Misperceived Effectiveness and the Demand for Psychotherapy*

Christopher Roth Peter Schwardmann Egon Tripodi

September 13, 2024

Abstract

While psychotherapy has been shown to be effective in treating depression, take-up remains low. In a sample of 1,843 depressed individuals, we document that effectiveness concerns are top-of-mind when respondents consider the value of therapy. We then show that the average respondent underestimates the effectiveness of therapy. An information treatment correcting this misperception increases participants' incentivized willingness to pay for a $320 therapy from $166 to $176. Our evidence suggests that while information can influence therapy demand by altering beliefs and shifting attention, it may not significantly increase demand unless substantial subsidies are provided.

Keywords: Mental Health, Depression, Psychotherapy, Beliefs, Effectiveness, Information policy

*Christopher Roth, University of Cologne, CEPR and ECONtribute, NHH Norwegian School of Economics, Max Planck Institute for Collective Goods; email: roth@wiso.uni-koeln.de; Peter Schwardmann, Carnegie Mellon University, email: pschwardmann@gmail.com; Egon Tripodi, Hertie School of Government, email: egontrpd@gmail.com. We thank Siyan Chen, Zining Chen, Maximilian Fell, Eleonora Guseletova, Apoorv Kanoongo, Frida Küper and Myung Won (Misha) Seong for excellent research assistance. We received ethics approval from the ethics committee of the University of Essex and the University of Cologne. The experiments were pre-registered in the AsPredicted registry (#107190 and #189189). Financial support from BA/Leverhulme Small Research Grants is gratefully acknowledged. Roth acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866. Tripodi acknowledges funding by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119). Online appendix and replication files are available at https://osf.io/vjdcm/.

1 Introduction

Depression is estimated to affect more than 264 million people worldwide, making it one of the primary causes of disability (Shorey et al., 2022; WHO, 2017) and a source of substantial suffering. However, depression is not insurmountable and can be effectively treated with psychotherapy, which aims to change dysfunctional beliefs, thoughts, and behaviors (Cuijpers et al., 2015).[1] In recent years, the availability of online therapy services has made psychotherapy affordable to an ever increasing number of people, while achieving similar mental health outcomes to in-person therapy (Lin et al., 2022). But despite robust evidence for the effectiveness of psychotherapy and its greater availability, a large fraction of the population suffering from mental illness still do not seek treatment (Cronin et al., 2023).

This paper brings causal evidence to bear on the question of whether misperceptions about the effectiveness of psychotherapy act as a meaningful barrier to therapy take-up and whether the provision of accurate information can help remove said barrier. There are at least two reasons why it is far from obvious that an information intervention should affect the demand for therapy. First, it is unclear whether considerations about effectiveness loom large in depressed individuals' decision to seek help. Second, the negative thinking patterns and rigid cognition associated with depression (Everaert et al., 2018; Lefebvre, 1981; Liknaitzky et al., 2017; Ridley et al., 2020) may stand in the way of individuals' belief updating in the face of accurate information about the effectiveness of therapy.

We conduct a pre-registered experiment with 1,843 Americans that suffer from depression symptoms and are not currently undergoing therapy to examine whether low perceived

[1] Psychotherapy was used by close to 10 percent of US adults in 2019 (Terlizzi and Norris, 2020).

effectiveness of therapy acts as a meaningful barrier to therapy demand. We measure demand as respondents' incentivized willingness to pay for online therapy and randomize respondents into one of three conditions. In *Pure control*, respondents complete a willingness to pay elicitation, without having received information or being prompted to think about the effectiveness of therapy. In *Info*, respondents are asked for their prior estimates of effectiveness and are then provided with research evidence on the effectiveness of psychotherapy. In particular, respondents are truthfully told that, according to a study by Cuijpers et al. (2010), 22 out of 22 clinical studies with at least one hundred participants found that therapy is an effective treatment for depression. In *Flag*, respondents are asked about their estimate of the effectiveness of therapy before completing the willingness to pay elicitation.[2] This condition draws attention to effectiveness as a potentially decision-relevant dimension of the demand decision without providing explicit information. Alongside a rich set of post-treatment beliefs, we collect qualitative data on the considerations on top of respondents' minds when deciding on their willingness to pay for therapy.

We document several findings. First, our findings reveal that, alongside financial (67 percent) and time costs (21 percent), effectiveness emerges as the most frequently cited factor (67 percent) when respondents assess their willingness to pay for therapy. Second, we find that beliefs about the effectiveness of therapy are pessimistic, with participants, prior to receiving information, underestimating both the number of studies that find support for the effectiveness of therapy in Cuijpers et al. (2010) and the fraction of depressed individuals that overcome depression with the help of therapy. Third, beliefs are malleable, with accurate

[2]Participants in *Flag* are also told that they will receive information about effectiveness at the end of the survey.

information increasing respondents' beliefs about both the objective effectiveness of therapy and about whether online therapy will work for them. Our fourth and main finding is that making participants more optimistic about the effectiveness of therapy by means of an information intervention increases their willingness to pay for therapy by \$9.83 (t-test, $p = 0.031$) relative to the *Pure control*, with a group mean of \$166.24.

To investigate whether information affects willingness to pay by changing beliefs or by drawing attention to effectiveness as a relevant consideration, we compare therapy demand in *Pure control* with demand in *Flag*, which is designed to raise the salience of effectiveness without correcting misperceptions. We find that willingness to pay for therapy is \$4.28 (t-test, $p = 0.34$) higher among respondents in *Flag*. This effect size corresponds to about 40 percent of the effect of the information treatment and is not statistically significant. Next we examine whether the effect of information over and above merely flagging effectiveness is heterogeneous in pre-treatment beliefs about effectiveness. Consistent with beliefs playing an important role, we find that the effect of information on therapy demand is driven by those with more pessimistic pre-treatment beliefs. Taken together, our findings are consistent with the effect of information being driven by both bringing attention to treatment effectiveness and by shifting beliefs about effectiveness.

Because we elicit our participants' willingness to pay for therapy, we can draw demand curves for each of our treatments. These demand curves allow us to investigate the complementarity of subsidies and information interventions in improving therapy take-up. We find that information campaigns only affect therapy demand when therapy is partially subsidized. Calculating the marginal value of public funds then suggests that, under some assumptions, information also only serves to raise welfare in combination with intermediate subsidies.

Our paper contributes to an interdisciplinary literature that seeks to understand the reasons behind the low take-up of mental health therapy. Using a structural model, Cronin et al. (2023) document that many people who might benefit from therapy do not seek it and that the financial and time costs of therapy can only account for a part of this treatment gap. Survey respondents in Newson et al. (2021) cite concerns about therapy effectiveness as an impediment to seeking treatment and respondents in Andrade et al. (2014) cite it as a cause for dropping out of therapy. In a sample of depressed individuals in India, Bhat et al. (2023) find that people underestimate the efficacy of psychotherapy and that exposure to therapy causally decreases such pessimism. Our paper provides the first causal evidence that pessimistic beliefs about the effectiveness of therapy result in lower demand for therapy.

