University of Cape Town

Masters in Economic Development

Half Dissertation

ECO5066W

Mental health and social decision making:

How depression alters the way we trust

Rowan Spazzoli

SPZROW001

Supervisor: Malcolm Keswell

In partial fulfilment of the requirements for
MASTERS OF COMMERCE
SPECIALISING IN ECONOMIC DEVELOPMENT

October 22, 2019

Contents

List of Tables

Acknowledgements

Thank you to the legion of people that encouraged me and cheered me on in the process of writing this dissertation. I experienced multiple setbacks over the period when this was written, including suffering from depression and anxiety, family problems, being the victim of a robbery and various issues in completing the thesis. I made it through thanks to the unrelenting support of my incredible family and friends.

The irony of writing a dissertation on depression while suffering from it is not lost on me. However, having people around me that openly talked about mental health problems made seeking help much easier. In January last year, I started on anti-depressant and anti-anxiety medication, which has changed my life. Thank you to my therapist and my psychiatrist for all the work they have done with me, I am eternally grateful. I encourage anyone suffering from mental health issues to speak up and not be afraid to seek help. You don't have to fight the battle alone.

Thank you to the many friends that helped me through the tough patches and celebrated with me during the successes. A special thank you to the following, who were always there when I needed them: Jared, Holly, Janine, Hana, Samantha, Jess, Edward, Marcelle, Charlotte, Kuhle, Kay, Steve, Michael, Jonathan, Nick, Abigail, Thabo, Ahmed, Renée, Grace, Suzie, Shelly, Dale, Alison, Tiisetso, Alexei, Anna, Tiang, Louis, Rhiannon, Cait, Steph, Ndumi and Tumi.

Thank you to my supervisor, Malcolm, for all the support with the dissertation and for being so patient with me throughout the process, especially when I was going through a difficult time last year. Your input has been invaluable and I have learnt so much from you.

Thank you to my wonderful family for all they have done for me. My dad, Lihor, for all the support he has provided me throughout my life, particularly in my academic endeavours. My brothers, Lorenzo and Fabio, for being such wonderful siblings. Thank you to my uncle Massimo, for being supportive of my dreams and thank you to Sue, Jacqui, Jules and Matt for inspiring me. Thank you to the Weston family, who have been there whenever I've needed them.

Finally, and most importantly, thank you to my incredible mother Debbie for everything she has done for me. You are always there when I need someone to chat to, always ready to help and you have sacrificed so much for me. You've experienced every part of the journey with me, from panicking at the registration line with my degree choice, to the teary phone call after the tests in 2015 and, now, to celebrating completing my masters and finding my way in life. Thank you from the bottom of my heart, I love you and truly would not have made it without you.

My journey as a student at UCT has come to the end after eight years, three degrees, thousands of hours studying and many late nights. Thank you to UCT for all the phenomenal memories and for shaping me into who I am today. It's time to close this chapter of my life. As the last line of my old school song says:

"Here our ship once anchor'd and here its course was set."

Abstract

Depression is one of the most prevalent causes of disease burden in the world, with a particularly high prevalence in South Africa. Significant evidence exists for how depression affects employment, income and education, but there is little research on how it affects social decision making. This dissertation addresses the direct impact of depression on the prosocial behaviours that influence economic outcomes, specifically trust and trustworthiness, rather than the economic outcomes themselves. Using experimental and survey data from a randomised control trial, I show that depression reduces trust but, counter-intuitively, increases the trustworthiness of individuals. Additionally, I show that the Activate! programme reduces depression in men and increases trust in all participants. These results have significant implications for how we consider depression in economics, particularly on how it affects poverty by influencing prosocial decision making.

1 Introduction

Depression is one of the most widespread health problems globally, with many individuals experiencing it at some point in their life. The prevalence of depression in South Africa is one of the highest in the world, and it disproportionately affects poor, female and black African individuals. A significant body of literature exists on depression and its effects on and interaction with health and poverty. Additionally, there has been shown to be a strong relationship between prosocial behaviours, which include preferences and beliefs regarding trust and trustworthiness, and poverty. However, there is little evidence tying these together and examining the relationship and causal pathways between depression on prosocial behaviours. This dissertation contributes to this research agenda.

A starting point in this dissertation is to examine the literature on depression. Major depressive disorder (interchangeably referred to as depression) is one of the most widespread health conditions in the world. According to the World Health Organisation, depression is characterised by sadness, loss of interest and/or pleasure, feelings of low self-worth or guilt, decreased energy and poor concentration (World Health Organization, 2017). It can be long-lasting and/or recurring and can significantly impair a person's ability to hold employment, receive an education or function in normal life. (World Health Organization, 2017). Depression is the second most prevalent cause of disease burden globally, in terms of years lost to disability, and second only to HIV/AIDS in Southern Sub-Saharan Africa (Vos et al., 2014). It is estimated that between 4% and 5 % of the global population suffers from depression at any given time, with a male prevalence of 3.21% and a female prevalence of 5.48% (Vos et al., 2014).

At an international level, depression has been shown to have a major economic impact. For example, in Europe, the total annual cost of depression was estimated to be 118 billion, comprised of 42 billion of direct costs and 76 billion of indirect costs, which in total corresponds to 1% of total European GDP (Sobocki et al., 2006). A 2003 study based on the American Productivity Audit calculated that depression accounts for 48% of lost productive time in US workers, amounting to an estimated cost of $44 billion per year to US employers (Stewart et al., 2003).

In a review of three global mental health reports across six low-income countries, Patel and Kleinman (2003) find an association between various indicators of poverty and the risk of mental health disorders, with low levels of education being the most consistent affiliation. However, they go on to show that there is weak evidence to suggest a specific link to low-income levels. Instead, they find that depression affects factors such as lack of security, social change, risks of violence/harm and poor physical health. They also show that the direct and indirect costs of poor mental health are likely to further impact a person's economic condition, in turn setting up a cycle of poor mental health and poverty.

Another study for the World Bank by Das et al. (2007) explored the relationship between poverty and mental health. They found that people who are female, older and in poor health are more likely to have worse levels of mental health. They find little evidence between higher education and mental health as well as no consistent relationship between consumption poverty and mental health. They conclude that changes in life circumstances are likely to have a greater impact on mental health outcomes than poverty levels. However, poor individuals are less likely to have security or be protected from adverse events. In line with Patel and Kleinman (2003), they recommend that policy should focus on targeted interventions that help protect individuals from the effects of shocks and adverse events.

An interesting paper by Quidt and Haushofer (2016) develops a theoretical economic model which predicts the symptoms of depression from economic primitives such as beliefs. They show in their model that when a person's beliefs of the returns to their efforts are affected exogenously by a shock, they will exhibit symptoms of depression and that if these effects are high enough they can create a poverty trap.

Examining the literature on depression in South Africa paints a stark picture of mental health in the country. The South African Stress and Health (SASH) study found that depression is one of the most prevalent individual lifetime disorders in the country, with 9.8% of their respondents being diagnosed with it using the DSM-IV system (Herman et al., 2009). Furthermore, the 12-month prevalence of depression was 4.8%, making it the most common mental disorder in the study. It is clear from this information alone that that depression is a major area of concern in South Africa and that there is a need for policymakers to formulate approaches to manage the problem. Apart from dealing with it because it is a problem in itself, there is evidence that depression is associated with other critical areas of concern in South Africa.

For instance, several papers have examined the impact of depression on other social outcomes. A study on the predictors of drug use among South African adolescents showed that those with higher levels of drug use reported higher levels of depression (Brook et al., 2006). Additionally, among South African adults, it has been shown that higher levels of depression are associated with lower levels of education, less income stability and a lower perceived social status (Hamad et al., 2008).

Research done at a primary care clinic in the Cape Flats, a low-income area in Cape Town, showed that depression was associated with a worse perception of overall health and had a strong, direct association with anxiety disorder (Triant, 2002). Furthermore, fatigue was a higher predictor of depression, along with symptoms such as joint pain, palpitations and nausea.

Additionally, depression can have an inter-generational nature and a recent paper by Eyal and Burns (2017) has shown the inter-generational nature in South Africa. Using the longitudinal NIDS data-set, they estimate that one parent experiencing depression increases the chances of having a child who is depressed by more than 7 times. Having two parents who are depressed raises the risk by 14 times. Transmission of paternal depression is shown to be higher than the transmission of maternal depression.

A paper by Ardington and Case (2010) gives a descriptive account of the relationship between mental health and socioeconomic status in South Africa, using data from the National Income Dynamics Study (NIDS). They find that depression scores in South Africa are higher than those found in other countries and that the scores for black African men and women are 39% and 43% higher than white men and women respectively. Sixty percent of this gap can be explained by socioeconomic status, with household expenditure, number of assets and educational attainment having a significant, negative correlation with depression. Finally, they find that physical health is positively correlated with mental health (Ardington and Case, 2010).

Based on the depression literature, there is significant evidence that poverty and depression are inextricably intertwined. However, many other factors that have been linked to poverty. Recent literature in the field

of behavioural economics has shown that poverty is strongly related to diminished prosocial behaviours and low community efficacy. In other words, communities that are poorer have been shown exhibit less trust. For example, on a macro level, Slemrod and Katuscak (2005) show that lower levels of trust are strongly related to lower levels of income. On a micro level, Ludwig et al. (2012) examine this relationship in their paper on "neighbourhood effects." They find that poverty affects "collective efficacy", which is the willingness of members of a community to work together in supporting shared norms. Furthermore, Ludwig et al. (2012) tie the neighbourhood effects and poverty to mental health, with poorer communities exhibiting lower mental and physical health. It is therefore important to understand the causal pathways between depression, trust, trustworthiness and poverty, and specifically the effect of depression on trust and trustworthiness.

Using data collected from a randomised control trial conducted in a youth leadership programme in South Africa, I investigate the relationship and causal pathways between depression, trust and trustworthiness. The key result of this investigation is that increases in depression are shown to increase trustworthiness. Additionally, the results show that the programme was able to increase various measures of trust among participants and reduce depression in male participants. A mechanism is then uncovered whereby the reduction of male depression caused the men who were treated in the programme to exhibit less trustworthiness. This dissertation contributes to a gap in the literature, as trustworthiness has been shown to increase with depression, but no evidence existed regarding the causal nature of this relationship.

The rest of the dissertation is arranged as follows. Section 2 unpacks the literature surrounding trust and trustworthiness, including the various measures of trust and how these can be interpreted. Section 3 discusses existing literature on the interaction between depression, trust and trustworthiness. Section 4 describes how the data used in the dissertation was collected and provides summary statistics for the key variables that were used. Section 5 explains the methods that were used in the data analysis and presents the results from this analysis. Section 6 summarises and discusses the key results and ties these to existing literature. Finally, Section 7 concludes the dissertation and draws together the main findings.

2 Trust and Trustworthiness

Trust is a crucial concept in economics, particularly because it is an important part of social and economic exchanges. Trust helps facilitate transactions in real-world markets, where incomplete contracts and information asymmetries would otherwise reduce their efficiency and hamper their functioning (Arrow, 1973).

There is a significant volume of literature on how trust has an impact and relationship with various economic outcomes. For example, it has been shown to be correlated with higher rates of economic growth and more equal incomes. Additionally, trust has been seen to strengthen institutions against predatory practices (Knack and Keefer, 1997).

There is evidence of the impact of trust on economic growth both empirically and in economic models. Zak and Knack (2001) use a general equilibrium growth model to show that investment decreases with lower amounts of trust, as informal institutions are weaker, and this, in turn, has a negative impact on income growth. They give strong empirical evidence from cross country work to support this, showing that trust and the social factors that influence it are important for growth in economies.

The other side of the trust relationship is **trustworthiness** (also sometimes referred to as reciprocity), where the trustee shows reciprocal behaviour in response to the person placing the trust (known as the trustor). There is less literature available that explores the effect of trustworthiness on economic outcomes than there is of trust on these outcomes. However, Bohnet and Baytelman (2007) argue that where there is a greater level of trustworthy behaviour in a society, there is more reason for individuals to be trusting. Furthermore, they argue that trust is endogenous in society, as it can only exist where there is an environment or culture of trustworthiness. They differentiate between two types of trustworthiness: intrinsic trustworthiness, which is culturally related, and conditional trustworthiness, which is brought on by strong institutions that punish cheating and encourage trustworthiness.

Breuer and McDermott (2010), use a question from the World Value Surveys (WVS) as a proxy for trustworthiness. This question asks parents for their subjective view on teaching their children respect and tolerance (note, this is not the same as the WVS question used in this dissertation). They find that trustworthiness is consistently significant in explaining increases in per capita output. They repeat this with various controls, and although the magnitude of the coefficient is reduced, the significance remains strong. Based on this, they conclude that increased trustworthiness, as well as trust, are equally important in generating economic output, with trust crucial to start transactions and trustworthiness important to realise output.

Another paper by Slemrod and Katuscak (2005) brings together trust and trustworthiness by investigating the relationship between these and an individual's income. This is done using both a theoretical model and empirical data from the World Values Survey. They find that higher levels of trust have a positive impact on an individual's income, while higher levels of trustworthiness have a negative impact on an individual's income. However, the payoff to trustworthiness increases when trust in a country is higher, with returns to trustworthiness being negative in countries with low levels of trust and positive in countries with high levels of trust. These results suggest that it is possible that some countries are in an equilibrium trap, where no one is willing to invest in trust or trustworthiness, resulting in potentially productive relationships not forming.

As trust and trustworthiness have been shown to be significant for growth, there is reason to investigate what influences it on an individual level. Evidence from a study of US localities in Alesina and La Ferrara (2002) shows that the strongest factors associated with low trust are a recent history of traumatic experiences, being part of a group that has been discriminated against (particularly for black African individuals), low levels of income or education and living in areas of income inequality.

