeking: (1) the desire to avoid being labeled as a mental patient by others (public-stigma avoidance), and (2) the desire to avoid experiencing shame and embarrassment (self-stigma avoidance).

8.1.3 self-stigma: effects on problem recognition and attention to mental health information

Between public-stigma and self-stigma, the latter is of focus in current study. As mentioned earlier, the precondition for help-seeking to occur is recognition of symptoms, self-identification of specific disease, and this stage involves gathering relevant information including symptomatology of the disease and informal self-administered diagnosis result. In this section, I review two potential mechanisms in which self-stigma might interfere with successful self-appraisal, which are the effects of self stigma on (1) denial or self-labeling avoidance, and (2) motivated inattention to mental health information.

First, out of the fear of experiencing self-stigma which would be experienced as a form of shame or embarrassment, individuals might be reluctant to attribute the current emotional disturbances to mental illness [12, 57, 65]. For example, newly-admitted psychiatric patients are observed to display denial behavior, that is, they are likely to refuse to characterize themselves as having a negative mental health condition [57]. Moreover, according to a study with subjects of those potentially depressed but untreated, there is an association between internalized stigma (support for public's stigmatizing attitude) and reduced likelihood of self-identification as having depressive symptoms [65]. This association may imply the negative emotional consequence resulting from admitting depression to oneself would be greater for those have more discriminating view against those with depression. Thus, this process of self-identification failure might not happen just because the subjects are irrational or lack sufficient knowledge. Indeed, researchers have hypothesized that 'ratio-

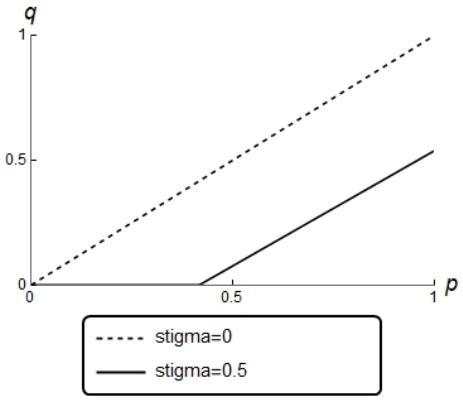
nal' rejection of unpreferable-identity might explain denial of identifying oneself as a mental health patient [57].

The second potential mechanism of how self-stigma avoidance motive may prevent potential patients from recognizing their mental health symptoms is its negative effects on the willingness to attend to relevant mental health information such as treatment options and causes, symptoms of specific mental illness or even information on one's own experiences which may imply one's mental health state and be gathered from recollection of one's daily life. As mentioned earlier, the problem-recognition stage includes self-recognition of symptom, self-identification of the type of disease and perception of need for treatment, and each of them involves gathering relevant information. In order for the treatment decision to be made, individuals first need to know whether their current emotional disturbance fits into the formal definition of mental disorder and whether it requires help from professionals [52]. For example, a majority of those eventually make a treatment decision are those who actively have sought out information [74]. For example, university students who highly internalized stigmatizing attitudes toward those with depression are less likely to seek counseling information even though the act of seeking information does not require them to see a mental health professional [52]. Also, it is observed that participants of a study are less likely to attend to health-promotion information regarding stigmatized health conditions (e.g., genital herpes or HIV) than stigma-neutral information (e.g., flu) [30]. To explain the association between stigma and information avoidance, those studies suggest mechanisms according to which individuals with high level of internalized stigma are likely to perceive the information to be a threat in maintaining positive self-image (individual level effect), or information on stigmatized diseases could be deemed more fearful to attend do (disease-specific effect). However, those explanations require more theoretical basis regarding why personal level stigma or stigma attached to specific disease are positively associated with fear of being informed even though the information itself does not force individuals to take further action.