While our qualitative elicitations reveal that no consideration looms larger than concerns about effectiveness and costs in our respondents' demand decisions, the previous literature has highlighted other information frictions that may curb therapy demand. For example, several studies have uncovered a negative correlation between perceived social stigma and the demand for therapy (see Corrigan et al. 2006 for a review). Contrary to this, Roth et al. (2024) show that higher perceived stigma can cause higher demand for therapy, likely due to it resulting in a higher perceived need for therapy (Andrade et al., 2014). Gulliver et al. (2010) highlight poor mental health literacy as an impediment to seeking therapy. This is confirmed by Acampora et al. (2022), who show how an information intervention aimed at improving mental health literacy increases demand for a mental health app among parts of their sample.

More generally, our findings relate to a burgeoning literature on the causes and consequences of depression and mental illness in economics (Allcott et al., 2020; Angelucci and Bennett, 2024; Banerjee et al., 2023; Biasi et al., 2021; McKelway et al., 2023; De Quidt and

Haushofer, 2016; Ridley, 2023; Ridley et al., 2020; Shreekumar and Vautrey, 2022). Our results suggest the presence of a depression trap: we find that pessimism about therapy effectiveness keeps depressed individuals from seeking the help they need; at the same time, their depression may be the cause of them holding pessimistic beliefs about effectiveness in the first place (Alloy and Ahrens, 1987; Bhat et al., 2023). While these forces may make it prohibitively hard for individuals to become unstuck by themselves, we find that an information intervention that increases the perceived effectiveness of therapy can lead to a meaningful increase in therapy demand.

2 Data

The data collection for the main study was pre-registered on AsPredicted (#107190). We pre-specified the sampling procedure, the main outcomes of interest and the main empirical specifications (see Online Appendix D).[3]

2.1 Design

2.1.1 Structure

At the start of the survey we elicit a series of background characteristics and introduce the Becker-DeGroot-Marschak (BDM) mechanism we use to elicit willingness to pay for therapy. We then randomly assign respondents to one of three treatment groups. Respondents in the *Pure control* group move straight to the elicitation of their willingness to pay for therapy.

[3]Our pre-registration outlines two additional treatments that we conducted at the same time with different samples. This second experiment measures and debiases beliefs about the social stigma associated with depression and is described in Roth et al. (2024) alongside two further experiments that explore stigma. The two experiments outlined in the pre-registration share a control group.

Respondents assigned to *Flag* are asked to state their beliefs about the effectiveness of therapy in treating depression. Respondents in *Info* state their beliefs and then receive information about research evidence on the actual effectiveness of therapy. Respondents in *Info* and *Flag* then state their posterior beliefs about the effectiveness of therapy before stating their willingness to pay for therapy. The structure of our design is summarized in Figure 1.

2.1.2 Becker-DeGroot-Marschak mechanism

After the initial background questions, we illustrate the BDM mechanism in the context of a hypothetical willingness to pay elicitation for a 1-month spa membership, which did not involve real stakes. We tell our respondents that we will ask them for the maximum amount of money they would be willing to pay for the membership. They are further told that they will not have to use their own money to buy the product and that after they stated their valuation, the computer will randomly pick a dollar amount between 0 and 300. Moreover, they learn that if this dollar amount is larger than their valuation, then the dollar amount will be paid out to them, while otherwise they will receive the spa membership. We emphasize that this rule means that it is in the respondent's best interest to state the maximum amount of money they would be willing to pay for the product. To ensure a high understanding of our respondents we include a control question that respondents need to correctly answer.

The hypothetical willingness to pay for an example good we elicit here serves as a useful control variable that can increase statistical power by controlling for idiosyncratic scale use (Dizon-Ross and Jayachandran, 2022).

2.1.3 Effectiveness beliefs

Pre-treatment beliefs Respondents in *Flag* and *Info* are told that researchers have conducted many clinical studies to estimate the effectiveness of psychotherapy for treating depression. They are then told about a comprehensive review that looked at the 22 studies with the largest number of participants. Next, they are asked to estimate how many of these 22 studies show that therapy is an effective treatment for depression.

Respondents are told that one of the questions in which they make quantitative estimates will be randomly selected for payment. They are informed that if their answer in the selected question is within 3 percent of the truth, they will receive a $0.50-dollar bonus.

Treatments Respondents in both *Flag* and *Info* are reminded of their quantitative beliefs about effectiveness. Subsequently, only respondents in *Info* receive research evidence about the effectiveness of therapy for treating depression (Cuijpers et al., 2010). They receive the following instructions:

> The review from Pim Cuijpers and co-authors shows that out of the 22 clinical studies with at least a hundred participants, all 22 studies find that therapy is an effective treatment for depression.

While respondents in *Info* are shown a chart contrasting their estimate with the true value, respondents in *Flag* are shown a chart displaying their guess and are told that they will receive information about effectiveness at the end of the survey. Given the finding of the research study, the treatment should make all respondents weakly more optimistic about therapy effectiveness.

Post-treatment beliefs To assess whether the information treatment changed beliefs, we elicit an incentivized quantitative belief about the effectiveness of therapy as a treatment for depression among respondents in *Flag* and *Info*. We elicit this belief on a different quantitative scale in order to mitigate anchoring and demand effects (Haaland et al., 2023). All respondents are told that the largest of the 22 studies in the review included 818 participants that were diagnosed with depression. We then ask them to estimate the percentage of study participants assigned to take part in psychotherapy that recovered from depression.

To assess whether people extrapolate from beliefs about the general effectiveness of therapy to beliefs about the effectiveness of online therapy for them personally, we ask respondents to guess how likely it is that completing online therapy would be effective for them in overcoming depression. This belief is measured on a 5-point scale ranging from (i) very unlikely to (v) very likely.

2.1.4 Willingness to pay elicitation

Finally, respondents proceed to the elicitation of our main outcome of interest: willingness to pay for a 4 week subscription to BetterHelp, one of the leading online therapy services in the United States. Respondents are given details about the services offered by BetterHelp. They are told that clients of BetterHelp can send audio, video, or text messages to their therapist at any time and that they can schedule weekly live sessions (30 to 45 min) with their therapist to communicate via phone, video, or live chat. To give people a sense of the value of the service, we tell them that the service is normally priced at $320 for 4 weeks. We then ask our respondents to state the maximum amount of dollars they are willing to spend on four weeks of therapy from BetterHelp. Our respondents are further truthfully told that the choice of 10

participants in this study will be implemented. Moreover, respondents are reminded about the details of the BDM mechanism that makes this elicitation incentive compatible.