In a paper that uses various experimental methods, Cox et al. (2016) examine trustworthiness and what drives it. They find that other-regarding preferences are the primary driver of trustworthiness. This includes altruism and inequality aversion toward the trustor. The next most important driver of trustworthiness that they identify is vulnerability responsiveness, where a trustee reacts to a trustor's vulnerability by implementing actions that will not harm the trustor. Additionally, as discussed previously, Bohnet and Baytelman (2007) distinguish between intrinsic trustworthiness and conditional trustworthiness, which means that it is both inherent and driven by institutions.

Given the historical context of South Africa, the evidence that trust is impacted by discrimination and income inequality is of great interest. The Activate! programme, from which the data was collected for this dissertation, aims to foster and grow pro-social behaviours, such as trust, in youth from marginalised communities. This dissertation, therefore, looks at the success of the programme in improving this trust and how trust may have been affected by other factors, specifically depression.

To investigate these relationships, trust needs to be defined and understood. Additionally, the various measures of trust need to be examined and explained in greater depth. This is done in the subsections below.

2.1 Defining Trust and Trustworthiness

To be able to investigate trust, we first need to define it. There have been many attempts in the economics literature to develop a conclusive definition of trust, though they all vary slightly. For this dissertation, I will be using the definition of trust from the seminal paper on the economics and biology of trust by Fehr (2009), which is based on the work of Coleman (1990). The definition reads as follows:

> "An individual (lets call her the trustor or investor) trusts if she voluntarily places resources at the disposal of another party (the trustee) without any legal commitment from the latter. In addition, the act of trust is associated with an expectation that the act will pay off in terms of the investors goals. In particular, if the trustee is trustworthy the investor is better off than if trust were not placed, whereas if the trustee is not trustworthy the investor is worse off than if trust were not placed" (Fehr, 2009, pg. 238)

This quote is useful as it allows us to conceptualise trust as a behaviour and then to be able to begin measuring it. Furthermore, it provides us with an explanation and determination of trustworthiness, which is the trustee's reciprocal behaviour when responding to the trust placed in them. In this dissertation, the terms reciprocity and trustworthiness will be used interchangeably.

2.2 Elicitation Methods of Trust and Trustworthiness

The importance of trust, in economics generally and as a component of social capital, means that its measurement has become a key topic in the literature. Two main approaches are taken when measuring trust, namely survey-based methods and experimental methods.

Various types of survey questions and methods have been used to ascertain an individual's trust. One of the most frequently used questions first appeared in Rosenberg (1956) and now features on the American General Social Survey, the European Values Survey and the World Values Survey. This question is generally referred to as the **World Values Survey (WVS)** trust question and reads as follows: Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people? (1) Most people can be trusted (2) Can't be too careful

Although this is the most common type of survey question, many variations exist. Examples from Miller and Mitamura (2003) include using the same question on a Likert scale of 1-5 or 1-7 and asking about how much a person trusts various groups of people including neighbours, people in their work/school, acquaintances and strangers. The reason for these variations in the survey-based measures is the debate as to whether the WVS is an adequate measure of trust. This is discussed in more detail in section 2.3.

There are also several experimental methods that have been used to measure trust. The most well known and commonly used of these is the "Investment Game" also known as the "Trust Game" developed by Berg et al. (1995). In the game, a person (called the investor) is given an initial endowment of 10 units of currency. They then have the choice to send none, some or all of this to another player (the trustee) with the knowledge that whatever they send will be doubled or tripled. The trustee then chooses whether to send anything back to the investor and can choose how much to send, which can range to nothing to the full amount. If an assumption of rationality or self regarding preferences is taken, the outcome of this game would be that investors would not send any positive amount across. This is because trustees would be expected to keep the entire amount they received, preferring to have more money than less. With this knowledge, investors would not invest, leading to a sub-game perfect equilibrium that is Pareto-inferior as the sum of both parties would be two or three times less than what it could have been (Cox, 2004).

However, when running this experiment Berg et al. (1995) found that investors in the game transferred \$5.16 (out of a possible \$10) and trustees returned \$4.66. Upon observing this, they conclude that reciprocity is a basic component of behaviour as evidenced by the return of the trustees and the trust extended by the investors. Additionally, they find that, when administering a second treatment, social history can strengthen the relationship between trust and reciprocity.

Other games have also been used either to measure trust or to control for other potential influences in the trust game. For example, Blount (1995) makes use of the "ultimatum game", where Player A is given an endowment and offers a portion to Player B. If Player B accepts the offer, they both keep their allocation but if Player B rejects it then they both receive nothing. She varied the selection of the first move being from Player A or being randomly allocated, and found that there were lower rejection rates by Player B when it was randomly allocation, demonstrating that a component of Player B's decision is whether to punish Player A for their allocation.

Another method that has been used by Cox (2004) is a triadic design with three variations of the trust game to identify trusting and reciprocating behaviour. The first game is the same as the standard trust game as developed by Berg et al. (1995). The second is the "Dictator Game" where Player A makes an allocation but Player B does not get to make any decision and must accept Player A's decision. In the third game, Player As do not have a decision to make and both Player A and B are allocated amounts from a previous "first game", with Player B deciding how much to send back. The third game is a reverse dictator which uses the random allocation concept from Blount (1995) to remove the potential of Player B punishing Player A. Cox (2004) finds evidence from this that altruistic preferences, beliefs about others' altruistic preferences and beliefs about others' reciprocating behaviour have an impact on both trust and reciprocity.

2.3 Interpreting the Various Measures of Trust

As can be seen, a variety of different methods have been developed to try and measure trust. The question is, what do each of these methods measure? Furthermore, what are the components that form part of these measurements?

The WVS question has been used extensively in exploring the relationship between trust and other factors,

with various researching attempting to explain whether it is a valid or accurate measure of trust. Johnson and Mislin (2012) compared the WVS question to results from the trust game and find that it is strongly, positively correlated with experimentally measured trust. This is true when implementing various experimental controls including anonymity, initial endowment and expected return and controlling for variables such as the region where the participants are from. Additionally, they find no relationship between the WVS measure and experimental trustworthiness.

However, despite this evidence and the frequent use of the question, it has also been shown to be a poor measure of a person's trust. Miller and Mitamura (2003) note that the WVS question is likely to be conflated with caution as it does not ask about trust vs distrust. Instead, it asks about trust vs caution, and the two terms are not necessarily opposites to one another. They further show that, in a cross-cultural context, the wording leads to misinterpretations concerning trust levels when compared to a number of other survey measures of trust. In their paper, all other measures of trust show that American college students are more trusting than Japanese college students, except in the WVS question, where the opposite result is observed. They note that although the student sample makes it difficult to generalise, there needs to be caution when using the question as a measure of trust.

Miller and Mitamura (2003) propose that a Likert Scale is more likely to help elicit trust, which would rule out having people agree with both answers. Doing so would involve asking the question on a 5 or 7 point scale, as opposed to having a binary option like in the WVS question. The question can also be phrased without mention of the word "caution", and instead ask the question from "do not trust at all" to "trust completely".

Additionally, Glaeser et al. (2000) combine two experimental methods and survey measures of trust, one of which is the Berg et al. (1995) trust game. They find that 10 different variations of survey questions about trust, including the WVS question, have no significant relationship or correlation with trust by Player A in the trust games that were conducted. However, they find that these questions do appear to predict trustworthiness by Player B in their experiments. A replication of this study by Lazzarini et al. (2003) showed the same results, where the WVS question did not significantly explain Player A trusting behaviour but was significantly, positively correlated with Player B trustworthiness.

In contrast to this, another study by Fehr et al. (2003) uses a survey of a large sample of German households that contains the WVS question and other questions on trust. Uniquely, the survey combined experimental methods by integrating the trust game into the survey. They did this by firstly asking half of the respondents to answer questions as "Player A" in the game. Then, they told half the respondents that they were "Player B" and would be matched to a "Player A". The Player B then made a decision for each possible transfer of Player A. The two players are matched up ex-post, with one A for every B, and the actual transfers are mailed to the respective players. In effect, Fehr et al. (2003) were able to play a trust game using a national level experiment and then compare it to the survey questions on trust. Their results showed that the sender's actions (Player A) in the trust game are positively correlated with survey-based trust measures. They also find that trustworthiness in the trust game is not correlated with the survey measures.

These contradictions raise questions about whether the trust game or the WVS are good indicators of trust and what exactly they are measuring. Sapienza et al. (2013) explore this, with the starting point that a sender's behaviour in the trust game is a combination of their beliefs about trustworthiness and preferences (such as risk aversion and altruism). They then run an investigation using both survey questions and a modified trust game where all participants play as both senders and receivers.

They find that the quantity sent in the trust game is correlated with both the sender's expectations of trustworthiness and his preferences. Furthermore, the sender's expectations are correlated with the WVS

trust question and other trust questions. Finally, they find a correlation between the sender's expectations of trustworthiness and the sender's actual trustworthiness when he plays as a receiver. They conclude that the best measure from the trust game is not the amount sent by Player A, but Player A's expectation of what Player B will return. This is because it is not influenced by preferences (such as risk preferences) and represents the closest measure of beliefs about the trustworthiness of others, which Sapienza et al. (2013) say represents the best measure of trust.

However, Fehr (2009) would disagree with this conclusion, stating that the best measure of trust is, in fact, the amount sent by the sender and not their expectation of trustworthiness. He explains that trust is not solely about an individual's beliefs, but their risk preferences too. An essential component to trust, as per his definition, is vulnerability, which is influenced by a person's appetite for risk. He concludes that the sender's expectations of trustworthiness best represent beliefs, but the amount sent is the best indicator of trust as it includes both beliefs and preferences.

A final consideration in this area is the role of altruism, also referred to as unconditional kindness, in trust and trustworthy behaviour. Ashraf et al. (2006) run trust and dictator games as well as measuring expectations of return and risk preferences across South African, Russian and American participants. This is done to understand the various motivations underlying trust and trustworthiness. They find that expectation of return is the major component of trust but that unconditional kindness explains some of the variance. On the other hand, trustworthiness is mostly explained by unconditional kindness with reciprocity playing a smaller role. Interestingly, people behave similarly across the various nationalities studied.

Brülhart and Usunier (2012) approach the question slightly differently, and test altruism by observing the difference between what trustors give to poor trustees compared to rich trustees. They find that they give no more to poor than rich, confirming that trust is the dominant motivation for their behaviour. However, this doesn't necessarily go against the findings of Ashraf et al. (2006), particularly as it only shows that altruism isn't the main effect, which means the findings are aligned.

To conclude, trust and trustworthiness are complex behaviours that are influenced by a multitude of factors. Based on a reading of the literature, it can be said that trust, as observed through trusting behaviour, is influenced by risk preferences, social preferences (including altruism), a person's beliefs about the trustworthiness of others and their own trustworthiness. Furthermore, trustworthiness is best measured by the amount returned in the trust game and is influenced to some degree by altruism and reciprocity. These relationships are summarised in Figure 1 below.

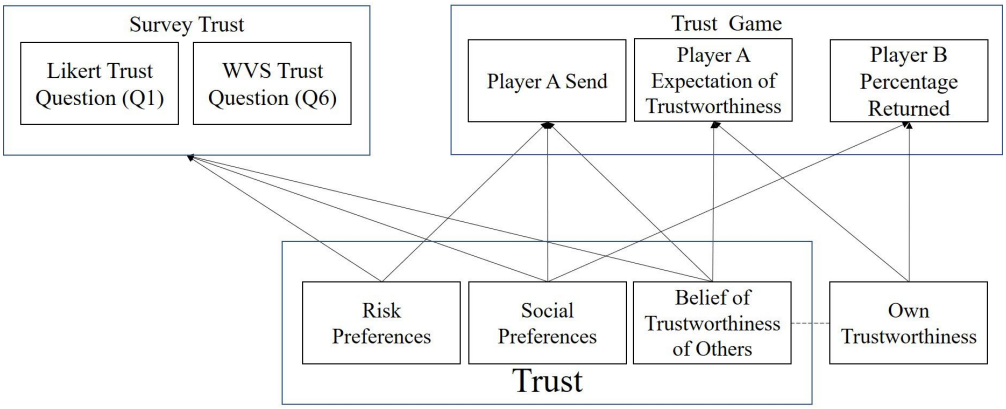

Figure 1: Summary of components and measures of trust based on the trust literature

3 The Influence of Depression on Trust and Trustworthiness

An area that has not been widely explored is the impact of depression on social behaviour, specifically trust and trustworthiness. Although many studies have been done on the impact of depression on economic outcomes, as referenced previously, there is little consideration of the impact on the social decision making that often determines economic outcomes.

Some evidence exists regarding the relationship between depression and trust. One example is a study on trust by Zak and Fakhar (2006), done at a country level and based on the World Values Survey questions. Their results show that happiness was significantly and positively correlated to trust while depression was significantly and negatively correlated to trust. Interestingly, they found that depression explained the largest variation in trust among the 80 different variables that were examined. A partial explanation for this could be that depressed individuals are potentially more risk-averse. A study by Kupferberg et al. (2016) investigated whether depressed individuals would avoid competition in favour of a cooperative payment scheme. Their key finding was that parties with depression were less likely to choose the competitive payment option when compared to the healthy controls, indicating that the depressed individuals were more risk-averse.

In contrast to the paper by Zak and Fakhar (2006), a thesis by Mellick (2014) examines the link between trust and depression at a micro-level. The trust game was played with 38 depressed inpatients and 38 healthy controls, all of which were adolescent girls. The results showed that the depressed girls made significantly higher offers in the trust game when compared to the healthy controls. Additionally, when the groups were divided by the degree of depression the severely depressed girls were shown to make higher offers in the trust game compared to the girls with lower levels of depression

A more significant volume of literature exists on the link between depression and trustworthiness. This is because the various components of trustworthiness have been explored in relation to depression, such as altruistic behaviour, cooperative behaviour and reciprocity. A paper by Fujiwara (2007) shows that people who exhibit higher levels of altruistic behaviour (AB), a component of trustworthiness, were more likely to be depressed. The study was conducted using a survey from the United States and was therefore unable to show causality of the association. The results showed that the relationship between AB and depression was significant and sizeable, with the highest AB group being 59% more likely to have depression than the lowest AB group.