Figure 4: psychological factors for risk underestimation

(a) Stigma

($\gamma = 1$, $W = 1.2$)



(b) Self-assessed Vulnerability to Depression

($\gamma = 1$, $Stigma = 0.3$)

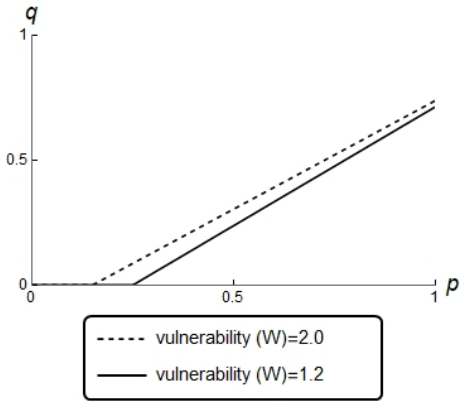
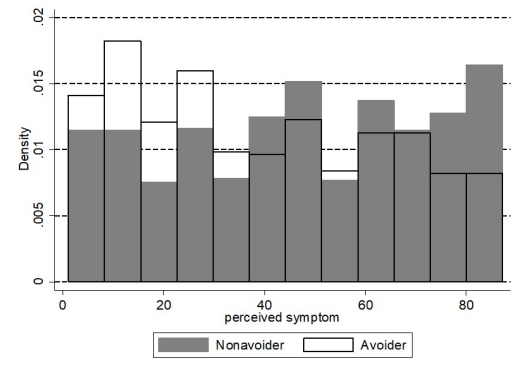


Figure 5: Willingness to know the depression test result

(a) Depression severity (percentile score)



(b) Stigma

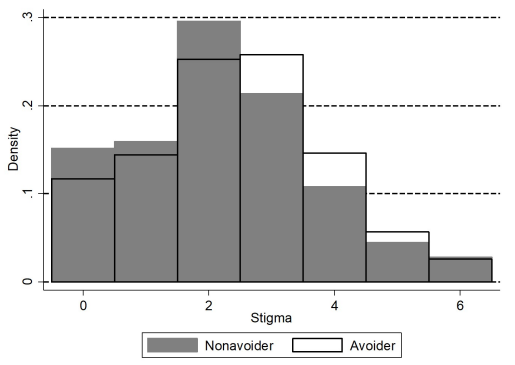
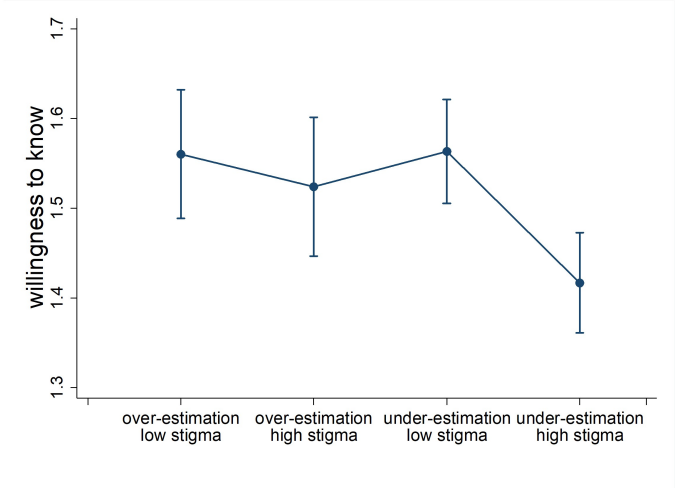


Figure 6: [Willingness to know] vs. [$stigma \times underestimation$]



Note. The higher value of percentile score implies more severe depressive symptom

Figure 7: Effects of the type of message (married & employed)

(a) Self-evaluation Bias

(b) Willingness to receive Information

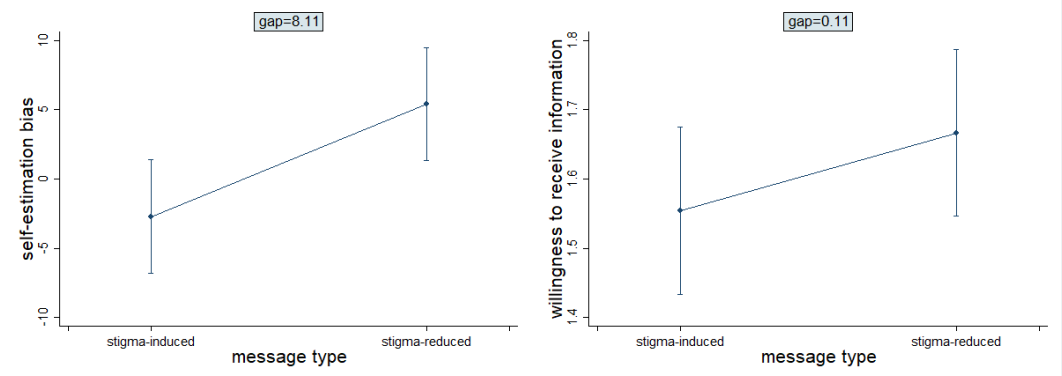


Figure 8: Effects of the type of message (married & employed & **non-denial**)