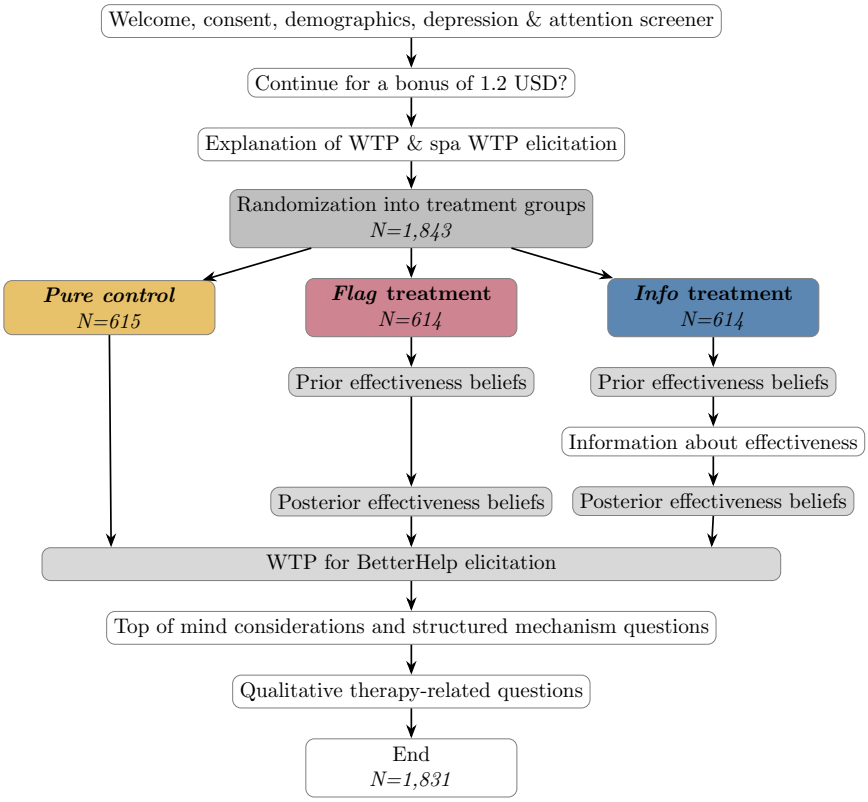

Figure 1: Experimental design

Notes: The study was advertised as a short survey, designed to end with the depression screeners. To subjects who passed both attention and depression screeners, we gave the choice to continue participating in the extended study for an additional payment, which almost everyone took up (only 2.9 percent of participants chose to not continue further). Attrition after randomization is also minimal, with only 12 out of 1843 participants dropping out.

2.1.5 Considerations on top of mind

To shed light on attentional mechanisms, we ask all respondents to write 2-3 sentences on the considerations they have on their mind while deciding on their willingness to pay, broadly following the methodology in Ericsson and Simon (1980).[4]

[4]On the subsequent page, we also ask respondents to select the considerations they had on their mind from a list of 14 considerations. We devised this list of considerations based on pilot data.

To analyze the unstructured text data, we instructed research assistants to assign the text responses into multiple categories. Online Appendix Table A.3 provides an overview of all categories, including an example response.

End-of-survey beliefs After our main outcomes, we elicit a series of additional beliefs and attitudes among all respondents. First, we ask respondents an open-ended question about the study purpose. We then elicit a series of beliefs related to respondents' self-image, the social cost of seeking therapy, and the perceived effectiveness of online therapy for the respondent personally.

2.2 Sample

We recruited 1,843 US respondents using the online platform Prolific, a survey provider commonly used in social science research (Eyal et al., 2021). To qualify for our study, respondents had to pass a standard attention screener. We further restricted our sample to only include respondents that suffer from depression, according to the personal health questionnaire (Kroenke et al., 2009),[5] and have never tried therapy before.

Representativeness Table A.1 examines the representativeness of our sample. As a benchmark we leverage the National Health and Nutrition Examination Survey (NHANES), which is a representative sample of the US population containing data on the PHQ8. We compare our sample to NHANES using the same PHQ8 cutoff. While our survey matches the NHANES in terms of the PHQ8, respondents in our sample are less likely to be female (56% vs 62%,

[5]The PHQ8 is a widely used scale to identify depression. We excluded respondents with a PHQ8 below 10.

$p = 0.03$) and significantly younger (30 vs. 50, $p < 0.01$). These differences in observable characteristics raise the question whether our results would generalize to more representative samples of individuals suffering from depression.

Moreover, our sample consists of depressed individuals who have never participated in therapy. Is it possible that these individuals have particularly pessimistic beliefs? To shed light on this question, we conducted an additional pre-registered survey on Prolific (AsPredicted, #189189). Our sample consists of 310 depressed individuals with and without therapy experience. We find that therapy experience is not significantly associated with effectiveness beliefs (see Online Appendix Table A.2).

3 Perceived Effectiveness and the Demand for Therapy

3.1 Which considerations are top of mind?

We begin by analyzing the considerations that are top of mind for respondents when they decide on their willingness to pay.[6] Online Appendix Figure A.1 displays the most commonly mentioned categories by respondents from the *Pure control* condition, who were not primed on effectiveness. Costs loom large on respondents' minds: 67 percent mention financial costs, 21 percent mention time and 12 percent insurance.[7] Next to these cost considerations, thoughts about effectiveness loom very large in respondents' minds: Approximately 67 percent of respondents mention effectiveness. Moreover, a large fraction of these respondents mention

[6]Note that the framing of the decision context likely affects the measured considerations. For example, if we asked respondents to consider why they haven't been to therapy without asking the WTP question, respondents might provide different responses that weigh costs more heavily.

[7]Approximately 47 percent of study participants report that their health insurance offers some coverage for psychotherapy (see Online Appendix Table A.1).

concerns about low effectiveness of therapy. The following response is representative of the type of responses participants gave:

> I thought about how much it could help me and how much of a difference it could make. [...]

Other commonly mentioned considerations include in-person therapy (10 percent) and other substitutes (8 percent). An explanation for not seeking therapy popular in the literature, related to social concerns and stigma (Corrigan and Rüsch, 2002), is mentioned very rarely (0.5 percent).[8]

Result 1. *Whether or not therapy is effective in reducing depression symptoms looms large in the self-reported considerations that inform respondents' valuation of therapy.*

3.2 How do respondents perceive therapy effectiveness?

To ease interpretation of our main results, we first examine pre-treatment beliefs about therapy effectiveness. Panel A of Figure 2 shows that respondents on average estimate that the number of studies showing therapy to be effective is 16.5 out of 22, compared to a true value of 22 out of 22. To assuage concerns that this underestimation result depends on this specific measure which ruled out overestimation by design, Panel B of Figure 2 displays beliefs about effectiveness with an alternative benchmark, the fraction of patients that overcame depression through therapy in the largest study included in Cuijpers et al. (2010). The figure, again, shows that respondents on average underestimate therapy effectiveness. Respondents in the *Flag* treatment, who respond to this question without learning new information, estimate the

[8]This might arise from the special nature of social concerns, which may be harder to verbalize and are commonly suppressed Smart and Wegner (1999).