Several studies have been done using the"**dictator game**" to examine the effect of depression on social decision making. Harlé et al. (2010) examined the role of depression on social-economic decision making by enrolling 15 depressed and 23 non-depressed participants in the dictator game. They found that, although depressed individuals reacted with a greater negative emotional response to unfair offers, they accepted a significantly higher number of these offers. They conclude that, in light of other literature, depression results in biasing of a person's decision making.

In another study, Destoop et al. (2012) had 39 patients diagnosed with depression and 22 healthy controls play a modified version of the ultimatum game, both in the role of proposer and responder against the same person. They found that patients with depression made significantly higher offers compared to the control group in the ultimatum game when compared to the healthy controls. One potential explanation for the higher offer made by depressed individuals is that depression has been shown to be linked to a decreased responsiveness to reward sensitivity, resulting in a lack of interested in maximising monetary returns (Henriques and Davidson, 2000). Additionally, Destoop et al. (2012) suggest that the higher offers by depressed individuals may be due to deficits in second-order mentalising, which involves imaginative mental activity that allows a person to interpret or perceive another person's behaviour. This corresponds

to Harlé et al. (2010) conclusion that a depressed individual's decision making is biased.

In contrast to both Harlé et al. (2010) and Destoop et al. (2012), Wang et al. (2014) used the dictator game to provided evidence that patients with depression showed impaired bargaining behaviour in a social decision-making situation. The acceptance rates of depressed individuals were lower than the control group. However, although the direction of the effect is different from the Destoop et al. (2012) and Harlé et al. (2010) studies, it is complementary to them. This is because both studies state that an individual's decision making is impaired by depression, which is in line with the conclusion by Wang et al. (2014).

Many studies that utilise the "**trust game**" found similar results for the effect of depression on trustworthiness. Ong et al. (2017) conducted a study on individuals with a history of mood disorders, specifically depression and bipolar disorder, using the trust game. They find that depressed individuals are more trustworthy, in that they return higher amounts to trustees when compared to healthy controls. They conclude that although people with mood disorders suffer from disruptions in their social decision making they may also display more prosocial behaviours, such as increased cooperation.

Clark et al. (2013) assess the impact of depression on cooperation using four different games, which were the trust game, the prisoner's dilemma game, the dictator game and the public goods game. They found that there were significant correlations between an individual's performance in the games and depressive symptoms. These showed a negative relationship between depressive symptoms and sustained reciprocal cooperation, even when controlling for characteristics such as ages, gender, education and marital status. They conclude that there is a need for further analysis in this area, particularly in establishing causality, and that these economic games are useful in assessing the association between depressive symptoms and impaired social functioning.

De Silva et al. (2005) conducted a review of 21 studies that examined the relationship between mental disorders and social capital, with 14 done at the individual level and 7 done at a macro level. They found that there was evidence for an inverse relationship between social capital and mental disorders such as depression. However, they concluded that the evidence is currently not sufficient to draw conclusions about the use of social capital interventions to treat mental health problems.

Another interesting paper by Caceda et al. (2014) examines the relationship between prosocial behaviours, gender and depression. In the background to the paper, they cite existing evidence that women normally display higher levels of prosocial behaviours, such as trustworthiness, than men do. Their study used a modified version of the trust game with depressed men, depressed women, healthy women and healthy men. They found that depressed men exhibited reciprocity to a significantly higher extent than healthy men. Furthermore, depression was shown to invert the self-centred, gender-specific behaviour. Based on this, they conclude that depression is related to the reversal of gender-specific prosocial behaviours.

Finally, King-Casas and Chiu (2012) argue that multiplayer exchange games, such as the trust game, are appropriate ways to study the effect of psychiatric illness. These games, combined with functional magnetic neuroimaging, provide a way of understanding the neurobiology of neural dysfunction.

To conclude, the literature shows that there is a relationship between depression and social behaviour. However, there are gaps in the literature in terms of understanding the causal nature of this relationship, as there are no studies that use the necessary tools to establish causation. Additionally, other factors, such as gender, need to be taken into account, particularly with the evidence that it can have effects on both social behaviours and mental health issues.

4 Data Collection and Description

The data used in this dissertation is from an evaluation that was conducted the Activate! Change Drivers programme. In order to make any type of statement about the impact of the Activate! programme, one needs a control group (i.e. the counterfactual outcome) to know what *would have* taken place in the lives of the beneficiaries had they not participated in the programme. Randomising assignment to the intervention is often the best way to achieve this counterfactual information. This study used a variant of this approach called "pipeline randomisation", which does not deny access to the intervention but merely delays it. This approach effectively creates a queue/pipeline, and one's place in the queue is randomly assigned. In the case of this study, successful applicants to the programme were randomly assigned to participate either in 2014 or in 2015. As such, the 2015 group forms the control/counterfactual group, whilst the 2014 group constitutes the treatment group. Importantly the outcomes of the control group (both survey-based outcomes as well as behavioural outcomes) were measured *before* the control group embarked on the programme (between January-July 2015). For the trust game, both the treatment and control groups were randomised into Player As and Player Bs. Thus, we are able to look at both trust (proxied by the behaviour of the Player As) as well as trustworthiness (proxied by the behaviour of the Player Bs).

4.1 Activate! programme Description

The Activate! Change Drivers programme is a youth development programme that is open to South Africans between the ages of 18 and 30. These individuals are brought together as "Change Drivers" who, throughout the programme, develop their sense of self, their network and their skill-set so that they are able to drive development and innovation in the country. The programme is structured as a set of three residential modularised workshops that run from Saturday to Sunday the following week, with gaps of around 8-10 weeks between the modules. The modules cover a number of areas including political engagement, self-belief, project management, innovative thinking, trust-building, teamwork and goal setting (Activate!, 2019).

The underpinning objective of the programme is to alter the preferences and beliefs of young people, particularly those from marginalised communities. This helps spark what can be described as active "social capital" which facilitates leadership from individuals so they can begin to positively change their communities.

There was no specific intervention within Activate! that targeted mental health or depression. However, the nexus of interventions used was thought to potentially influence depression. On that basis, questions were included in the baseline and endline survey that could be used to construct a measure of depression.

4.2 Randomised Control Trial

Unpacking the causal effects of a programme such as Activate! is a difficult task. Improvements that are seen may have taken place regardless of whether participants had taken part in the programme. Ideally, we would need to have a counterfactual = to be able to draw causality, where we could know what would have happened to the same participants if they hadn't participated. The problem is that an individual is either in the programme or not, which means that we cannot observe both scenarios applied to the same person. Observing an individual over time will not give a reliable estimate of the effect of a programme on that individual, as other factors that affect the outcomes might have changed after the programme is introduced (Duflo et al., 2007).

Though observing a person in two different states is not possible, a credible way of creating a coun-

terfactual is to randomise selected participants to treatment and control groups. This makes it possible to measure the average difference of the treatment between two similar groups, the group who were part of the programme and those who were not. It is important to note that the comparison group needs to be similar to the treatment group, as the difference in traits between groups would affect the outcome. For example, if the groups differ by factors such as location or income level, the difference in their results would be as a result of both the programme as well as the pre-existing difference, known at the selection bias. (Duflo et al., 2007). Therefore, the treatment and control group need to be as similar as possible.

One way of doing this is to select the people who qualify for the programme first. Then, given that this group is selected based on similar traits, randomise who receives the treatment first and who receives it second. This type of randomisation is known as a "pipeline" randomisation, where individuals are not denied the programme but instead join it at different times. The "treatment" group embarked on the Activate! programme first, followed by the "control" group, who participated 6 months later. The baseline surveys were conducted at the start of the first programme and the endline surveys were done between the end of the first and the beginning of the second.

The randomisation to treatment and control is done to remove selection bias entirely. According to Duflo et al. (2007), when a randomised evaluation is designed and implemented correctly, it provides an unbiased estimate of the impact of that programme on the sample that is being studied. In this way, we are able to attribute the differences that arise between the groups to the programme. By having a control group we can make causal inferences about the effect of the programme and not to any unobservable factors

This design helps us to be confident about the measured impacts of the programme for two reasons. First, the focus only on *successful* applicants means that the study sample will be homogeneous with respect to the selection criteria – leadership and innovation potential, as well as a commitment to the public good. Second, by randomising the year of training, the preferences, behaviours and other outcomes of the study participants should be statistically the same in the treatment and control groups, thus allowing us to attribute any *differences* that arise after completion of the training programme by the treatment group to the intervention itself and not other unobserved factors.

4.3 Baseline and endline Survey

Both the control and the treatment groups were surveyed at the beginning of the first programme. This survey included questions that screened for depression as well as demographic traits, the educational levels of participants and their parents, their employment history and a number of other characteristics. This section gives an overview of the summary statistics from the baseline survey that was done with participants as well as some of their survey health data. **Table 1** shows the summary statistics of various characteristics of the participants.

Overall there are more male than female participants, with 58.3% male overall and 55.8% in the control group compared to 60.5% in the treatment group. The participants are predominantly black African, with 97.7% of the control participants and 96.3% of the treatment participants being black African, resulting in 97% of the total participants being black African. Coloured individuals make up 2.31% of the control group and 2.99% of the treatment group, meaning that 2.87% of the total participants are coloured. There are no Indian participants in the control group, though they make up 1% of the treatment group. There are no white participants in either group. The demographic composition of the participants is likely because the population is drawn from disadvantaged areas, which consist mostly of black African and coloured individuals in South Africa due to the history of the country.

The next part of Table 1 shows the break down of the home language of the participants. There are a low number of people with English as their home language, with only 2.32% of the control and 3.99% of the treatment group reporting that it is their primary language. The majority of the participants responded that they speak either Zulu, Xhosa, Sepedi and Setswana as their home language. In the control group, 27.4% speak Zulu, 21.6% speak Xhosa, 15.4% speak Sepedi and 10.0% speak Setswana as their home language. The proportions are similar in the treatment group, with responses indicating that 29.2% speak Zulu, 16.3% speak Xhosa, 13.6% speak Sepedi and 11.0% speak Setswana as their home language.

The share of home languages spoken is similar in proportion to the national distribution, except for English and Afrikaans which are mostly spoken by white individuals as a first language. According to the 2011 census, 22.7% speak Zulu, 16.0% speak isiXhosa, 9.1% speak Sepedi and 8.0% speak Setswana. Therefore the languages spoken in the sample set is reflective of the demographic make-up across South Africa, except for English and Afrikaans, as there are no white participants in the programme.

The data on the relationship status of the individuals of the programme shows that 96.5% of the control group and 97.3% of the treatment group are not married or have not been married before. Additionally, less than 2% of the participants are either married or living with a partner. Based on the age of the participants, who are between 18 and 30, this result is expected.

Table 1 also show the education statistics of the group. The completion rate of matric amongst the sample set is low, with 53.5% of the control and 46.7% of the treatment having completed matric. This is understandable, as the group is drawn from primarily disadvantaged backgrounds. In terms of tertiary education, 33.7% of the control group and 39.0% of the treatment group have completed some type of tertiary training after matric. The overall education level of the group is in line with what we would expect, as Activate! is specifically for individuals who show leadership potential but might not be able to realise this due to their background.

Table 1: Summary Statistics: Demographic Data

	Control	Treatment	Total
Gender (1=Male)	0.562	0.601	0.583
	(0.497)	(0.490)	(0.494)
African (1=yes)	0.977	0.963	0.970
	(0.150)	(0.188)	(0.172)
Coloured (1=yes)	0.0231	0.0299	0.0267
	(0.150)	(0.171)	(0.161)
Indian (1=yes)	0	0.00664	0.00357
	(0)	(0.0814)	(0.0597)
White (1=yes)	0	0	0
	(0)	(0)	(0)
English	0.0232	0.0399	0.0321
	(0.151)	(0.196)	(0.177)
Afrikaans	0.0154	0.0299	0.0232
	(0.124)	(0.171)	(0.151)
IsiZulu	0.274	0.292	0.284
	(0.447)	(0.456)	(0.451)
IsiXhosa	0.216	0.163	0.188
	(0.412)	(0.370)	(0.391)
Sesotho	0.0927	0.0864	0.0893
	(0.291)	(0.281)	(0.285)
Tshivenda	0.0309	0.0266	0.0286
	(0.173)	(0.161)	(0.167)
IsiTsonga	0.0386	0.0532	0.0464
	(0.193)	(0.225)	(0.211)
SiSwati	0.0386	0.0399	0.0393
	(0.193)	(0.196)	(0.194)
Setswana	0.100	0.110	0.105
	(0.301)	(0.313)	(0.307)
Sepedi	0.154	0.136	0.145
	(0.362)	(0.344)	(0.352)
IsiNdebele	0.0116	0.0233	0.0179
	(0.107)	(0.151)	(0.133)
Other	0.00386	0	0.00179
	(0.0621)	(0)	(0.0423)
Married (1=yes)	0.0116	0.00664	0.00893
	(0.107)	(0.0814)	(0.0942)
Living with a Partner (1=yes)	0.0193	0.0166	0.0179
	(0.138)	(0.128)	(0.133)
Never Married (1=yes)	0.965	0.973	0.970
	(0.183)	(0.161)	(0.172)
Completed matric (1=yes)	0.535	0.497	0.515
	(0.500)	(0.501)	(0.500)
Completed some type of tertiary training after matric (1=yes)	0.337	0.390	0.366
	(0.473)	(0.489)	(0.482)
Unemployed (1=yes)	0.139	0.131	0.134
	(0.346)	(0.337)	(0.341)
Observations	610		

1. * $p < 0.01$, ** $p < 0.05$, *** $p < 0.10$.
2. Std. errors reported in parenthesis, robust to individual clustering.