(a) Self-evaluation Bias

(b) Willingness to receive Information

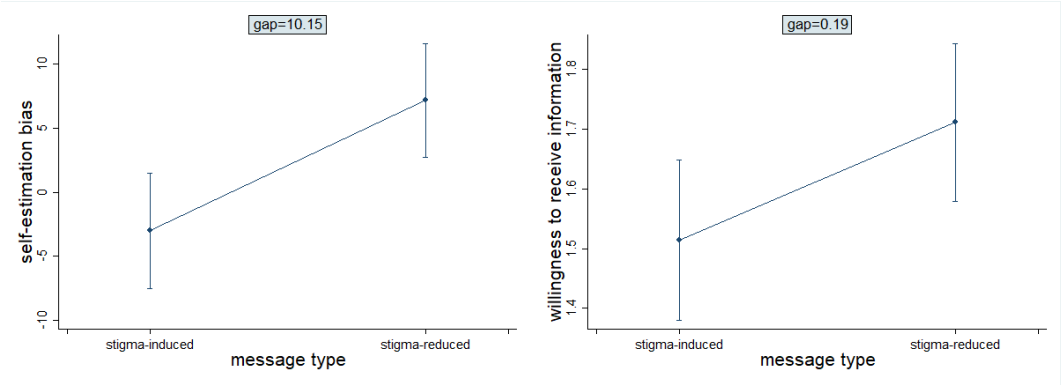


Table 1: Underestimation of depression percentile score

	percentile score: *self-estimation*		
	(1)	(2)	(3)
percentile score: *true score*	0.792***	0.794***	0.801***
	(0.018)	(0.019)	(0.020)
stigma		-1.081**	-0.986**
		(0.487)	(0.496)
self-efficacy		-0.861*	-0.834*
		(0.492)	(0.495)
age			0.065
			(0.042)
male			0.962
			(1.048)
married			-0.141
			(1.063)
non-caucasian			-3.094***
			(1.138)
years of education			-0.187
			(0.242)
income ($1,000)			0.006
			(0.013)
employed			-0.021
			(1.042)
Observations	1372	1372	1372

Notes. In parentheses are robust standard errors of OLS regression. The dependent variable is self-estimated percentile ranking of one's own CES-D test score. The higher value represents more severe degree of depressive symptom. Two independent variables *stigma* and *self-efficacy* are standardized. ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.)

Table 2: Information avoidance

	Willingness to Receive information						intention to utilize mental healthcare
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
percentile score: *true score*	0.015***	0.015***	0.013***	0.013***	0.014***	0.014***	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
stigma	-0.144***	-0.134**	-0.112**	-0.187***	-0.124*	-0.113*	0.133**
	(0.055)	(0.057)	(0.057)	(0.062)	(0.063)	(0.065)	(0.058)
optimistic bias	-0.011***						-0.003
	(0.003)						(0.003)
optimistic bias		-0.014***	-0.014***	-0.014***	-0.012***	-0.012***	
× high stigma		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	
optimistic bias		-0.009**	-0.009**	-0.009**	-0.008**	-0.009**	
× low stigma		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	
self-efficacy			-0.159***	-0.137**	-0.139**	-0.145**	-0.213***
			(0.058)	(0.060)	(0.061)	(0.061)	(0.054)
trust toward					0.282***	0.293***	0.268***
mental health services					(0.060)	(0.060)	(0.059)
trust toward					0.174***	0.182***	0.199***
depression screening test					(0.059)	(0.059)	(0.056)
intention to utilize				0.169**	0.095	0.061	
mental healthcare				(0.070)	(0.072)	(0.072)	.
intention to utilize				0.022	0.022	0.025	1.410***
physical healthcare				(0.069)	(0.069)	(0.070)	(0.067)
perceived discrimination				0.222***	0.198***	0.193***	-0.233***
toward depressed people				(0.060)	(0.060)	(0.061)	(0.055)
previous mental				0.004	-0.006	0.004	0.135***
health services use				(0.048)	(0.049)	(0.050)	(0.046)
age						-0.000	-0.010**
						(0.005)	(0.005)
male						0.018	-0.024
						(0.123)	(0.113)
married						-0.305**	0.003
						(0.124)	(0.114)
non-caucasian						-0.052	0.408***
						(0.135)	(0.125)
years of education						0.014	-0.007
						(0.028)	(0.027)
income ($1,000)						0.002	-0.004**
						(0.002)	(0.001)
employed						-0.343***	0.037
						(0.123)	(0.113)
Observations	1372	1372	1372	1372	1372	1372	1372