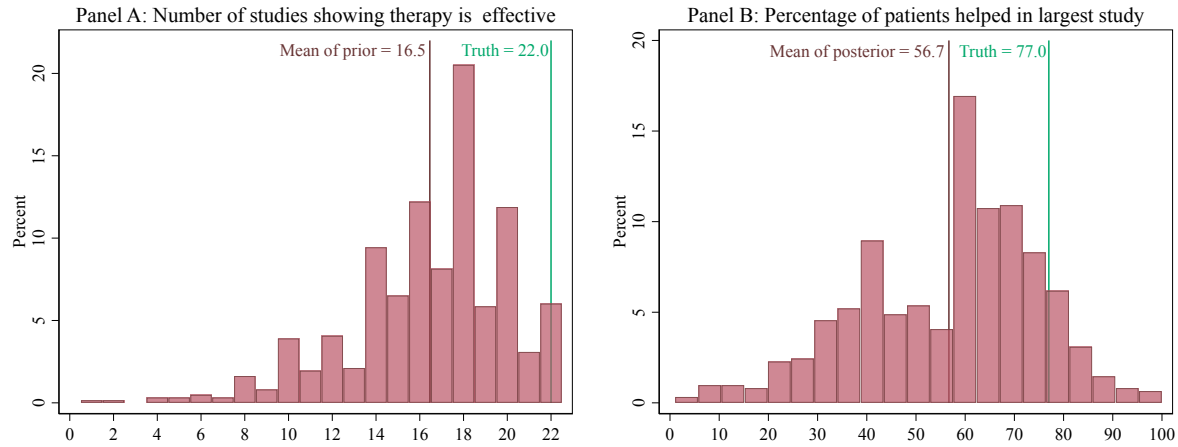

Figure 2: Distribution of effectiveness beliefs in *Flag*

Notes: Panel A reports the distribution of prior effectiveness beliefs that were measured by asking how many of the 22 studies in a meta-analysis find significant positive effects of psychotherapy. Panel B reports the distribution of posterior effectiveness beliefs that were measured by asking what fraction of participants in the largest study of the meta-analysis recover from depression in the treatment group. These distributions refer to the 614 observations in the *Flag* treatment.

fraction to be 56.7 on average, much below the true number (77). Beliefs are widely dispersed with 87 percent of respondents in *Flag* underestimating therapy effectiveness and 11 percent of respondents overestimating effectiveness. Overall, these data highlight the potential of information interventions to shift these beliefs.

We conducted an additional pre-registered survey on Prolific, with 310 depressed and 813 non-depressed individuals. We find that depressed individuals are slightly and significantly more pessimistic about the effectiveness of psychotherapy than non-depressed individuals (see Online Appendix Table A.2). In this sense, our participants' beliefs exhibit the kind of pessimism about therapy effectiveness that has been hypothesized to be symptomatic of depression more generally (Beck et al., 1961).

3.3 Did the research evidence shift beliefs?

Next, we turn to the effects of the information treatment on participants' beliefs. Panel A of Figure 3 shows that information about the high number of studies that find therapy is effective increases respondents' incentivized perceptions of therapy effectiveness. While respondents in *Flag* estimate that, according to one such study, 56.7 percent of participants receiving therapy manage to overcome depression, respondents in *Info* estimate this fraction to be 74.6 percent. The latter is remarkably close to the 77 percent rate of recovery estimated in the referenced study (Rahman et al., 2008), suggesting that the information treatment was successful at debiasing beliefs.

Did the research evidence also change respondents' beliefs about the effectiveness of therapy for them personally? We find that *Info* increases respondents' belief that online therapy would be effective for them personally in overcoming depression by 0.18 of a standard deviation ($p < 0.01$), compared to respondents from *Flag*. This effect is critical, as beliefs about personal therapy effectiveness are plausibly the more decision-relevant beliefs for seeking therapy.

Panel A of Figure 4 illustrates analogous patterns for a qualitative measure of perceived effectiveness of online therapy elicited among all respondents at the end of the survey: Respondents in *Info* are 0.31 standard deviations more optimistic about online therapy for them personally compared to respondents in *Pure control*.[9] Respondents in *Flag* are also somewhat more optimistic about therapy effectiveness (0.13 standard deviations), suggesting that merely asking respondents to think about effectiveness induces some moderate optimism.

[9]Analyses presented in Figure 4 were indicated in the pre-registration as exploratory.

Taken together, our data show that beliefs about effectiveness are malleable and that the information treatment significantly shifted respondents' perceived effectiveness of therapy.

Result 2. *Prior to receiving information, respondents are pessimistic about therapy effectiveness. Research evidence then significantly increases respondents' optimism about how effective therapy can be in general and for them personally.*

3.4 Did the information shift willingness to pay?

We now turn to the question of whether observed treatment effects on beliefs translate into changes in the willingness to pay for therapy, by estimating the following equation:

$$Y_i = \alpha_0 + \alpha_1 \text{Flag}_i + \alpha_2 \text{Info}_i + X_i + \varepsilon_i \tag{1}$$

Flag_i takes value 1 for respondents in *Flag*, and Info_i takes value 1 for respondents who receive research evidence on effectiveness. The omitted category captures respondents in the *Pure control* condition. As pre-specified, our regressions include as controls, X_i, all variables that are elicited pre-treatment, including interest in therapy, PHQ8 score, age, gender, and willingness to pay for the example good.

Panel B of Figure 3 presents the main finding of this paper. We see that information about effectiveness significantly increases willingness to pay for therapy by \$9.83 ($p = 0.031$) relative to the *Pure control* group mean of \$166.24. This corresponds to a moderate effect size of 0.10 of a standard deviation.