Table 2 summarises some of the general physical and mental health data that was captured during the survey. The participants were asked to report on their current present health, with the options including excellent, very good, good, fair or poor. The table shows the participants that answered yes to excellent, with 45.0% of the control group and 44.9% of the treatment group reporting that they were in excellent health. Additionally, most participants reported being in very good or good health, with less than 5% in both the treatment and control group reporting that their present health was fair or poor.

The survey also asked about the frequency of exercise of the participants. This is of particular interest as there is evidence across multiple studies that has shown the bi-directional relationship of depression and physical exercise, showing that low amounts of exercise can reduce depression and that depression reduces the likelihood that a person will exercise (Teychenne et al., 2008). In total, 11.2% of the control group and 13.0% of the treatment group reported to never exercise. Conversely, 17.7% of the control group and 14.6% of the treatment group reported that they exercise more than 3 times a week. Overall, the amount of exercise that the participants do is fairly distributed between low, moderate and high frequency per week.

With regards to the use of alcohol, it is interesting to note that a high number of the participants do not drink alcohol, with 41.1% of the control and 39.6% of the treatment group responding that they abstain from it. However, according to the World Health Organisation, South Africa has one of the highest proportion of people who abstain from drinking, with 65% of South Africans never having drunk alcohol, despite the quantity that is consumed on a per capita basis being relatively high (World Health Organization, 2017). Therefore, more participants in the Activate! data appear to consume alcohol than the national average.

The proportion of individuals in the study that have smoked cigarettes in the past are 15.9% of the control and 13.9% of the treatment group, and those that are currently smoking are 10.4% of the control and 14.0% of the treatment group. This is slightly lower than the percentage of people that smoke nationally, which is estimated to be 17.6% (Reddy et al., 2015). Therefore the proportion of smokers in the Activate! study is similar to the national estimate.

In total, 84.6% of the control group and 88.0% of the treatment group reported that they have taken an HIV test. This is significantly higher than the proportion of South Africans who have reported having gone for an HIV test, with 65.5% of individuals in the country responding in survey data that they have been tested for HIV at least one (Johnson et al., 2015).

Included in the general health data summary statistics are 4 questions that can be used as general indicators of a person's mental health. The first two concern the willingness of a person to stay or leave the neighbourhood that they currently reside in. This is of interest as there is evidence of neighbourhood effects on mental health, where parents who moved to low poverty areas report less distress and boys who moved reporting fewer depressive and anxious symptoms than those who stay (Leventhal and Brooks-Gunn, 2003). The proportion of participants who have a strong desire to stay in their neighbourhood is fairly low, with 30.5% of the control group and 29.0% of the treatment group indicating that they want to stay. Additionally, 22.3% of the control group and 18.0% of the treatment group indicated that they have a strong desire to leave their neighbourhood.

The last two questions in table 2 ask about how happy and how in control the participants feel regarding their life. When asked whether they were happy or very happy with life 68.4% of the control group and 69.2% of the treatment group indicating that they were. Finally, with 80.1% of the control group and 77.4% of the treatment group indicated that they felt they were either totally in control or had control over most things in their lives.

Table 2: Summary Statistics: General Health Data

	Control	Treatment	Total
Present Health - Excellent (1=yes)	0.450	0.449	0.449
	(0.498)	(0.498)	(0.498)
Exercise - Never	0.112	0.130	0.121
	(0.315)	(0.336)	(0.327)
Exercise - Less than once a week	0.104	0.0963	0.0998
	(0.306)	(0.296)	(0.300)
Exercise - Once a week	0.108	0.143	0.127
	(0.311)	(0.351)	(0.333)
Exercise - 3 or more times a week	0.177	0.146	0.160
	(0.382)	(0.354)	(0.367)
Don't drink alcohol (1=yes)	0.411	0.396	0.403
	(0.493)	(0.490)	(0.491)
Smoke cigarettes (1=yes)	0.104	0.140	0.123
	(0.306)	(0.347)	(0.329)
Smoke cigarettes in the past (1=yes)	0.159	0.139	0.148
	(0.366)	(0.347)	(0.356)
HIV test (1=yes)	0.846	0.880	0.864
	(0.361)	(0.326)	(0.343)
Strong preference to stay in neighbourhood (1=yes)	0.305	0.290	0.297
	(0.461)	(0.454)	(0.457)
Strong preference to leave neighbourhood (1=yes)	0.223	0.180	0.200
	(0.417)	(0.385)	(0.400)
Happy or very happy with life (1=yes)	0.684	0.692	0.689
	(0.466)	(0.462)	(0.463)
Totally in control or control over most things in life (1=yes)	0.801	0.774	0.787
	(0.400)	(0.419)	(0.410)
Observations	610		

1. * $p < 0.01$, ** $p < 0.05$, *** $p < 0.10$.
2. Std. errors reported in parenthesis, robust to individual clustering.

4.4 Trust Data

In order to measure trust in the Activate! study, a combination of survey questions and experimental methods were used. The questions were administered at endline alongside other survey questions. A variety of questions were used and were intended to measure perceptions of trust as well as perceptions of self and others.

A slight variation of the WVS survey question mentioned in section 2.2 is used (Q6), with the difference being that the second part of the question is "most people can't be trusted" as opposed to "you can't be too careful". As Miller and Mitamura (2003) show, the use of "careful" can be confounding, so the modified question was done to take this into account. Additionally, the "Likert Scale" question is used as Q1 in the survey and is done on a 7 point scale. As discussed, Miller and Mitamura (2003) propose that a Likert Scale is more likely to help elicit trust, which would rule out having people agree with both answers. Doing so would involve asking the question on a 5 or 7 point scale, as opposed to having a binary option like in the WVS question

The questions that were used in this thesis were Q1 and Q6. However, a wide number of questions on trust were included. The full list of questions used around trust and perceptions were as follows:

Q1 On a scale of 1 (not at all) to 7 (trust completely), how much can other people be trusted?

Q2 On a scale of 1 (never) to 7 (always), how much do you trust people in your home community?

Q3 On a scale of 1 (low trust) to 7 (high trust), how much do you think other people trust you?

Q4 Friend/Neighbour unlikely to return wallet (1=yes)

Q5 Stranger unlikely to return wallet (1=yes)

Q6 Generally speaking, would you say most people can be trusted or that most people can't be trusted?

Q7 I have been disappointed in trusting others (1 = strongly agree or agree)

For the experimental measure of trust, the standard trust/investment game is used (Berg et al., 1995) in order to elicit trust. The game is structured as follows: there are two players, Player A (known as the trustor) and Player B (known as the trustee) that participate in the game. Player A is given R50 by the experimenter and is then asked whether they would like to send any of the money to Player B. This can be done in increments of R1 and can be any amount, including R0 and R50. They are informed that whatever they send will be doubled by the experimenter before being sent to Player B.

In the second stage, Player B has the option to send some money back to Player A, and again they can choose to send any amount in R1 increments. Player A is aware of this when they make their decision about sending to Player B. The amount that Player B sends back is then retained by Player A, at which point the game ends. As discussed previously, the socially optimal outcome of this game is for Player A to send the entire allocation and for Player B to return half. However, the sub-game perfect equilibrium dictates that Player A should send 0 as a "rational" Player B would prefer more money to less and therefore not respond.

There are several measures that are taken around these games, and these are as follows:

Sent A: the amount sent by Player A to Player B

Fraction Offered: the % of the initial endowment that A sends, calculated as "Sent A" divided by 50

Expected Return: amount that A expects B to return (measures belief of trustworthiness)

Fraction Expected Return: subjective expectations as a fraction of B's total (Player A)

Sent B the amount returned by Player B to Player A

Fraction Returned: the % returned by Player B, calculated as "Sent B" divided by "Sent A"

Amount Kept: the amount kept by Player B. This is measured as "Sent A" minus "Sent B"

The combination of survey and experimental measures of trust give a rich set of variables and data to explore. Furthermore, the various components of trust, discussed in the literature, can all be observed.

4.4.1 Trust Summary Statistics

Table 3 summarises the results of the questions that were captured during the survey regarding trust and the experimental trust measures. It is important to note that this data is from the endline survey and experiments, meaning that the treatment group had undergone exposure to the programme and the control had not. Additionally, the summary statistics include t-tests for significance of the differences between the averages of the control and the treatment groups, to indicate whether there is a significant difference between them. The full discussion around the impact of the programme is discussed in the results section, where various forms of regressions, which include control variables, are run to see the impact of the programme.

The results from survey questions are listed in the first part of the table, with the two questions that are used in the results section listed under Q1 and Q6. For **Q1**, which uses the "Likert" trust question, the average response by the control group is lower than the treatment group, with scores of 3.850 and 3.027 respectively. The difference between the groups is significant at the 10% level, indicating that the treatment group trust more than the control group at the end of the programme.

This result is confirmed in the "WVS" survey question (**Q6**), where the treatment group on average responded that they trust people more than the control group do. The table shows that 56.2% of the treatment group say that most people can be trusted while only 46.1% of the control group say most people can be trusted. Additionally, this result is significant at the 1% level when running a t-test.

This result extends to how much each group trusts people in their home community (**Q2**). The treatment group responded with an average score of 4.082 out of 7 while the control group responded with 3.918 out of 7. The difference between these scores is significant at the 10% level.

There is also a significant difference between the treatment and control group in how much they think other people trust them. In **Q3**, the control group has an average response of 5.10 while the average response of the treatment is 5.274. This is significant at the 5% level in the t-test.

The final question where there is a significant difference is **Q5**, where the participants were asked whether they thought a stranger would be unlikely to return a lost wallet to them. The results show that 62.3% of the control group and 56.2% of the treatment group responded that they didn't expect a stranger to return their wallet. This indicates that the control group trusted strangers to do the right thing *less than* the treatment group did. This result is significant at the 10% level.

There is no significant difference between the treatment and control group for **Q4** (Friend/Neighbour unlikely to return wallet) or **Q7** (I have been disappointed in trusting others). However, in both of these questions, the averages show that the treatment group is more trusting than the control group, though not significant.

Therefore, the results from the survey questions show that the treatment group is more trusting than the control group. This is the case amongst all 7 questions included here, with 5 of these differences being significant. Additionally, **Q6**, which is the most commonly used survey measure of trust shows that the treatment trust more than the control. This will be explored in more detail in the results section.

The experimental data from the trust game is also captured in Table 3. It is important to note though that the sample size for the experiment was smaller than the survey. This is because the survey includes all participants, while in the experiment half of the participants play as Player A or Player B. The first four measures capture how the "Player A's" performed in the game. There is no significant difference in the average amounts sent by the treatment and control groups, with the average offer being R22.82 and R22.27 respectively. The other measures are not significantly different either. The treatment group offered 45.6% of their endowment and the control group offered 44.5%. Additionally, the control group expected the other player to return 59.5% and the treatment group expected 61.4%.

The most interesting summary statistics from the trust game experiment concerns "Player B" in the game. The absolute amount returned by the Player Bs in the control group is higher than that of the treatment group, though this is not significant. However, the fraction returned by the control group is 25.9% and by the treatment group is 22.9%. The t-test shows a result of 0.1038, which is not significant at the 10% level but it would be significant at a slightly higher alpha.

The final summary statistic, being the amount kept by Player B's, shows that the treatment group kept more than the control group, with R65.18 and R59.90 being kept respectively. The difference between the means is significant at the 5% level, indicating that the control group is more trustworthy than the treatment group

This result is counter-intuitive. How did the treatment group become less trustworthy than the control group, given the nature of the Activate! programme? This question is answered in section 5 (Analysis and Results), where the impact of the programme on trustworthiness and the impact of depression on trustworthiness are explored in detail.

Table 3: Summary Statistics: Trust Data at Endline

	Control	Treatment	Total
Trust Survey Questions			
Q1 How much can other people be trusted? (1-7)	3.850	4.027*	3.946
	(1.434)	(1.238)	(1.334)
Q2 How much do you trust people who live in your home community? (1-7)	3.918	4.082*	4.007
	(1.448)	(1.406)	(1.427)
Q3 How much do you think other people trust you? (1-7)	5.100	5.274**	5.194
	(0.501)	(0.498)	(0.499)
Q4 Friend/Neighbour unlikely to return wallet (1=yes)	0.488	0.450	0.467
	(0.501)	(0.498)	(0.499)
Q5 Stranger unlikely to return wallet (1=yes)	0.623*	0.562	0.590
	(0.486)	(0.497)	(0.492)
Q6 Most people can be trusted (1=yes)	0.461	0.562***	0.524
	(0.499)	(0.497)	(0.500)
Q7 I have been disappointed in trusting others (1 = strongly agree or agree)	0.901	0.894	0.897
	(0.300)	(0.309)	(0.305)
Trust Game Experiment			
Amount sent (Player A)	22.27	22.82	22.58
	(11.64)	(13.11)	(12.47)
Fraction offered (Player A)	0.445	0.456	0.452
	(0.233)	(0.262)	(0.249)
Subjective expectations (Player A)	28.56	30.70	29.76
	(20.87)	(21.81)	(21.40)
Subjective expectations as a fraction of B's total (Player A)	0.595	0.614	0.606
	(0.213)	(0.221)	(0.217)
Amount returned (Player B)	22.97	21.35	22.22
	(20.68)	(20.25)	(20.46)
Fraction returned (Player B)	0.259	0.229	0.245
	(0.201)	(0.184)	(0.194)
Amount kept (Player B)	59.90**	65.18	62.33
	(20.79)	(20.56)	(20.82)
Observations	610		

1. * $p < 0.01$, ** $p < 0.05$, *** $p < 0.10$.
2. Std. errors reported in parenthesis, robust to individual clustering.