Notes. In parentheses are robust standard errors of ordered logit regression. The dependent variable is willingness to receive the diagnostic information. The independent variables *stigma*, *self-efficacy*, *trust toward mental health service*, *trust toward depression screening test*, *intention to utilize mental healthcare*, *intention to utilize physical healthcare* and *perceived discrimination toward depressed people* are standardized.
($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.)

Table 3: Effect of message provision on willingness to receive information
(whole samples)

	(1)	(2)
	self-evaluation bias (OLS)	willingness to receive information (Ordered logit)
stigma-inducing:	0.109	0.083
(compared to stigma-reducing)	(1.406)	(0.163)
percentile score:	-0.249***	0.012***
true score	(0.031)	(0.004)
Observations	703	703

Notes. In parentheses are robust standard errors of each regression model.
($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.)

Table 4: Effect of message provision on self-evaluation bias
(samples: married & employed as regular worker)
model: OLS

	self-evaluation bias		self-evaluation bias (excluding denial samples)	
	(1)	(2)	(3)	(4)
stigma-inducing:	-7.022**	-7.198**	-8.792***	-8.601***
(compared to stigma-reducing)	(2.852)	(2.883)	(3.148)	(3.199)
percentile score:	-0.220***	-0.218***	-0.226***	-0.214***
true score	(0.064)	(0.065)	(0.074)	(0.076)
age		0.057		0.089
		(0.138)		(0.148)
male		0.239		-1.192
		(2.997)		(3.281)
non-caucasian		-4.855		-0.084
		(3.813)		(4.496)
years of education		-0.280		-0.739
		(0.701)		(0.804)
income ($1,000)		0.004		0.026
		(0.040)		(0.045)
Observations	185	185	150	150

Notes. In parentheses are robust standard errors of OLS regression. The dependent variable self-evaluation bias is self-estimated percentile score minus true depression percentile score. The higher value of the bias represents more optimistic evaluation.
($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.)

Table 5: Effect of message provision on willingness to receive information
(samples: married & employed as regular worker)
model: Ordered Logit

	willingness to receive information		willingness to receive information (excluding denial samples)	
	(1)	(2)	(3)	(4)
stigma-inducing:	-0.412	-0.509	-0.703**	-0.844**
(compared to stigma-reducing)	(0.312)	(0.326)	(0.352)	(0.371)
percentile score:	0.001	0.003	0.004	0.007
true score	(0.007)	(0.007)	(0.008)	(0.009)
age		-0.002		-0.002
		(0.015)		(0.016)
male		-0.605*		-0.518
		(0.328)		(0.366)
non-caucasian		-1.062***		-1.216**
		(0.406)		(0.475)
years of education		0.078		0.087
		(0.079)		(0.091)
income ($1,000)		0.010		0.011*
		(0.005)		(0.006)
Observations	185	185	150	150

Notes. In parentheses are robust standard errors of ordered logit regression. ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.)

Table 6: Effect of Receiving CES-D test result on the willingness to receive the CES-D percentile score
(model: Logit)

	receive the percentile score					
	treatment group ($N = 334$)			control group ($N = 1566$)		
	(1)	(2)	(3)	(4)	(5)	(6)
first signal:	0.750*	2.310***	2.311***	-0.185	-0.211	-0.198
depression	(0.393)	(0.790)	(0.691)	(0.613)	(0.606)	(0.774)
disappointment		-2.314**			-0.037	
($N = 193$)		(0.988)			(0.988)	
relief		5.086*			0.239	
($N = 141$)		(2.996)			(2.678)	
disappointment			-1.709			0.038
low stigma			(1.396)			(1.335)
disappointment			-3.730**			-0.315
high stigma			(1.639)			(1.251)
relief			4.873			-0.366
low stigma			(4.448)			(4.277)
relief			5.907**			2.227
high stigma			(2.829)			(1.727)

Notes. In parentheses standard errors computed by bootstrap. ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.)