Result 3. *Research evidence on high therapy effectiveness increases the demand for online therapy.*

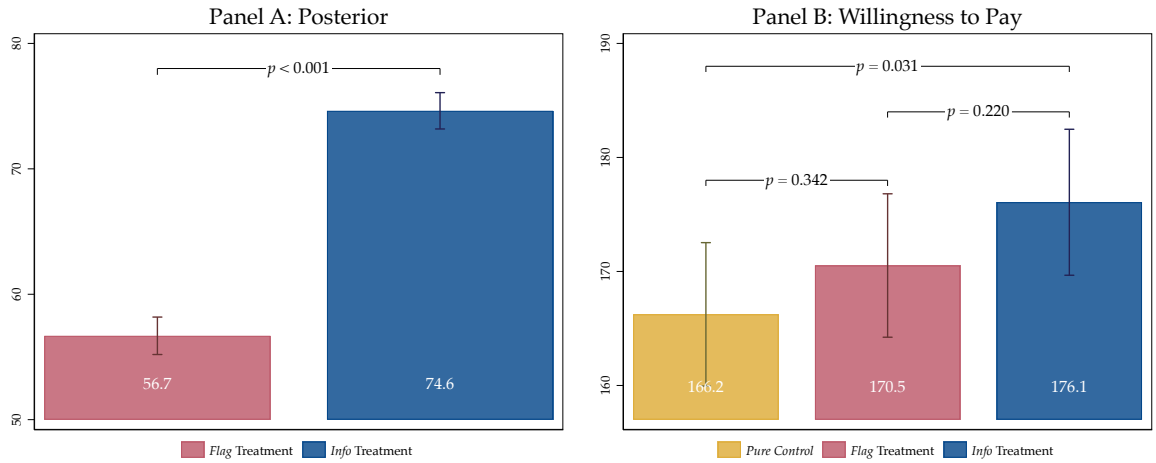

Figure 3: Treatment effects on posterior beliefs and willingness to pay

Notes: This figure presents predicted values derived from pre-registered regression analyses. Panel A focuses on post-treatment quantitative beliefs in the effectiveness of psychotherapy (not elicited in *Pure control*); the control variables used in the regressions include age, gender, the PHQ8 score, and an indicator for previous consideration of individual online therapy for depression. Panel B reports the willingness to pay for BetterHelp; the control variables used in the regressions include age, gender, the PHQ8 score, willingness to pay for a month of spa membership, an indicator for prior awareness of BetterHelp, and an indicator for previous consideration of individual online therapy for depression. 95 percent confidence intervals and p-values are computed using robust standard errors from relevant regressions.

4 Mechanisms

This section examines the respective roles played by attention and belief movements in shaping the observed effects. We then discuss the potential role of cross-learning about other decision-relevant variables in driving the effects of the information intervention.

4.1 Attention versus beliefs

The *Flag* treatment. Recent evidence suggests that information treatments may affect behavior not only by changing beliefs, but also by changing how people allocate their attention across various decision-relevant dimensions (Conlon, 2023). To isolate the effect of attention, we want to estimate the effect of the *Flag* condition.

Willingness to pay in *Flag* is at \$170.5. It is thus \$4.28 higher compared to respondents in *Pure*

control ($p = 0.34$). This, in turn, suggests that directing people's attention to effectiveness per se does not cause a significant increase in people's willingness to pay for therapy. Average willingness to pay for therapy is approximately \$5.50 larger in *Info* compared to *Flag*, but not significantly so ($p = 0.22$). This effect is relatively small given the substantial differences in effectiveness beliefs among respondents in *Info* and *Flag*. The small effect of effectiveness perceptions on therapy demand might stem from financial constraints, social concerns or alternative help-seeking options, which may dampen the impact of increased perceived effectiveness. Taken together, our findings are consistent with the effect of information likely being driven by both bringing attention to treatment effectiveness and by shifting beliefs about effectiveness.

Heterogeneity by pre-treatment beliefs. To provide further evidence for the primacy of belief movements in driving our main treatment effect, we examine the heterogeneity of the treatment effect by pre-treatment beliefs about effectiveness. Here, we use data from respondents in *Flag* and *Info* , the two conditions in which pre-treatment beliefs were elicited. We estimate the following specification:

$$Y_i = \alpha_0 + \alpha_1 \text{Info}_i + \alpha_2 \text{belief}_i + \alpha_3 \text{belief}_i \times \text{Info}_i + \varepsilon_i \tag{2}$$

where belief_i is a measure of pre-treatment beliefs about effectiveness.

Figure A.3 depicts evidence of heterogeneous treatment effects by pre-treatment beliefs. Panel A shows that the research evidence more strongly boosts incentivized beliefs about therapy effectiveness among respondents with a more pessimistic prior. Panel B shows that,

among respondents with below-median pre-treatment beliefs about therapy effectiveness, we observe an increase in willingness to pay of \$11.92 ($p = 0.070$), compared to only \$0.09 ($p = 0.988$) among those with above median beliefs.

Online Appendix Table A.4 also shows that these patterns are robust to a specification that interacts continuous pre-treatment beliefs about therapy effectiveness with the treatment indicator: the interaction coefficient is negative and marginally significant both when the outcome is the posterior belief ($p = 0.035$) and when it is willingness to pay ($p = 0.082$), suggesting that the treatment effect is stronger for respondents with more pessimistic pre-treatment beliefs about effectiveness.

Treatment effects on considerations Panel B of Figure 4 shows that the issues that respondents pay attention to while deciding on their willingness to pay do not vary across treatments. The figure features the top 6 most-cited considerations from the open-ended data. Most importantly, the *Info* treatment does not significantly increase attention to therapy effectiveness compared to *Pure control* and *Flag*.[10]

Taken together, the results in this section lead us to the following conclusion.

Result 4. *The effect of the information treatment is likely driven by both bringing attention to treatment effectiveness and by shifting beliefs about effectiveness.*

[10]Our complementary survey question on structured considerations allows us to better understand the surprising decrease in attention to effectiveness. In *Info*, subjects are significantly less likely than in *Pure control* ($p = 0.042$) to report low effectiveness as a consideration that affected their willingness to pay decision.