4.5 Depression Data

Diagnosing a person with depression would usually include a full psychiatric or psychological evaluation with a trained professional. Doing this within the Activate! study, which is a group of over 600 individuals, at two different measurement points would be an expensive task. Fortunately, there is a survey-based measure that is frequently used as an indicator of depression.

The Centre for Epidemiologic Studies Depression Scale (CES-D) is a short, self-reported scale that is commonly used as an indicator of depression. The 20 questions that form part of the scale were developed based on symptoms of depression and validated using longer scales and internal testing (Radloff, 1977). Participants respond to the questions on a scale of 0-3, with 0 indicating they experienced the symptom rarely and 3 indicating they experience it most of or all of the time. Since being developed, a number of shortened versions have been tested and validated. For example, a CES-D 10 scale was developed, using 10 of the 20 questions, and showed good predictive accuracy compared to the full version (Andresen et al., 1994).

The CES-D 10 scale has been investigated in a South African context. A recent study aimed to establish the validity and reliability of the scale in Zulu, Afrikaans and Xhosa groups across South Africa. Using the NIDS panel data, the study was able to demonstrate that the scale was a valid and reliable screening tool across this demographic and that there were adequate levels of consistency across samples and when compared to other depression measures (Baron et al., 2017)

Another version is the CES-D 8 where two questions from the CES-D 10 scale are excluded. These are on how hopeful and how happy the person feels. This is because both questions are answered in the positive, which may confound a participant. The CES-D 8 question scale has also been shown to be a valid and reliable instrument for screening depression (Karim et al., 2015).

Data on depression was collected using a slightly modified version of the CES-D scale. In the questionnaire, 8 questions were asked of the participants. However, these 8 questions were not the same as the ones in this CES-D 8 scale mentioned above. Instead, the CES-D 10 scale was used but two questions were excluded on the basis that it may be difficult for the participant to understand what the questions were asking, as many of them are not first language English speakers. The first excluded question was "I felt that everything I did was an effort" the second was "I could not get going". The table below summarises which questions that were used in the Activate! survey compared to the CES-D 8 and CES-D 10 scale:

It is worth noting that because the questions that were asked took language barriers into account, they would have resulted in a more accurate estimate on depression than if the full 10 questions were used. Furthermore, the two questions that are usually excluded in the CES-D 8 scale are done so because they are asked in the positive. However, these two questions are quite easy to understand, and because the answers are recorded by researchers there is a smaller probability of error. Once recorded, the two positive questions are inverted such that higher responses are given lower scores. Therefore, with the combination of simplified questions to remove the language barrier and use of researchers to record the questions, the CES-D score that is calculated from these data presents a less error-prone estimate of depression.

4.5.1 Building the CES-D Score

A minor complication in building comparable CES-D scores from the beginning and end of the programme was that the baseline and endline surveys used two different scales. The baseline survey used a 3 point scale from 1-3 while the endline survey was on the normal 4 point scale from 0-3. To take this into account, the baseline scale was adjusted down to 0-2 by subtracting one point from each answer. When the endline score is compared to the baseline it is taken on a scale of 0-3. This is the score that is used for the majority of this

Table 4: Questions used Activate! compared to the CES-D 8 and CES-D 10 scales

Question	CES-D 10	CES-D 8	Activate!
1. I was bothered by things that usually dont bother me	X	X	X
2. I had trouble keeping my mind on what I was doing	X	X	X
3. I felt depressed	X	X	X
4. I felt that everything I did was an effort	X	X	
5. I felt hopeful about the future	X		X
6. I felt fearful	X	X	X
7. My sleep was restless	X	X	X
8. I was happy	X		X
9. I felt lonely	X	X	X
10. I could not get going	X	X	

dissertation and has been labelled "End CES-D" in the results tables.

To form the total score, the results from the questions were added together. The questions on happiness and hope were inverted before being added. This is because they are responses in the positive, meaning that a lower number is reflective of greater levels of depression. The result was a baseline CES-D score between 0 and 16 and the endline CES-D score is between 0 and 24. Although the baseline scale is compressed, it is a classical measurement error meaning that the effect size will only be understated. Therefore there is no concern that compressing the scale will cause false positives (Angrist and Pischke, 2009).

4.5.2 Depression Summary Statistics

Table 5 summarises the results of the questions that were captured during the baseline and endline survey regarding depression, based on the CES-D scale mentioned previously. Additionally, for the endline score, the treatment group had undergone exposure to the programme and the control had not.

The summary statistics include t-tests for significance of the differences between the means of the control and the treatment groups, to indicate whether there is a significant difference between them. The full discussion around the impact of the programme can be found in the results section, where various forms of regressions, which include control variables, are run to see the impact of the programme on depression as well as the impact of depression on trust and trustworthiness.

The results from the baseline survey questions are listed in the first part of the table, with the CES-D total listed first and the individual questions listed thereafter. For the total baseline CES-D score there is no significant difference between the control and treatment, with the average being 4.335 and 4.310 respectively. This extends to each of the questions, where there is no significant difference between the treatment and control for any of the eight individual questions

It is interesting to note how similar the means of the two groups are in each of the questions. For example, in the first question, the average response is 0.632 for the control group and 0.639 for the treatment group. For the "I felt depressed" response, the control group average was 0.527 compared to the treatment group average of 0.528. And the response on "I was happy" had an average score of 1.465 and 1.467 for the control group and treatment group respectively. Based on this, there is evidence that there was no difference in the

level of depression between the control and treatment groups, in aggregate or at an individual question level, at the start of the programme.

The second part of table 5 shows the summary statistics of the endline CES-D score. Again it is important to note that these results are on a different scale to the baseline, as the baseline survey asked the questions on a scale of 0-2 while the endline asked the questions on a scale of 0-3.

The results at endline shows that the treatment group has a lower CES-D score than the control group. However, this difference is not significant at the 10% level. In the individual questions, almost all of them show that the treatment responded with less evidence of depression than the control, except for the question around sleep restlessness. However, the only response where there was a significant difference between the treatment and control was for "I felt hopeful about the future", where the mean for the treatment group was 2.460 and the mean for the control group was 2.351.

Based on this, there is some evidence that the programme had a positive effect on the mental health of the participants. This is explored in more detail in the results section, where various regressions are run using control and interaction variables to examine this impact.

Table 5: Summary Statistics: Depression Data

	Control	Treatment	Total
Baseline CES-D Score	**4.335**	**4.310**	**4.321**
	(2.463)	(2.253)	(2.349)
I was bothered by things that usually don't bother me	0.632	0.639	0.635
	(0.605)	(0.587)	(0.595)
I had trouble keeping my mind on what I was doing	0.556	0.526	0.540
	(0.591)	(0.574)	(0.581)
I felt depressed	0.527	0.528	0.528
	(0.607)	(0.550)	(0.577)
I felt hopeful about the future	1.603	1.571	1.586
	(0.660)	(0.632)	(0.645)
I felt fearful	0.601	0.604	0.603
	(0.530)	(0.560)	(0.546)
My sleep was restless	0.533	0.502	0.516
	(0.631)	(0.625)	(0.628)
I was happy	1.465	1.467	1.466
	(0.544)	(0.556)	(0.550)
I felt lonely	0.593	0.581	0.586
	(0.593)	(0.575)	(0.583)
Endline CES-D Score	**6.255**	**6.031**	**6.135**
	(4.077)	(3.957)	(4.011)
I was bothered by things that usually don't bother me	0.925	0.869	0.895
	(0.919)	(0.840)	(0.877)
I had trouble keeping my mind on what I was doing	0.871	0.798	0.831
	(0.928)	(0.857)	(0.890)
I felt depressed	0.668	0.631	0.648
	(0.838)	(0.857)	(0.848)
I felt hopeful about the future	2.351	2.460*	2.410
	(0.948)	(0.866)	(0.905)
I felt fearful	0.786	0.827	0.808
	(0.814)	(0.891)	(0.856)
My sleep was restless	0.743	0.802	0.775
	(0.899)	(0.944)	(0.923)
I was happy	2.089	2.134	2.113
	(0.873)	(0.819)	(0.844)
I felt lonely	0.682	0.736	0.711
	(0.873)	(0.845)	(0.858)
Observations	610		

1. * $p < 0.01$, ** $p < 0.05$, *** $p < 0.10$.
2. Std. errors reported in parenthesis, robust to individual clustering.

4.5.3 Correlates of Baseline Depression in the Activate! Data

Using the baseline survey, a set of regressions were run to observe which of the variables were correlated with depression at baseline. Based on the literature, several variables were chosen, such as gender, unemployment, exercise and health. These regressions can be seen in **Table A.1**, where eight regressions are run with each of the variables being introduced one at a time.

Gender is used as the first variable in model (1). The variable is a dummy variable with 1 representing male participants. The coefficient is negative but insignificant, meaning that there is no significant difference in depression between men and women in the study. There is significant evidence in the literature that has shown that women have higher reported levels of depression than men (Das et al., 2007). However, it has also been shown that men are less likely to report depression and symptoms of depression and that they are less likely to seek treatment for mental health problems (Seidler et al., 2016)

The next variable to be introduced in the model is unemployment at baseline. Across all the models where it is included, unemployment is not significantly correlated to depression though the coefficient is positive. However, in the literature, it has been shown that people with depression were more likely to become unemployed, have a higher job turnover and report higher absenteeism rates (Lerner et al., 2004).

The next regression in the table is run with the exercise variable added. The exercise variable is a dummy variable with 1 representing a response of the person exercising 3 times or more per week and 0 indicating all values lower than that. The result shows that there is a significant, negative correlation between depression and exercise. This means that higher levels of exercise are associated with lower levels of depression. The literature supports this relationship between depression and exercise. Evidence across multiple studies has shown the bi-directional relationship of depression and physical exercise, showing that low amounts of exercise can be protective against depression and that depression reduces the likelihood that a person will exercise (Teychenne et al., 2008). Salmon (2001) shows that exercise has successfully been used as a psychological treatment for depression and North et al. (1990) use evidence from a comprehensive review of the literature to show that exercise has an antidepressant effect.

Present health is introduced in the fourth regression model onwards. The variable was constructed using the question "On the scale of 1 (poor) to 5 (excellent), how would you describe your health at present?". This was then turned in to a dummy variable, with 1 being "excellent" and 0 being all other responses. The table shows an inverse relationship between present health and depression, which is significant at the 1% level across 4 regressions where the variable is included. In other words, lower levels of current health are associated with higher levels of depression. This is consistent with the depression literature, where poor mental health at a national level has been shown to predict lower reported health levels (Ardington and Case, 2010). On a micro level, a study of patients at a community clinic in Cape Town showed lower levels of overall health with higher levels of depression (Triant, 2002).

After this, a variable for smoking cigarettes is added to the model, with 1 indicating that a person smokes and 0 indicating that they don't. In the group of participants, there were 46 smokers and 342 non-smokers. From table Table A.1, the regression models show that smoking is directly and significantly correlated with depression. The relationship is large, with smoking predicting an approximately 8% higher CES-D score. This relationship has been seen in the literature, with people with a history of depression being twice as likely to smoke than those who have never suffered from depression (Mendelsohn, 2012). This may be causal, with smoking being used as self-medication to elevate mood, though it has also been shown that shared genetic and family factors predispose people to both depression and smoking (Kendler et al., 1993).

The "Staying in Neighbourhood" variable was constructed using the question "Is there a high or very high

chance of staying in neighbourhood (1=yes)?" Of the responses recorded, 34% of participants said there was a high chance that they'd stay in their neighbourhood and 66% said there wasn't. In the regression model, there is a significant relationship showing that wanting to leave the neighbourhood is associated with higher depression levels. This could be for a number of reasons, such as having feelings of wanting to escape or being in an environment that negatively impacts mental health. There is evidence of neighbourhood effects on mental health, where parents who moved to low poverty areas report less distress and boys who moved reporting fewer depressive and anxious symptoms than those who stay (Leventhal and Brooks-Gunn, 2003).

Computer literacy is added in regression (8). The coefficient is negative and significant at the 5% level, indicating that lower levels of depression are associated with computer literacy. A search of the literature reveals little evidence of the relationship between computer literacy and depression, though it might have an effect through other channels of disadvantage, such as through neighbourhood effects, inequality and income levels (Ross, 2000).

The last control variable is race, used with 1 indicating black African South Africans. Out of the 388 participants in the study, there were 374 black African, 14 coloured and no white individuals. From the regression table, the race variable is positive but isn't significant. It has, however, been shown that black African individuals in South Africa have higher rates of depression than white individuals, with coloured/Indian individuals falling in between the two (Adjaye-Gbewonyo et al., 2016)

These regressions show the correlation between depression and a number of other characteristics. There is significant literature that supports the relationships that have been observed. Further analysis of the endline depression with other variables is discussed in section 5.

5 Analysis and Results

This section examines the effect of the Activate! programme on depression, trust and trustworthiness. Additionally, the relationship between trust and depression is explored, followed by an analysis of the causal impact of depression on trustworthiness and trust.

Two types of regression models were used to analyse and explore the various relationships and causal effects in the data. The first is a set of ordinary least squares regression models that are used to predict the value of the dependent variable from multiple independent variables. The second type of model used is a two-stage least squares regression. This has been done to estimate the impact of depression on trust and trustworthiness. This section describes these empirical techniques and how they were applied to the Activate! data as well as the results from using these techniques.