References

[1] Akerlof, G. A., and Dickens, W. T. The economic consequences of cognitive dissonance. The American economic review 72, 3 (1982), 307–319.

[2] Akerlof, G. A., and Kranton, R. E. Economics and identity. The Quarterly Journal of Economics 115, 3 (2000), 715–753.

[3] Andersen, R. M. Revisiting the behavioral model and access to medical care: does it matter? Journal of health and social behavior (1995), 1–10.

[4] Augsberger, A., Yeung, A., Dougher, M., and Hahm, H. C. Factors influencing the underutilization of mental health services among asian american women with a history of depression and suicide. BMC health services research 15, 1 (2015), 542.

[5] Baumann, A. E. Stigmatization, social distance and exclusion because of mental illness: the individual with mental illness as a 'stranger'. International review of psychiatry 19, 2 (2007), 131–135.

[6] Bell, D. E. Disappointment in decision making under uncertainty. Operations research 33, 1 (1985), 1–27.

[7] Bénabou, R. Groupthink: Collective delusions in organizations and markets. Review of Economic Studies 80, 2 (2012), 429–462.

[8] Bénabou, R., and Tirole, J. Self-confidence and personal motivation. The Quarterly Journal of Economics 117, 3 (2002), 871–915.

[9] Bénabou, R., and Tirole, J. Identity, morals, and taboos: Beliefs as assets. The Quarterly Journal of Economics 126, 2 (2011), 805–855.

[10] Bénabou, R., and Tirole, J. Mindful economics: The production, consumption, and value of beliefs. Journal of Economic Perspectives 30, 3 (2016), 141–64.

[11] Bernheim, B. D., and Thomadsen, R. Memory and anticipation. The Economic Journal 115, 503 (2005), 271–304.

[12] Bharadwaj, P., Pai, M. M., and Suziedelyte, A. Mental health stigma. Economics Letters 159 (2017), 57–60.

[13] Brocas, I., and Carrillo, J. D. Dual-process theories of decision-making: A selective survey. Journal of economic psychology 41 (2014), 45–54.

[14] Brown, J. D. Understanding the better than average effect: Motives (still) matter. Personality and Social Psychology Bulletin 38, 2 (2012), 209–219.

[15] Brunnermeier, M. K., and Parker, J. A. Optimal expectations. American Economic Review 95, 4 (2005), 1092–1118.

[16] Camerer, C., and Lovallo, D. Overconfidence and excess entry: An experimental approach. American economic review 89, 1 (1999), 306–318.

[17] Caplin, A., and Leahy, J. Psychological expected utility theory and anticipatory feelings. The Quarterly Journal of Economics 116, 1 (2001), 55–79.

[18] Caplin, A., and Leahy, J. V. Wishful thinking. NBER Working paper (2019).

[19] Carrillo, J. D., and Mariotti, T. Strategic ignorance as a self-disciplining device. The Review of Economic Studies 67, 3 (2000), 529–544.

[20] Cauce, A. M., Domenech-Rodríguez, M., Paradise, M., Cochran, B. N., Shea, J. M., Srebnik, D., and Baydar, N. Cultural and contextual influences in mental health help seeking: a focus on ethnic minority youth. Journal of consulting and clinical psychology 70, 1 (2002), 44.

[21] Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rüsch, N., Brown, J., and Thornicroft, G. What is the impact of mental health-related stigma on help-seeking? a systematic review of quantitative and qualitative studies. Psychological medicine 45, 1 (2015), 11–27.

[22] Compte, O., and Postlewaite, A. Confidence-enhanced performance. American Economic Review 94, 5 (2004), 1536–1557.

[23] Corrigan, P. How stigma interferes with mental health care. American psychologist 59, 7 (2004), 614.

[24] Corrigan, P. W., Edwards, A. B., Green, A., Diwan, S. L., and Penn, D. L. Prejudice, social distance, and familiarity with mental illness. Schizophrenia bulletin 27, 2 (2001), 219–225.

[25] Corrigan, P. W., and Rao, D. On the self-stigma of mental illness: Stages, disclosure, and strategies for change. The Canadian Journal of Psychiatry 57, 8 (2012), 464–469.