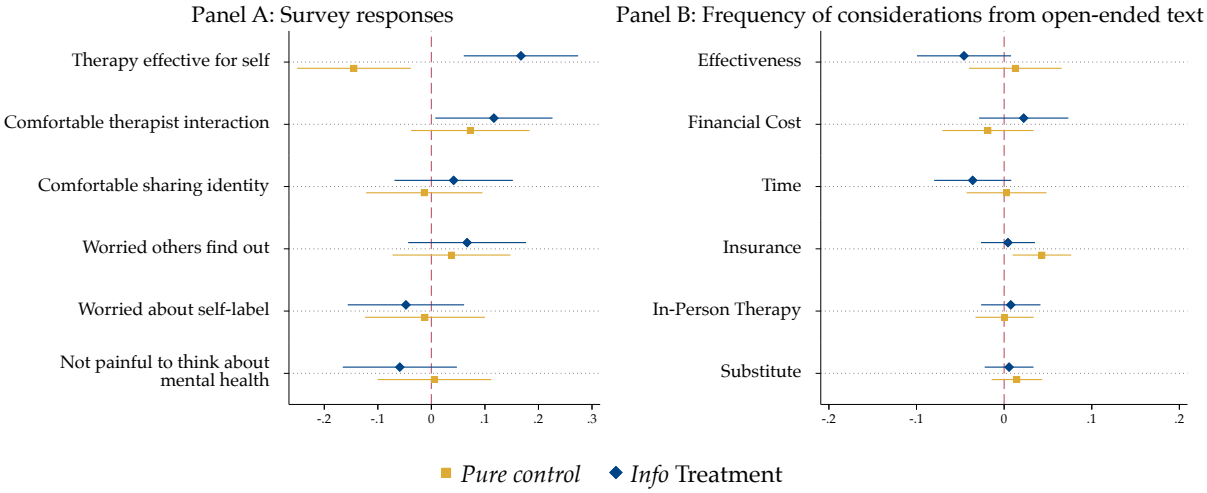

Figure 4: Treatment effects on mechanism outcomes

Notes: The dashed line denotes the *Flag* treatment. Panel A focuses on mechanism questions measured using 5-points Likert scales after the measurement of willingness to pay for therapy; these include a qualitative posterior of effectiveness that we included for all treatments—Therapy effectiveness for self. Panel B focuses on the Top 6 most cited considerations in open-ended considerations that were top of mind for participants when formulating their willingness to pay for therapy. The control variables in the regressions include age, gender, the PHQ8 score, and an indicator for previous consideration of individual online therapy for depression. For comparability of effect sizes, all outcome variables in panel A are standardized by subtracting the mean and dividing by the standard deviation of the variable in the overall sample. 95 percent confidence intervals are based on robust standard errors.

4.2 Cross-learning

One threat to the interpretation of the treatment effects arises from cross-learning i.e., the possibility that respondents update beliefs about other aspects of their mental health, seeking help, and psychotherapy. Panel A of Figure 4 provides evidence of a limited quantitative importance of cross-learning. The information treatment does not change (i) how comfortable respondents are with therapist interactions, (ii) how comfortable they are about sharing they identity, (iii) how worried they are about others finding out about depression, (iv) how worried people are about self-labeling as depressed, and (v) how painful it is to think about mental health related issues. While the above null effects are not very precisely estimated, we can rule out large cross-learning effects, suggesting that the information primarily operates through changes in beliefs about and attention to therapy effectiveness.

5 External validity

In this section we discuss several potential concerns about the external validity of our findings.

Substitution to in-person therapy. In our experiment, we provided respondents with research evidence on the effectiveness of therapy in general, yet our outcome measure is the willingness to pay for online therapy specifically. Although our post-treatment perception measures show that individuals seem to generalize the perceived effectiveness of therapy to online therapy, it is possible that the research evidence, which did not mention, but was based on in-person therapy, could shift people from seeking online to in-person therapy. This mechanism would work against us finding effects and would imply that the treatment effect we observe is a lower bound. We think it is relatively unlikely that this mechanism plays a meaningful role as the open-ended data on consideration shows that there is no differential mention of the in-person therapy substitute across the different treatment arms (see Panel B of Figure 4).

Does willingness to pay translate into actual help seeking? It is unclear whether increases in willingness to pay translate into changes in actual therapy take-up. Evidence from a field experiment in a related paper Roth et al. (2024), suggests that an increase in willingness to pay for online group therapy is significantly positively associated with a higher likelihood of signing up for and scheduling online group therapy. Several papers find willingness to pay to be a valid predictor of actual good purchasing (e.g. Allcott and Taubinsky, 2015).

Experimenter demand effects. While it is possible that the treatments systematically shape respondents' perceptions of our study's purpose, there are several reasons suggesting that experimenter demand is not a major threat to our results. First, the patterns of heterogeneity by pre-treatment optimism about therapy effectiveness imply that one would have to invoke heterogeneously occurring demand effects. Second, our main outcome measure of willingness to pay involves real stakes, making demand effects less plausible. Third, participants in experiments more generally tend to show only weak reactions to clear indications of the experimenter's expectations (de Quidt et al., 2018).

To further mitigate concerns, we leverage participants' open-ended responses to the following question: "Please describe in a few words what you think the aim was of the research conducted through this survey." Online Appendix Figure A.4 reveals that approximately one percent of respondents thought that we were interested in measuring the effect of effectiveness perceptions or information on the demand for therapy. Excluding those respondents leaves our main results unaltered.[11]

6 Welfare and policy implications

Because we elicit each participant's willingness to pay for therapy, we can draw the demand curves for our treatment and control conditions. This facilitates speaking about the policy implications of our findings. We note that the demand curves may not be completely represen-

[11] In other studies that used a similar open-ended question to measure beliefs about the study purpose, approximately 10 percent of respondents correctly guessed the hypothesis (Andre et al., 2023; Jäger et al., 2024; Schwardmann et al., 2022). This suggests that the study purpose was relatively difficult to guess. Indeed, most responses are more generic: 50 percent of respondents mention that the survey tried to measure people's valuation of therapy. 16 percent mention perceptions of online therapy, 15 percent mention opinions about therapy, and close to 9 percent discuss determinants of therapy demand. 7 percent of respondents explicitly indicate being unsure about the purpose.

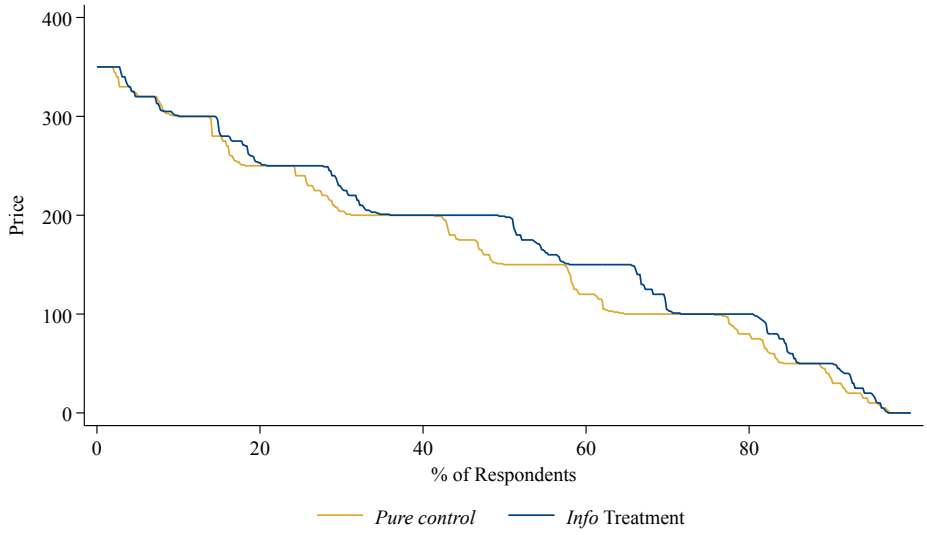

Figure 5: Willingness to Pay in the *Pure control* and *Info* conditions

Notes: A Kolmogorov-Smirnov test rejects the hypothesis that the two Willingness to Pay distributions are the same ($p = 0.011$).

tative. For example, Prolific participants tend to be less wealthy than the general population. Moreover, we do not measure the demand for any depression therapy, but for the services provided by BetterHelp in the presence of substitutes. Finally, the scale of our willingness to pay elicitation, which ranges from zero to $30 above the market price of BetterHelp services, is likely to underestimate the subjective utility benefits of demanding the service, because participants may be reluctant to state a willingness to pay at the top end of the scale.