5.1 Impact of the Activate! Programme

The first component of this analysis examines the effect of the Activate! programme on a number of variables. These are depression, survey trust, experimental trust and experimental trustworthiness. These regressions are shown in the tables contained in each section, namely Table 6, Table 7, Table 8 and Table 9. There are also tables included at the end of this paper that include the same regressions with added controls. The use of the treatment variable in these regressions means that causation can be inferred between the effects of the treatment on the dependent variables. This is because randomisation has been used in the data collection, as described in subsection 4.2. The key results from the regressions are:

Depression is reduced for men who participate in the programme

Survey trust (WVS) increases for the treatment group

Experimental trust (Player A send) increases for the treatment group when controlling for Player A's expected return

Trustworthiness decreases in men who participate in the programme

The first three key results are expected from the programme. The group therapy nature of Activate! is hypothesised to reduce depression in men, as per the paper by Seidler et al. (2016). Although this was not the direct intention of the programme, the direction and gendered nature of the results can be understood from the literature. Additionally, the effect of the programme on both the survey and experimental measures of trust is expected. Activate! targets various components of a person's belief mechanisms, by influencing their perceptions of self and others, with the goal of improving how they trust. Therefore, the individuals that underwent the treatment are expected to trust more than those that weren't part of the programme.

However, the effect of the programme on reducing trustworthiness of men is unexpected. This is because the programme is not hypothesised to affect this behavioural measure. Trustworthiness is a component of intention that operates on belief mechanisms that sit outside of the domain of the self-actualisation elements of the training and therefore should not be affected by Activate!. Additionally, there is little literature to suggest that a programme such as this would negatively impact trustworthiness. This is discussed in more detail in section 5.2 and section 5.3, where the effect of depression on trustworthiness is examined using various statistical techniques.

5.1.1 Programme Impact on Depression

In order to have an unbiased estimate of the impact of the programme on depression and trust we would need to know what would have happened to the participants had they not been in the programme. As discussed

previously, a pipeline randomisation was used on the Activate! programme, meaning that the control group can be used as our counterfactual to assess the impact.

The Activate! programme had modules that targeted several areas but had no specific intervention for depression. Despite this, the programme may have improved depression outcomes through other mechanisms, as discussed in section 3. To investigate whether the programme had an impact on depression, several least squares regressions were run with the endline CES-D as the dependent variable. The endline CES-D score, which ranges from 0 to 24, is based on the 8 questions that were asked on the endline survey. This variable is then regressed on the treatment variable, gender variable baseline CES-D and a treatment interaction term. The interaction term is between treatment and gender which shows the impact of the programme on male participants only.

Furthermore, controls are added for present health, race, frequency of exercise and whether an individual smokes cigarettes, which are known to be correlated with depression. **Table 6** shows the regressions without the introduction of control variables while **Table A.2** shows all the regressions that were run with the control variables being added in each subsequent regression. The key result from these regressions is that the treatment is shown to reduce depression levels for men.

Table 6: Regressions showing the Impact of the Activate! programme on Endline Depression

	(1)	(2)	(3)	(4)
	End CES-D	End CES-D	End CES-D	End CES-D
Treatment	-0.224	-0.134	-0.141	0.792
	(0.331)	(0.347)	(0.326)	(0.497)
Gender (1=Male)		-0.231	0.0307	0.892*
		(0.350)	(0.329)	(0.477)
Base CES-D			0.576***	0.570***
			(0.0687)	(0.0684)
Treatment x Gender				-1.626**
				(0.656)
Constant	6.255***	6.403***	3.744***	3.303***
	(0.242)	(0.318)	(0.428)	(0.462)
Observations	593	547	529	529

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

The first regression (1) in **Table 6** is a simple linear regression model which uses the treatment dummy as the independent variable. This variable equals 1 if a participant was part of the treatment and 0 if they were part of the control. The results of the regression show a negative relationship between the treatment and depression, though this is not significant.

The next regression model (2) introduces the gender variable, in which 1 indicates male and 0 indicates female. Again, this is not significant and the treatment variable is unaffected.

In model (3) a baseline CES-D score is added as a control variable to control for depression from the start of the programme. The coefficient on this is highly significant, at the 1% level, and it has a positive

relationship with the dependent variable. This can be understood as the impact of a person's previous level of depression on their current level of depression. It shows that having depression at the start of the programme is predictive of having depression at the end of it. The magnitude of this coefficient is also of importance, as it indicates that an increase of one point on the CES-D scale at baseline is predictive of just over half a point at endline. It must be noted though that the baseline CES-D score is on a scale of 0-16 as opposed to 0-24 at endline, meaning that the actual magnitude of the coefficient is likely to be slightly smaller if they were on the same scale. Additionally, in model (3) the treatment variable is still not significant.

In the fourth regression model (4), a treatment gender interaction variable is added. This interaction term is created by multiplying the gender variable by the treatment variable. When this interaction term equals 1 it can be interpreted as the effect of the treatment on male participants in the programme. From the table, the coefficient on the interaction term is significant and negative, reducing the predicted depression level by -1,626. The treatment variable remains insignificant in this regression model.

However, the gender variable is also significant, and therefore needs to be taken into account. The effect for a male participant on the programme is, therefore, the difference between the interaction term and the gender variable. This gives a net effect of a reduction of 0,734 on the CES-D scale. Therefore, the programme is predicted to have reduced the depression of the male participants by 3,06% when compared to the controls.

In **Table A.2** several other control variables are added to explore whether they have an effect on the regressions. The first four models in this table are the same as **Table 6** and the control variables are added in models (5) to (8). The controls that are used are dummy variables for health, race, exercise and smoking. These variables were chosen as there is evidence in the literature that they might affect depression. Additionally, present health, smoking and frequency of exercise are shown to be correlated with baseline depression in **Table A.1**. In all 4 models where the control variables are used, the findings were robust to the inclusion of the controls.

Model (10) in **Table A.2** contains all the variables and shows that, once the controls have been added, the net impact of the programme on depression is 0,703 or 2,93%. Therefore, after controlling for a number of variables there is still a significant, positive effect of the programme on depression in men.

5.1.2 Programme Impact on Survey Trust

Using various regression models, I examined the impact of the programme on one of the measures of survey trust, the "WVS Trust Question". The questions states "Generally speaking, would you say most people can be trusted or that most people cant be trusted?" with 1 indicating that people can be trusted and 0 indicating people can't be trusted. This variable is then regressed on the treatment variable, gender variable, baseline CES-D and race variable.

Furthermore, controls are added for risk, altruistic behaviour and expected return. **Table 7** shows the regressions with these control variables added individually while **Table A.3** shows various combinations of these measures being added. The key result from these regressions is that programme positively impacted trust as measured by the WVS survey question, with a significant increase for individuals that participated in the programme.

Table 7: Regressions showing the Impact of the Activate! Programme on the WVS Trust Question

	(1) Q6	(2) Q6	(3) Q6	(4) Q6	(5) Q6
Treatment	0.117***	0.110**	0.127***	0.119**	0.127*
	(0.0404)	(0.0425)	(0.0444)	(0.0462)	(0.0665)
Base CES-D		-0.0177**	-0.0190**	-0.0133	-0.0207
		(0.00901)	(0.00932)	(0.00968)	(0.0143)
Gender (1=Male)		0.116***	0.124***	0.106**	0.102
		(0.0429)	(0.0445)	(0.0466)	(0.0676)
Race (1=black African)		-0.235*	-0.227*	-0.187	-0.165
		(0.121)	(0.121)	(0.134)	(0.161)
Mean Row (Risk)			-0.00193		
			(0.0124)		
Dictator Game Offer				-0.215*	
				(0.117)	
Expected Return					0.182
					(0.151)
Constant	0.461***	0.694***	0.693***	0.700***	0.517***
	(0.0297)	(0.130)	(0.142)	(0.147)	(0.195)
Observations	609	542	501	457	230

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Q6. Would you say most people can be trusted or that most people can't be trusted?

CES-D: a short, self reported scale used as an indicator of depression

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

The first regression (1) in **Table 7** is a simple linear regression model which uses the treatment dummy as the independent variable. The results show that the treatment has a positive and significant effect on the WVS trust measure. This indicates that, according to this measure, the Activate! programme increases the degree to which the participants trust.

In model (2) the baseline CES-D score, gender and race variables are added to control for the depression and demographic factors. The coefficient on baseline CES-D is significant at the 5% level and it has a negative relationship with the dependent variable. This negative relationship between the CES-D score and WVS trust is explored in more detail in section 5.2. The gender variable is significant at the 1% level and shows that men responded that they trusted more than the women did. Finally, the race variable is significant and negative, showing that there is a difference between the racial groups in the study in responding to the WVS question. With these controls added, the treatment variable is still significant at the 5% level.

Additional controls are included in regression (3) to (5) in **Table 7**, with various combinations of these controls used in regressions (2) to (8) in **Table A.3**. This has been done to control for other factors that may be influencing trust. As described in Fehr (2009), risk preferences, social preferences (e.g. altruism) and expected return all have an influence on trust. Therefore, I use three variables to control for the effect of these influences. The first is mean switching row in the Holt-Laury multiple price lists where individuals are asked to choose 15 times between two options, A and B, where an extremely risk-averse individual would always choose option A and an extreme risk-seeking individual would always choose B. In between this, it

is expected that participants will change from A to B with later switching indicating a higher degree of risk aversion (Sapienza et al., 2013). The second is the dictator game offer, which can be used as a measure of altruistic behaviour. The third control is Player A's expected return, which is measured as the percentage that Player A expects Player B to return.

The risk measure (mean row) is added in regression (3). The coefficient on it is negative but is not significant at the 10% level. As a result of adding the risk measure, the significance and magnitude of the treatment coefficient increases, with the coefficient being 0.127 significant at the 1% level. The magnitude and significance of the other control variables remain consistent with the previous regression.

In regression (4) the dictator game offer is included to control for altruistic behaviour. The coefficient on it is negative and is significant at the 10% level. This is a particularly interesting result, as it indicates that a higher offer in the dictator game, meaning a higher level of altruistic behaviour, is negatively related to trust as measured by the survey question. Examining the effect of adding this control on treatment, we see that the significance and magnitude of the treatment remain similar to that of regression (2). The magnitude and significance of the other control variables change slightly, with the base CES-D and race variables no longer being significant.

The final regression (5) in **Table 7** introduces the expected return of Player A as a control. It is important to note that the expected return variable only measures half the participants, as the other half of the participants act as Player B. This results in the sample size and, therefore, the power of the regression being reduced. Despite this, the coefficient on the treatment variable is still significant at the 10% level. The other variables have similar coefficients to the other regressions but are no longer significant. The expected return variable is not significant, but is positive, indicating that there may be a positive relationship between how much a person expects to receive back and how much they trust

The results show that, even when controlling for various combinations of risk, altruism and expected return, the Activate! programme has a significant and positive impact on the WVS trust variable of programme participants. This outcome is expected, as the programme specifically targeted the beliefs of individuals.

5.1.3 Programme Impact on Experimental Trust

The next step in understanding the impact of the Activate! programme on trust is to investigate the impact it had on trusting behaviour. This is measured by the percentage of the initial endowment that Player A sent in the trust game. It is important to note that the number of participants in this regression is lower than the other regressions, as only half of the treatment and control groups are assigned to being Player A. Regressions are run using the % sent as the dependent variable, with the independent variable being the treatment dummy variable. Three control variables are then added, which are the gender variable, baseline CES-D and race variable.

Controls are also added for risk, altruistic behaviour and expected return. **Table 8** shows the regressions with these control variables added individually while **Table A.4** shows various combinations of these measures being added. The key result from these regressions is that the programme increased trusting behaviour, as measured by the percentage sent by Player A when controlling for Player A's expected return as well as when controlling for altruistic behaviour.

The first regression (1) in **Table 8** is a linear regression model which uses the treatment dummy as the independent variable. Additionally, baseline depression, gender and race are added as controls. The result shows that there is a positive impact of the treatment on the amount sent by Player A but that this is not significant.

In model (2) the baseline CES-D score, gender and race variables are added to control for the depression

and demographic factors. The coefficient on baseline CES-D is positive but is not significant. Interestingly, there is a significant difference between the amount sent by the men and women, with men sending a higher amount than the women. The gender variable is significant at the 5% level and shows that men offered approximately 6.8% more than the women in the trust game. Finally, the race variable is positive but not significant. With these controls added, the treatment variable is still not significant at the 10% level.

Table 8: Regressions showing the Impact of the Activate! Programme on % Sent of Player A (trust game)

	(1)	(2)	(3)	(4)	(5)
	% Sent	% Sent	% Sent	% Sent	% Sent
Treatment	0.0109	0.0348	0.0327	0.0443	0.0515*
	(0.0296)	(0.0310)	(0.0324)	(0.0331)	(0.0307)
Base CES-D		0.00460	0.00490	0.00627	0.00604
		(0.00680)	(0.00691)	(0.00731)	(0.00660)
Gender (1=Male)		0.0677**	0.0676**	0.0753**	0.0611*
		(0.0313)	(0.0317)	(0.0334)	(0.0312)
Race (1=black African)		0.0956	0.0943	0.0759	0.0885
		(0.0754)	(0.0760)	(0.0805)	(0.0743)
Mean Row (Risk)			0.00357		
			(0.00950)		
Dictator Game Offer				0.0598	
				(0.0802)	
Expected Return					-0.0779
					(0.0698)
Constant	0.445***	0.277***	0.262***	0.268***	0.341***
	(0.0223)	(0.0836)	(0.0930)	(0.0934)	(0.0899)
Observations	289	256	253	228	230

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

% Sent: the % that Player A sends in the trust game

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

Additional controls are included in regression (3) to (5) in **Table 8**, with various combinations of these controls used in regressions (2) to (8) in **Table A.4**. As described in Fehr (2009), risk preferences, social preferences (e.g. altruism) and expected return all have an influence on trust. Therefore, these three variables are used to control for the effect of these on the amount sent by Player A. The first is mean switching row in the risk game, with later switching indicating a higher degree of risk aversion. The second is the dictator game offer, which can be used as a measure of altruistic behaviour. The third control is Player A's expected return, which is measured as the percentage that Player A expects Player B to return.