[26] Corrigan, P. W., and Watson, A. C. Understanding the impact of stigma on people with mental illness. World psychiatry 1, 1 (2002), 16.

[27] Coutts, A. Testing models of belief bias: An experiment. Games and Economic Behavior 113 (2019), 549–565.

[28] Crabb, R., and Hunsley, J. Utilization of mental health care services among older adults with depression. Journal of clinical psychology 62, 3 (2006), 299–312.

[29] Czyz, E. K., Horwitz, A. G., Eisenberg, D., Kramer, A., and King, C. A. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. Journal of american college health 61, 7 (2013), 398–406.

[30] Earl, A., Nisson, C. A., and Albarracín, D. Stigma cues increase self-conscious emotions and decrease likelihood of attention to information about preventing stigmatized health issues. Acta de investigacion psicologica 5, 1 (2015), 1860–1871.

[31] Eil, D., and Rao, J. M. The good news-bad news effect: asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics 3, 2 (2011), 114–38.

[32] Eiraldi, R. B., Mazzuca, L. B., Clarke, A. T., and Power, T. J. Service utilization among ethnic minority children with adhd: A model of help-seeking behavior. Administration and Policy in Mental Health and Mental Health Services Research 33, 5 (2006), 607–622.

[33] Eisenberg, D., Downs, M. F., Golberstein, E., and Zivin, K. Stigma and help seeking for mental health among college students. Medical Care Research and Review 66, 5 (2009), 522–541.

[34] Eliaz, K., and Schotter, A. Experimental testing of intrinsic preferences for non-instrumental information. American Economic Review 97, 2 (2007), 166–169.

[35] Eliaz, K., and Schotter, A. Paying for confidence: An experimental study of the demand for non-instrumental information. Games and Economic Behavior 70, 2 (2010), 304–324.

[36] Evans-Lacko, S., Knapp, M., McCrone, P., Thornicroft, G., and Mojtabai, R. The mental health consequences of the recession: economic hardship and employment of people with mental health problems in 27 european countries. PloS one 8, 7 (2013), e69792.

[37] Feiler, L. Testing models of information avoidance with binary choice dictator games. Journal of Economic Psychology 45 (2014), 253–267.

[38] Frey, D. Recent research on selective exposure to information. In Advances in experimental social psychology, vol. 19. Elsevier, 1986, pp. 41–80.

[39] Ganguly, A., and Tasoff, J. Fantasy and dread: The demand for information and the consumption utility of the future. Management Science 63, 12 (2016), 4037–4060.

[40] Gary, F. A. Stigma: Barrier to mental health care among ethnic minorities. Issues in mental health nursing 26, 10 (2005), 979–999.

[41] Golman, R., Hagmann, D., and Loewenstein, G. Information avoidance. Journal of Economic Literature 55, 1 (2017), 96–135.

[42] Golman, R., and Loewenstein, G. Curiosity, information gaps, and the utility of knowledge. Information Gaps, and the Utility of Knowledge (April 16, 2015) (2015).

[43] Holton, B., and Pyszczynski, T. Biased information search in the interpersonal domain. Personality and Social Psychology Bulletin 15, 1 (1989), 42–51.

[44] Kahneman, D. A perspective on judgment and choice: mapping bounded rationality. American psychologist 58, 9 (2003), 697.

[45] Karlsson, N., Loewenstein, G., and Seppi, D. The ostrich effect: Selective attention to information. Journal of Risk and uncertainty 38, 2 (2009), 95–115.

[46] Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E., Wang, P., Wells, K. B., and Zaslavsky, A. M. Prevalence and treatment of mental disorders, 1990 to 2003. New England Journal of Medicine 352, 24 (2005), 2515–2523.

[47] Kohn, R., Saxena, S., Levav, I., and Saraceno, B. The treatment gap in mental health care. Bulletin of the World health Organization 82 (2004), 858–866.

[48] Kőszegi, B. Health anxiety and patient behavior. Journal of health economics 22, 6 (2003), 1073–1084.

[49] Kőszegi, B. Ego utility, overconfidence, and task choice. Journal of the European Economic Association 4, 4 (2006), 673–707.

[50] Kőszegi, B. Utility from anticipation and personal equilibrium. Economic Theory 44, 3 (2010), 415–444.

[51] Kőszegi, B., and Rabin, M. Reference-dependent consumption plans. American Economic Review 99, 3 (2009), 909–36.