With these caveats in mind, we can use our demand curves to shed some light on the effect of subsidies or (partial) insurance coverage of therapy services and their interaction with our information treatment. Figure 5 shows that demand curves are downward sloping everywhere and that any increase in subsidies would result in fairly constant marginal increases in demand.

Comparing the demand curves of *Pure control* and *Info* we see that effectiveness informa-

tion only affects therapy demand for intermediate prices, but not for fully subsidized services or for services that charge market rate. Online Appendix Table B.2 makes this point explicit. Information increases demand by 6.8 percentage points (t-test, $p = 0.008$) at a price of \$150 and by 5.8 percentage points (t-test, $p = 0.016$) at a price at \$100, but has at most a modest effect at high or low prices. Information therefore complements therapy subsidies initially, but its effect disappears for full subsidies.

In Online Appendix B we use our demand estimates and estimated treatment effects to deliver a more formal analysis of the potential effects of subsidizing therapy or of providing information about its effectiveness. Taking our elicited willingness to pay for therapy at face value, we calculate the marginal value of public funds (Hendren and Sprung-Keyser, 2020) to more comprehensively accounts for the costs and benefits associated with different policy mixes.

A key benefit of a subsidy or an information campaign lies in correcting the internality that stems from individuals underestimating the effectiveness of therapy. However, we also consider that the take-up of therapy may be associated with a positive externality on an individual's employer or on their friends and family. Relatedly, we allow for the fact that therapy and any subsequent decrease in depression may increase productivity and may therefore result in more taxable income.

On the cost side we allow for the fact that information campaigns will be directed at many individuals who do not react to them. In the case of subsidies, a misallocation can arise. The higher a subsidy, the higher the share of individuals who buy therapy despite the social value that their therapy take-up generates being below the cost of providing therapy.

We confirm that information campaigns generally produce the highest marginal value of

public funds at intermediate subsidy levels. At the same time, the misallocation associated with subsidies results in a marginal value of public funds that is decreasing in the size of the subsidy, an effect that trumps the internality-correcting effect of intermediate subsidies.

7 Conclusion

Our experiment examines whether perceptions of therapy effectiveness impede help seeking. We first establish that respondents significantly underestimate the effectiveness of therapy. An information treatment correcting this misperception increases participants' incentivized willingness to pay for therapy from \$166 to \$176 for a therapy that costs \$320. Thus, the information is unlikely to affect demand in the absence of substantial subsidies. Information likely affects therapy demand both by changing beliefs and by shifting attention.

The significant effect of information on willingness to pay for therapy we observe contrasts with a series of studies that document rather muted effects of information about the effectiveness of goal-directed behaviors in other domains, ranging from climate-friendly consumption to political participation (Gerber et al., 2020; Hager et al., 2024; Imai et al., 2022). Future work might further fine-tune the exact content and delivery mode of information. For example, we conjecture that qualitative and personal success stories may be even more potent in persistently shifting beliefs about therapy effectiveness than statistical evidence (Graeber et al., 2024). In addition, the source of information (e.g. people with prior therapy experience) may matter crucially for shaping its impact (Alsan and Eichmeyer, 2023). However, it is important to consider that the impact of information interventions may diminish over time (Gerber et al., 2011; Hill et al., 2013), especially when presented numerically (Graeber et al.,

2024). This raises the question of whether the information interventions we study may also lack persistence, especially given the well-documented pessimistic beliefs associated with depression.[12]

[12]Our main study also collected another measure of interest in therapy. A few days after the experiment we sent participants direct messages inviting them to click-through to a survey in which they had a chance to win one of two further BetterHelp vouchers. Roughly 16 percent of participants clicked through, with no significant difference between treatments (F-test, $p = 0.874$). We worry that the low click-through rate was due to participants not seeing or engaging with the direct message, a flaw in our design that could yield any analysis of treatment differences woefully underpowered.

References

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

Allcott, Hunt and Dmitry Taubinsky, "Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market," *American Economic Review*, 2015, *105* (8), 2501–2538.

_ **, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow**, "The welfare effects of social media," *American Economic Review*, 2020, *110* (3), 629–676.

Alloy, Lauren B and Anthony H Ahrens, "Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others.," *Journal of personality and social psychology*, 1987, *52* (2), 366.

Alsan, Marcella and Sarah Beate Eichmeyer, "Experimental evidence on the effectiveness of nonexperts for improving vaccine demand," *American Economic Journal: Economic Policy*, 2023, *16* (1), 1–21.

Andrade, Laura Helena, J Alonso, Z Mneimneh, JE Wells, A Al-Hamzawi, G Borges, E Bromet, Ronny Bruffaerts, G De Girolamo, R De Graaf et al., "Barriers to mental health treatment: results from the WHO World Mental Health surveys," *Psychological medicine*, 2014, *44* (6), 1303–1317.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart, "Narratives about the Macroeconomy," 2023.

Angelucci, Manuela and Daniel Bennett, "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy," *American economic review*, 2024, *114* (1), 169–98.

Banerjee, Abhijit, Esther Duflo, Erin Grela, Madeline McKelway, Frank Schilbach, Garima Sharma, and Girija Vaidyanathan, "Depression and Loneliness among the Elderly in Low-and Middle-Income Countries," *Journal of Economic Perspectives*, 2023, *37* (2), 179–202.

Beck, Aaron T, Calvin Ward, Moch Mendelson, J Mock, and JJAGP Erbaugh, "Beck depression inventory (BDI)," *Arch Gen Psychiatry*, 1961, *4* (6), 561–571.

Bhat, Bhargav, Jonathan de Quidt, Johannes Haushofer, Vikram H Patel, Gautam Rao, Frank Schilbach, and Pierre-Luc P Vautrey, "The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes," Technical Report, National Bureau of Economic Research 2023.

Biasi, Barbara, Michael S Dahl, and Petra Moser, "Mental Health, Creativity, and Wealth," Technical Report, National Bureau of Economic Research 2021.

Conlon, John J, "Attention, Information, and Persuasion," 2023.

Corrigan, Patrick W, Amy C Watson, and Leah Barr, "The self–stigma of mental illness: Implications for self–esteem and self–efficacy," *Journal of social and clinical psychology*, 2006, *25* (8), 875–884.

_ **and Nicolas Rüsch**, "Mental illness stereotypes and clinical care: do people avoid treatment because of stigma?," *Psychiatric Rehabilitation Skills*, 2002, *6* (3), 312–334.