The control that is introduced in regression (3) is the risk measure (mean row). The coefficient on it is negative but is not significant at the 10% level. As a result of adding the risk measure, the magnitude of the treatment coefficient increases but remained insignificant. The magnitude and significance of the other control variables remain consistent with the previous regression.

In regression (4) the dictator game offer is included to control for altruistic behaviour. The coefficient

on it is positive but not significant at the 10% level. Adding this control treatment does not change the significance of the treatment variable or the other variables, though the magnitudes change slightly.

The final regression (5) in **Table 7** introduces the expected return of Player A as a control. The introduction of the expected return variable results in the coefficient on the treatment variable being significant at the 10% level. The variable is positive and shows that the treatment group sent approximately 5.2% more than the control group, once expected return is included. The other variables have similar coefficients and significance levels as the other regressions. The expected return variable is not significant but is negative.

In all the regressions in **Table A.4** the treatment variable is positive, but in (4) and (7) it is significant at the 0.10 level. In both of these regressions expected return is used as a control variable. This indicates that the programme had a positive impact on the behavioural trust measure of the participants when controlling for their expected return.

5.1.4 Programme Impact on Trustworthiness

To explore the impact of the Activate! programme on the amount returned by Player B (herein referred to as trustworthiness), a set of regressions was run. The results from these regressions are shown in **Table 9**. The dependent variable used in the regressions is the percentage returned by Player Bs in the trust game. This is calculated as the amount they sent back to Player A divided by the amount that was sent to them by Player A. This variable is then regressed on the treatment variable, control variables and several treatment interaction terms.

Table 9: Regressions showing the Impact of the Activate! Programme on Return of Player B (trust game)

	(1)	(2)	(3)	(4)	(5)
	% Returned	% Returned	% Returned	% Returned	% Returned
Treatment	-0.0295	-0.0307	-0.0307	0.0350	0.0370
	(0.0233)	(0.0235)	(0.0236)	(0.0349)	(0.0349)
Base CES-D		0.00159	0.00166	0.00110	0.000872
		(0.00470)	(0.00473)	(0.00469)	(0.00469)
Gender (1=Male)			0.00401	0.0582*	0.0590*
			(0.0237)	(0.0317)	(0.0317)
Treatment x Gender				-0.119**	-0.122***
				(0.0470)	(0.0470)
Race (1=black African)					0.0866
					(0.0767)
Constant	0.259***	0.249***	0.247***	0.219***	0.135*
	(0.0158)	(0.0263)	(0.0303)	(0.0318)	(0.0811)
Observations	276	254	254	254	254

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
CES-D: a short, self reported scale used as an indicator of depression
% Returned: the fraction returned by Player B in the trust game

The first regression (1) is a simple linear regression model which uses the treatment dummy as the independent variable. The results of the regression show a negative relationship between the treatment and trustworthiness, though this is not significant.

The next regression model (2) introduces baseline CES-D as a control variable. The reason this is introduced is that in a number of instrumental variable regression models, discussed in subsection 5.2, depression is shown to increase the amount returned by Player B. It is therefore important that this is incorporated

into the model. This is explored in more detail in subsection 5.2.

In the third regression model (3), the gender dummy variable is added. This variable is not significant in this regression and does not affect the significance or the other variables.

The "Treatment x Gender" interaction term is added in regression (4). The coefficient of the interaction term is negative and significant at the 5% level. The gender variable is positive and significant at the 10% level, meaning that the net effect is that men who were part of the Activate! programme are predicted to return 6.08% less of the amount that was offered to them. This can be interpreted as the programme having a negative impact on the trustworthiness of men.

However, this may be the result of a separate mechanism, as the programme was not targeted or theorised to have a negative impact on trustworthiness and there is little literature on a programme such as Activate! reducing trustworthiness in men. Further on in this dissertation, in subsection 5.3, depression is shown to increase trustworthiness. This means that because the programme reduced male depression, as seen in section 5.1.1, the reduction of depression would cause a decrease in the trustworthiness of men, which produces the result seen here, where men who are treated show a decrease in trustworthiness. This is discussed further in subsection 5.3.

The final regression (5) incorporate race as a control variable. In this model, the "Treatment x Gender" interaction term is significant at the 1% level, while the gender variable is still significant at the 10% level. The net result of this is that the men that participated in the programme returned 6.30% less in the trust game than the other players. Additionally, it means that the men in the control group returned 5.9% more in the trust game. The mechanisms and causes of this are explored in section 5.3.

5.2 Relationship between Depression and Trust

In subsection 5.1 we saw that the Activate! programme had an impact on depression for male participants. Furthermore, the programme was shown to have impacted various measures of trust and trustworthiness. We can take this a step further and investigate the relationship between depression and trust.

The first step in doing this is to analyse the relationship between endline depression and the measures of trust. This is done by running a least squares multiple regression using the trust measures as dependent variables. The results from this show significant evidence that higher levels of depression are correlated with reduced trust, measured through the trust survey questions. However, we are unable to draw conclusions about the causal pathways here, as these regressions show relationships between the variables, not the impact of one on the other. When using the treatment variable, we were able to draw causality as a result of the randomisation process that took place. However, we do not have this for the endline depression variable

Instead, another strategy was used to estimate the impact of depression on trust. A number of two-stage least square regressions were run using treatment and baseline CES-D scores as instrumental variables. Details of this are discussed in subsection 5.3.

The results of the regressions between the trust variables and depression are shown in **Table 10** below. The detailed tables that include various controls are **Table A.7**, **Table A.8**, **Table A.9** and **Table A.10**, and these are in the appendix at the end of this paper. As highlighted, these regressions show the relationship or correlation between the variables which means that causation cannot be inferred. The key results from the regressions are:

Higher Endline CES-D scores are significantly and negatively correlated with WVS trust with an approximately 2-3% decrease in WVS trust per point on the CES-D scale, depending on the controls used, significant at the 1% level

Higher Endline CES-D scores are significantly correlated with lower Likert trust with baseline CES-D also being negatively correlated to Likert trust

Higher Endline CES-D scores are negatively correlated with experimental trust (Player A send), significant at the 10% level when controlling for expected return

Trustworthiness (% returned) and expected return are not significantly correlated with Endline CES-D, but the coefficient is positive on both variables

The results of these regressions show that depression is negatively associated with both survey and experimental measures of trust. However higher levels of depression are not shown to be significantly correlated to Player B's return in the trust game, which may be a result of the size of the sample set. These results are discussed in detail in the sections below, with a deeper analysis conducted on the causal nature of these relationships in subsection 5.3.

Table 10: Summary of regressions between endline depression and various trust measures

	(1) Q6	(2) Q6	(3) Q1	(4) Q1	(5) % Sent	(6) % Sent	(7) % Return	(8) % Return
End CES-D	-0.0228***	-0.0198***	-0.0643***	-0.0559***	-0.00174	-0.00683*	-0.000178	0.00138
	(0.00504)	(0.00571)	(0.0144)	(0.0162)	(0.00353)	(0.00399)	(0.00311)	(0.00331)
Treatment		0.105**		0.155		0.0466		-0.0292
		(0.0426)		(0.121)		(0.0308)		(0.0243)
Base CES-D		-0.00657		-0.0472*		0.0124*		0.00130
		(0.00957)		(0.0271)		(0.00716)		(0.00512)
Gender(1=M)		0.122***		0.380***		0.0536*		0.00170
		(0.0430)		(0.122)		(0.0314)		(0.0243)
Exp. Return						-0.0749		
						(0.0713)		
Constant	0.662***	0.539***	4.408***	4.256***	0.462***	0.447***	0.248***	0.241***
	(0.0369)	(0.0600)	(0.106)	(0.170)	(0.0261)	(0.0604)	(0.0227)	(0.0334)
Observations	593	529	593	529	283	224	267	247

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

Q6. Generally speaking, would you say most people can be trusted or that most people can't be trusted?

Q1. On a scale of 1 (not at all) to 7 (trust completely), how much can other people be trusted?

% Sent: the % that A sends in the trust game

% Return: the fraction returned by Player B in the trust game

5.2.1 Regressions of Depression on Survey Measures of Trust

In the first regression (1) in **Table 10**, endline CES-D is regressed against the WVS trust question. The coefficient on endline CES-D is highly significant and shows that every 1 point increase on the CES-D scale is associated with a 0,0228 or approximately 2.3% decrease on the WVS question. This coefficient remains negative and significant at the 1% level in regression (2) when treatment, base CES-D and the gender dummy variable are added. When additional controls are introduced in **Table A.7**, the coefficient is still highly significant and negative.

The third regression (3) in **Table 10** uses the Likert trust question as the dependent variable, which is also regressed against the endline CES-D score. From the table, the coefficient is significant and negative.

The coefficient remains negative in regression (3) and significant at the 1% level when treatment, base CES-D and the gender dummy variable are added. When additional controls are introduced in **Table A.8**, the coefficient is still highly significant and negative.

Taken in combination, it is clear across the two survey measures of trust used here that higher levels of depression are related to lower responses in these questions. It is also interesting to note the gender difference here, with the coefficient on the male dummy variable being positive and significant at the 1% level.

5.2.2 Regressions of Depression on the Trust Game

Moving on to the trust game, the fifth column (5) in **Table 10** shows a regression using the percentage sent by Player As in the trust game as the dependent variable and depression as the independent variable. The coefficient on endline CES-D is not significant in this first regression. However, in regression (6), when controls including expected return and treatment are introduced, the coefficient on endline depression is negative and significant at the 10% level. This result appears in Table A.9 only when expected return is added as a control. Additionally, the relationship between gender and the % sent is positive, as it was in the survey trust measures, indicating that male participants are predicted to show higher levels of trust than female participants.

The sixth regression (6) uses the percentage returned by Player B (% return) to Player A as the dependent variable and endline CES-D as the independent variable. The coefficient is not significant at the 10% level. When controls are introduced, including treatment, baseline CES-D and gender, the coefficient shows a positive relationship between % returned and endline depression, though this is also not significant.

Based on these regressions, there appears to be a negative relationship between endline depression and the various measures of trust. There is no significant correlation between trustworthiness and depression, but this might be due to the low power of the regressions and the potential endogeneity in the model. In order to explore the causal nature of these regressions and remove endogeneity, a more advanced statistical technique, the two-stage least squares regression model, is used in the next section.

5.3 Two Stage Least Squares Regression Model

From the regression models, there was evidence that there is a strong relationship between endline depression and the various tenets of trust, with a potentially interesting effect on trustworthiness. However, we cannot draw conclusions about the causal nature of these relationships from the regressions. This is because there is endogeneity in the models, meaning that there may be unobserved factors that influence both depression and the trust variables. There may also be reverse causation, with the dependent variables of trust causing depression. The end effect is that this causes biased results.

As discussed in subsection 5.1, the Activate! programme was designed to target change the preferences and beliefs of the participants, in order to improve their active social capital. A specific component of this that was targeted was the trust of the participants, and the effect of this was seen in subsection 5.1, where trust increased for the treatment group. Additionally, the group therapy nature of the programme exogenously varied the depression of men in the treatment, as can be seen in subsubsection 5.1.1. This is seen in the literature, as Brownhill et al. (2002) show that a programme of this nature would have the effect of reducing depression in men, as depressed men respond well to group activities as a form of therapy.

However, in Table 9 there was a result that was not intended in the programme, as the trustworthiness of men in the treatment group ***decreased***. There was no specific intervention that would have caused this, particularly as the programme wasn't designed to target trustworthiness, much less reduce it in the

participants.

Some of the research discussed in the literature presents strong evidence that there is a negative correlation between trust and depression (e.g. Zak and Fakhar (2006) using the WVS question). There have been attempts to try to establish the direction of this, for example, Mellick (2014) played the trust game with 38 depressed patients and 38 healthy controls. Ong et al. (2017) played the trust game with people with a history of mood disorders alongside healthy controls and showed that formerly depressed individuals were more trustworthy in the game than their healthy counterparts. However, in both cases, depression is not exogenously varied and therefore does not establish causality. In other words, a third, unobserved factor may be correlated with both depression and the current behaviour in the trust game. Therefore, there is potentially endogeneity and reverse causation present in the model.

A method that can be employed to correct for this endogeneity is the use of instrumental variables. An instrumental variable is associated with the independent variable but not the dependent variable, which means it can correct for biased results. It is important to note that, when selecting an instrumental variable, it needs to be correlated with the regressor, uncorrelated with the error term and should not be a direct cause of the dependent variable (Wooldridge, 2009).

Once the appropriate IVs have been selected, we can use two-stage least stage regressions (2SLS) to estimate the causal relationships. The first stage of a 2SLS is to run a regression, using the chosen instrumental variables, to find the fitted value of the independent variable. This results in an estimation, or fitted values, of the independent variable, which means that this variable is assumed to no longer be endogenous. These fitted values are then regressed against the dependent variable using OLS, and this is known as the second stage regression (Wooldridge, 2009). It is important to note the terminology that is used for each of the variables in this process. The dependent variable and the independent variable that is being instrumented are known as the endogenous variables as there may be reverse causality and observably affecting both of them. To treat the independent variable as exogenous is to instrument it, meaning that it is replaced with fitted values that are estimated in the first stage using the instrumental variables. (Angrist and Pischke, 2009).

The 2SLS method requires a selection of IVs that are used to find the fitted values of the variable in the first stage. Each one of the instruments that have been chosen should, in theory, be related to the endline depression variable while not directly affecting the trust variables. The following IVs were chosen to instrument for endline depression:

Treatment: with 1 being the treatment, 0 being the control

Baseline CES-D: measure of depression at baseline

Gender: with 1 being men, 0 being women

Treatment x Gender: interaction variable where 1 indicates treated men

These four variables are the same as the ones used in **Table 6** regression (4), where the programme was shown to have affected depression in men. This is, therefore, the first stage of the 2sls estimate of the effect of depression on the trust and trustworthiness variables. Additionally, the treatment and treatment-gender interaction variables are used as instrumental variables for two reasons. Firstly, the random assignment of the treatment and control groups means that the endline CES-D score was varied exogenously and not as a result of unobservable factors. Secondly, the treatment-gender variable is shown in Table 6 to have an exogenous and significant effect on endline depression. Furthermore, the baselines CES-D score is shown to be a strong predictor of endline depression and is therefore included to improve the fitted values in the first stage of the 2sls regression.