[52] Lannin, D. G., Vogel, D. L., Brenner, R. E., Abraham, W. T., and Heath, P. J. Does self-stigma reduce the probability of seeking mental health information? Journal of Counseling Psychology 63, 3 (2016), 351.

[53] Link, B. G. Understanding labeling effects in the area of mental disorders: An assessment of the effects of expectations of rejection. American sociological review (1987), 96–112.

[54] Loomes, G., and Sugden, R. Disappointment and dynamic consistency in choice under uncertainty. The Review of Economic Studies 53, 2 (1986), 271–282.

[55] Magaard, J. L., Seeralan, T., Schulz, H., and Brütt, A. L. Factors associated with help-seeking behaviour among individuals with major depression: A systematic review. PLoS One 12, 5 (2017), e0176730.

[56] Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. Managing self-confidence. NBER Working paper (2014).

[57] O'Mahony, P. Psychiatric patient denial of mental illness as a normal process. British Journal of Medical Psychology 55, 2 (1982), 109–118.

[58] Oster, E., Shoulson, I., and Dorsey, E. Optimal expectations and limited medical testing: evidence from huntington disease. American Economic Review 103, 2 (2013), 804–30.

[59] Paradise, M., Cauce, A. M., Ginzler, J., Wert, S., Wruck, K., and Brooker, M. The role of relationships in developmental trajectories of homeless and runaway youth.

[60] Puri, M., and Robinson, D. T. Optimism and economic choice. Journal of Financial Economics 86, 1 (2007), 71–99.

[61] Rabin, M., and Schrag, J. L. First impressions matter: A model of confirmatory bias. The quarterly journal of economics 114, 1 (1999), 37–82.

[62] Schomerus, G., and Angermeyer, M. C. Stigma and its impact on help-seeking for mental disorders: what do we know? Epidemiology and Psychiatric Sciences 17, 1 (2008), 31–37.

[63] Schwardmann, P. Motivated health risk denial and preventative health care investments. Journal of health economics 65 (2019), 78–92.

[64] Sharot, T. The optimism bias. Current biology 21, 23 (2011), R941–R945.

[65] Stolzenburg, S., Freitag, S., Evans-Lacko, S., Muehlan, H., Schmidt, S., and Schomerus, G. The stigma of mental illness as a barrier to self labeling as having a mental illness. The Journal of nervous and mental disease 205, 12 (2017), 903–909.

[66] Sweeny, K., Melnyk, D., Miller, W., and Shepperd, J. A. Information avoidance: Who, what, when, and why. Review of general psychology 14, 4 (2010), 340–353.

[67] Taylor, S. E., and Brown, J. D. Illusion and well-being: a social psychological perspective on mental health. Psychological bulletin 103, 2 (1988), 193.

[68] Thurston, I. B., Phares, V., Coates, E. E., and Bogart, L. M. Child problem recognition and help-seeking intentions among black and white parents. Journal of Clinical Child & Adolescent Psychology 44, 4 (2015), 604–615.

[69] Vogel, D. L., Wade, N. G., and Hackler, A. H. Perceived public stigma and the willingness to seek counseling: The mediating roles of self-stigma and attitudes toward counseling. Journal of Counseling Psychology 54, 1 (2007), 40.

[70] Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bruffaerts, R., Chatterji, S., Chiu, W. T., De Girolamo, G., Fayyad, J., et al. Delay and failure in treatment seeking after first onset of mental disorders in the world mental health survey initiative. In The WHO World Mental Health Surveys: Global perspectives on the epidemiology of mental disorders. Cambridge University Press, 2008, pp. 522–533.

[71] Wang, P. S., Berglund, P. A., Olfson, M., and Kessler, R. C. Delays in initial treatment contact after first onset of a mental disorder. Health services research 39, 2 (2004), 393–416.

[72] Wong, E. C., Collins, R. L., Cerully, J., Seelam, R., and Roth, B. Racial and ethnic differences in mental illness stigma and discrimination among californians experiencing mental health challenges. Rand health quarterly 6, 2 (2017).

[73] Wu, S. Sickness and preventive medical behavior. Journal of health economics 22, 4 (2003), 675–689.

[74] Ybarra, M. L., and Eaton, W. W. Internet-based mental health interventions. Mental health services research 7, 2 (2005), 75–87.