Cronin, Christopher J, Matthew P Forsstrom, and Nicholas W Papageorge, "What Good are Treatment Effects Without Treatment? Mental Health and the Reluctance to use Talk Therapy," *Review of Economic Studies*, 2023.

Cuijpers, Pim, Erica Weitz, Eirini Karyotaki, Judy Garber, and Gerhard Andersson, "The effects of psychological treatment of maternal depression on children and parental functioning: a meta-analysis," *European child & adolescent psychiatry*, 2015, *24* (2), 237–245.

_ **, Filip Smit, Ernst Bohlmeijer, Steven D Hollon, and Gerhard Andersson**, "Efficacy of cognitive–behavioural therapy and other psychological treatments for adult depression:

meta-analytic study of publication bias," *The British Journal of Psychiatry*, 2010, *196* (3), 173–178.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Dizon-Ross, Rebecca and Seema Jayachandran, "Improving Willingness-to-Pay Elicitation by Including a Benchmark Good," in "AEA Papers and Proceedings," Vol. 112 2022, pp. 551–55.

Ericsson, K Anders and Herbert A Simon, "Verbal reports as data.," *Psychological review*, 1980, *87* (3), 215.

Everaert, Jonas, Michael V Bronstein, Tyrone D Cannon, and Jutta Joormann, "Looking through tinted glasses: Depression and social anxiety are related to both interpretation biases and inflexible negative interpretations," *Clinical Psychological Science*, 2018, *6* (4), 517–528.

Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina, "Data Quality of Platforms and Panels for Online Behavioral Research," *Behavior Research Methods*, 2021, pp. 1–20.

Gerber, Alan, Mitchell Hoffman, John Morgan, and Collin Raymond, "One in a million: Field experiments on perceived closeness of the election and voter turnout," *American Economic Journal: Applied Economics*, 2020, *12* (3), 287–325.

Gerber, Alan S, James G Gimpel, Donald P Green, and Daron R Shaw, "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment," *American Political Science Review*, 2011, *105* (1), 135–150.

Graeber, Thomas, Florian Zimmermann, and Christopher Roth, "Stories, statistics, and memory," *Quarterly Journal of Economics*, 2024.

Greenberg, Paul E, Andree-Anne Fournier, Tammy Sisitsky, Mark Simes, Richard Berman, Sarah H Koenigsberg, and Ronald C Kessler, "The Economic Burden of Adults with Major

Depressive Disorder in the United States (2010 and 2018)," *PharmacoEconomics*, 2021, *39* (6), 653–665.

Gulliver, Amelia, Kathleen M Griffiths, and Helen Christensen, "Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review," *BMC psychiatry*, 2010, *10* (1), 1–9.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," *Journal of Economic Literature*, 2023, *61* (1), 3–40.

Hager, Anselm, Lukas Hensel, Johannes Hermle, and Christopher Roth, "Political Activists are Not Driven by Instrumental Motives: Evidence from Two Natural Field Experiments," *British Journal of Political Science*, 2024.

Hendren, Nathaniel and Ben Sprung-Keyser, "A Unified Welfare Analysis of Government Policies*," *The Quarterly Journal of Economics*, 03 2020, *135* (3), 1209–1318.

Hill, Seth J, James Lo, Lynn Vavreck, and John Zaller, "How quickly we forget: The duration of persuasion effects from mass communication," *Political Communication*, 2013, *30* (4), 521–547.

Imai, Taisuke, Davide D Pace, Peter Schwardmann, and Joël J van der Weele, "Correcting Consumer Misperceptions About CO 2 Emissions," 2022.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer, "Worker beliefs about outside options," *Quarterly Journal of Economics*, 2024.

Kroenke, Kurt, Tara W Strine, Robert L Spitzer, Janet BW Williams, Joyce T Berry, and Ali H Mokdad, "The PHQ-8 as a measure of current depression in the general population," *Journal of affective disorders*, 2009, *114* (1-3), 163–173.

Lefebvre, Mark F, "Cognitive distortion and cognitive errors in depressed psychiatric and low back pain patients.," *Journal of consulting and clinical psychology*, 1981, *49* (4), 517.

Liknaitzky, Paul, Luke D Smillie, and Nicholas B Allen, "Out-of-the-blue: Depressive symptoms are associated with deficits in processing inferential expectancy-violations using a novel cognitive rigidity task," *Cognitive Therapy and Research*, 2017, *41*, 757–776.

Lin, Tao, Timothy G Heckman, and Timothy Anderson, "The efficacy of synchronous teletherapy versus in-person therapy: A meta-analysis of randomized clinical trials.," *Clinical Psychology: Science and Practice*, 2022, *29* (2), 167.

McKelway, Madeline, Abhijit Banerjee, Erin Grela, Frank Schilbach, Miriam Sequeira, Garima Sharma, Girija Vaidyanathan, and Esther Duflo, "Effects of Cognitive Behavioral Therapy and Cash Transfers on Older Persons Living Alone in India: A Randomized Trial," *Annals of Internal Medicine*, 2023, *176* (5), 632–641.

Newson, J, V Pastukh, J Taylor, and T Thiagarajan, "Mental Health has Bigger Challenges Than Stigma: Mental Health Million Project 2021," 2021.

Quidt, Jonathan De and Johannes Haushofer, "Depression for economists," Technical Report, National Bureau of Economic Research 2016.

Rahman, Atif, Abid Malik, Siham Sikander, Christopher Roberts, and Francis Creed, "Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised controlled trial," *The Lancet*, 2008, *372* (9642), 902–909.

Ridley, Matthew, "Mental Illness Discrimination," 2023. Working paper.

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

Roth, Christopher, Peter Schwardmann, and Egon Tripodi, "Depression Stigma," Technical Report, CRC TRR 190 Rationality and Competition 2024.

Schwardmann, Peter, Egon Tripodi, and Joël J Van der Weele, "Self-persuasion: Evidence from field experiments at international debating competitions," *American Economic Review*, 2022, *112* (4), 1118–1146.

Shorey, Shefaly, Esperanza Debby Ng, and Celine HJ Wong, "Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis," *British Journal of Clinical Psychology*, 2022, *61* (2), 287–305.

Shreekumar, Advik and Pierre Luc Vautrey, "Managing emotions: The effects of online mindfulness meditation on mental health and economic behavior," Technical Report, Tech. Rep., MIT 2022.

Smart, Laura and Daniel M Wegner, "Covering up what can't be seen: concealable stigma and mental control.," *Journal of personality and social psychology*, 1999, *77* (3), 474.

Terlizzi, Emily P and Tina Norris, "Mental health treatment among adults: United States, 2020," 2020.

WHO, "Depression and Other Common Mental Disorders: Global Health Estimates," 2017.