The 2sls regressions were run on three different measures, namely WVS trust, Likert trust and trustworthiness. The experimental trust measure is not used in the 2sls as the programme directly targeted this behaviour. On the other hand, WVS and Likert trust reflect a measure of trust that includes various beliefs which may have been influenced by depression. Additionally, tests for endogeneity and tests of over-identifying restrictions were done on the 2sls regressions that were conducted on trustworthiness and survey trust, with the tests not being rejected.

The first and second stages of these regressions are shown in the tables at the end of this paper, with the second stages seen in **Table A.11**, **Table A.12** and **Table A.13**. The use of the instrumental variables in these regressions means that causation can be inferred between endline depression and the dependent variables. The key results from the second stage of the 2sls regressions are:

WVS Trust: depression is shown to reduce trust measured by the WVS question

Likert trust depression reduces Likert trust, significant at the 1% level

Trustworthiness: an increase in depression *increases* the amount returned by Player B

The results of the 2sls regressions show that an increase in depression has a negative impact on both survey measures of trust and, counter-intuitively, **increases** the trustworthiness of a person. These results are discussed in detail in the sections below.

5.3.1 Impact of Depression on Survey-Based Trust Measures

The first set of 2sls regressions were run to investigate the effect of endline depression on the "**WVS Trust Question**". Controls are then added, which include health, computer literacy and employment. The first stages of the 2sls regressions can be seen in **Table A.11** and the second stage in **Table A.11**. The key result from these regressions is that an increase in depression has a negative impact on trust, as measured by the WVS survey question.

In the first regression (1) in **Table A.11**, fitted values of endline CES-D are regressed against the WVS trust question. The coefficient on endline CES-D is significant at the 5% level and shows that for every 1 point increase on the CES-D scale the effect is a 0,0309 or approximately 3.1% decrease predicted on the WVS question. This coefficient remains negative and significant at the 5% level when additional controls are introduced in the subsequent columns.

The second set of 2sls regressions were run to investigate the effect of endline depression on the "**Likert Trust Question**". Controls are then added, which include health, computer literacy and employment. The first stages of the 2sls regressions can be seen in **Table A.12** and the second stage in **Table A.12**. The key result from these regressions is that an increase in depression has a significant and negative impact on trust as measured by the Likert survey question.

In the first regression (1) in **Table A.12**, fitted values of endline CES-D are regressed against the Likert trust question. The coefficient on endline CES-D is significant at the 1% level and shows that for every 1 point increase on the CES-D the effect is a 0,126 decrease predicted on the Likert question. This coefficient remains negative and significant at the 1% level when additional controls are introduced in the subsequent columns.

5.3.2 Impact of Depression on Trustworthiness

The third set of 2sls regressions were run to investigate the effect of endline depression on trustworthiness, which is the percentage returned by Player B in the trust game. Controls are then added, which include

health, computer literacy and employment. The first stages of the 2sls regressions can be seen in **Table A.13** and the second stages in **Table A.13**. The key result from these regressions is that an increase in depression has a significant and positive impact on trustworthiness as measured by the percentage returned by Player B.

In the first regression (1) in **Table A.13**, fitted values of endline CES-D are regressed against the % returned. The coefficient on endline CES-D is significant at the 10% level and shows that for every 1 point increase on the CES-D scale the effect is a 0,0139, or approximately 1.4% increase in the amount that Player B returns to Player A. This coefficient remains positive and significant at the 10% level when additional controls are introduced in the subsequent columns.

This result explains how the programme seemed to have a negative impact on the trustworthiness of men. The programme was not targeted or theorised to have a negative impact on trustworthiness and there is little literature on a programme such as Activate! reducing trustworthiness in men. This means that because the programme reduced male depression, as seen in section 5.1.1, the reduction of depression would cause a decrease in trustworthiness of men, which produces is the result seen in subsubsection 5.1.4, where men who are treated show a decrease in trustworthiness.

6 Discussion and Interpretation of Results

This paper examines the interactions, correlations and causal relationships between many variables. A summary of these results is listed below.

Impact of the Activate! Programme on depression, trust and trustworthiness

1. Depression is reduced for men who participate in the programme
2. Survey trust (WVS) increases for the treatment group
3. Experimental trust (Player A send) increases for the treatment group
4. Trustworthiness decreases in men who participate in the programme

Correlations between endline depression and various trust measures

5. WVS trust is significantly and negatively correlated with endline CES-D
6. Likert trust is significantly and negatively correlated with endline CES-D
7. Experimental trust (Player A send) is negatively correlated with endline CES-D
8. Trustworthiness (% returned) is not significantly correlated with endline CES-D

Estimated impact of depression on trust and trustworthiness using 2sls regressions

9. WVS Trust: depression is shown to reduce trust measured by the WVS question
10. Likert trust: depression reduces Likert trust
11. Trustworthiness: an increase in depression *increases* the amount returned by Player B

Of these results, most are consistent with the objective of the programme and the literature. Of particular interest though are the results listed under point 1, point 4 and point 11 above. These results are the interaction between the Activate! programme, endline depression and trustworthiness, which are discussed in detail below.

6.1 Understanding why Activate! reduces depression in men

The evidence that the Activate! programme reduces depression in men is of great interest, especially given the importance of depression in economics, as highlighted in the introduction. The literature on male depression helps explain this result, particularly in the context of detecting depression in men, utilisation of traditional mental health services by men and the use of group therapy on reducing male depression.

A starting point regarding male depression is exploring the difficulties of detecting depression in men. In a study that made use of a series of focus groups, Brownhill et al. (2002) found that men have been socially conditioned to suppress emotional pain, which means that emotional distress often presents itself as physical illness such as chest pains or behaviours such as alcohol/drug abuse. This is reflected in the tendency for men to delay help-seeking and may also cause depression symptoms being overlooked when physicians read the physical symptoms instead of the underlying psychological ones. There is other evidence that also shows that men under-utilise traditional helping services, which is likely as a result of cultural or ideological norms (Addis and Mahalik, 2003).

This is further supported by Warren (1983), who shows that depression in men is a private experience that is hidden from others and that men try to cope with it without external help. An explanation for this is that men are intolerant of depression, as it is not compatible with masculinity. This makes male depression

more complicated to detect and implies that it would be difficult to encourage and administer individual therapy to treat depression.

In a paper on the role of masculinity in help-seeking for depressed men, Seidler et al. (2016) review 37 studies that include 18 quantitative, 17 qualitative and 2 mixed-method studies. In line with the other literature, they find that men have difficulty in recognising and communicating their experience or symptoms of depression. Furthermore, masculine norms conflict with depression, increasing the stigma around it, reducing help-seeking and increasing the use of poor coping mechanisms (such as alcohol or drugs).

Of great interest is that Seidler et al. (2016) find that men prefer interventions that are collaborative and include action-orientated problem-solving. Additionally, men are shown to prefer short-term, group-based workshops as a form of therapy to individual/one-on-one treatment. Based on the evidence of this, Primack et al. (2010) outline a method of group therapy for men called "The Men's Stress Workshop", which is a form of targeted cognitive behavioural therapy for depressed men. The programme consists of eight groups sessions of around 1.5hours and is framed as a treatment focused on trying to understand one's self better and make better decisions/choices. They find that all participants on the programme reported a decrease in the severity of their depression and that all men on the treatment found it to be highly beneficial.

Therefore, Activate! was able to reduce depression in men because the programme acted as a form of group therapy. As shown in the literature, this is both an effective way to approach the treatment of depression in men and is potentially preferred by them. Furthermore, the Activate! programme had elements that were similar to that of "Men's Stress Workshop" in Primack et al. (2010), such as the focus on goal setting and on understanding oneself.

The implications of this are twofold. Firstly, future Activate! programmes could increase the impact of the programme on depression by using lessons from the literature and by adding a mental health professional to the team. Secondly, this helps advise policy and other programs on how group therapy can be used as an effective treatment for male depression.

6.2 Understanding why depression increases trustworthiness

The result from the 2sls regressions that depression increases trustworthiness is of great interest, especially given the counter-intuitive nature of this result. The literature on the relationship between depression and trustworthiness helps explain this result, particularly in the context of signalling, utilisation of traditional mental health services by men and the use of group therapy on reducing male depression.

A possible explanation for the greater prosocial behaviours displayed by depressed individuals is that they are using signalling. As discussed by Lotem et al. (2003) signalling benefits of altruistic acts can establish a stable generosity by high-quality individuals that no longer depends on the probability of future reciprocation or punishment. This may be because a depressed individual trusts less, as is seen in the other results, meaning that they use their own trustworthiness as a signalling tool to show that they are not like most people.

Another possible explanation is from Destoop et al. (2012), who had 39 patients diagnosed with depression and 22 healthy controls play a modified version of the ultimatum game, both in the role of proposer and responder against the same person. They found that patients with depression made significantly higher offers compared to the control group in the ultimatum game when compared to the healthy controls. One suggested explanation for the higher offer made by depressed individuals is that depression has been shown to be linked to a decreased responsiveness to reward sensitivity, resulting in a lack of interested in maximising monetary returns (Henriques and Davidson, 2000). Additionally, Destoop et al. (2012) suggest that the higher offers by depressed individuals may be due to deficits in second-order mentalizing, which involves imaginative mental

activity that allows a person to interpret or perceive another person's behaviour. This corresponds to the Harlé et al. (2010) conclusion that a depressed individual's decision making is biased.

Ong et al. (2017) found similar results for the effect of depression on trustworthiness when conducting a study on individuals with a history of mood disorders, specifically depression and bipolar disorder, using the trust game. They find that depressed individuals are more trustworthy, in that they return higher amounts to trustees when compared to healthy controls. They conclude that although people with mood disorders suffer from disruptions in their social decision making they may also display more prosocial behaviours, such as increased cooperation.

A final potential explanation includes gender as a factor. Caceda et al. (2014) examines the relationship between prosocial behaviours, gender and depression. In the background to the paper, they cite existing evidence that women normally display higher levels of prosocial behaviour, such as trustworthiness, than men do. Their study used a modified version of the trust game with depressed men, depressed women, healthy women and healthy men. They found that depressed men exhibited reciprocity to a significantly higher extent than healthy men. Furthermore, depression was shown to invert the self-centred, gender-specific behaviour. Based on this, they conclude that depression is related to the reversal of gender-specific prosocial behaviours.

Beyond the direct impact of depression on trustworthiness, there is the impact that it has on income and poverty. The paper bySlemrod and Katuscak (2005), which investigated the relationship between trustworthiness, trust and an individual's income helps understand why this might be the case. They find that **higher levels of trust have a positive impact on an individual's income**, while **higher levels of trustworthiness have a negative impact on an individual's income**. Based on the evidence in this paper, depression decreases trust (survey measure) and increases trustworthiness. Taken in combination, in the context of Slemrod and Katuscak (2005), this means that depression has a negative impact on a depressed individual's income in two ways.

Firstly, the effect of depression lowering their trust means that their income is reduced, which is likely as a result of the lack of willingness to invest in potentially beneficially economic activities. Secondly, because depression increases their trustworthiness their income is also reduced, as they are likely to reciprocate and be altruistic beyond the norms of their society. This causal pathway can be explored further in future research.

7 Conclusion

Using data collected from a randomised control trial conducted in a youth leadership programme in South Africa, I examine the effect of the programme on mental health and prosocial behaviours. The results showed that the programme was able to increase various measures of trust among participants and, interestingly, reduce depression in male participants. Additionally, the programme had an unintentional effect of reducing trustworthiness in men who were treated. This prompted an investigation into the effect of depression on trust and trustworthiness using instrumental variables to run a two-staged least square regression. The important result from these regressions was that higher levels of depression increased trustworthiness. Uncovering this showed the mechanism whereby the leadership programme reduced depression in men which in turn caused their trustworthiness to decrease. This result contributes to a gap in the literature, as trustworthiness has been shown to increase with depression, but no evidence existed regarding the causal nature of this relationship.

An important lesson from these results is that social programs that deal with how to modify behaviour, particularly negative behaviours for young people living in distressed or poverty-stricken environments, have to contend with potentially opposing effects where programs could be affecting outcomes that operate in multiple domains. Specifically, improving mental health outcomes may affect trustworthy behaviour as changing mental health can regulate and moderate how individuals are situated socially. Future research should investigate how to improve mental health outcomes while monitoring or mitigating the impact prosocial behaviours, such as trustworthiness while considering the effect of both factors on poverty.

References

Activate! (2019), 'About - Activate!'.
URL: *https://www.activateleadership.co.za/what-we-do/*

Addis, M. E. and Mahalik, J. R. (2003), 'Men, Masculinity, and the Contexts of Help Seeking', *American Psychologist* **58**(1), 5–14.

Adjaye-Gbewonyo, K., Avendano, M., Subramanian, S. and Kawachi, I. (2016), 'Income inequality and depressive symptoms in South Africa: A longitudinal analysis of the National Income Dynamics Study', *Health & Place* **42**, 37–46.

Alesina, A. F. and La Ferrara, E. (2002), 'Who Trusts Others?', *Journal of Public Economics* **85**(2), 207–234.

Andresen, E., Malmgren, J., Carter, W. and Patrick, D. (1994), 'Screening for depression in well older adults: Evaluation of a short form of the CES-D', *American Journal of Preventive Medicine* **10**(2), 77–84.

Angrist, J. D. and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton.

Ardington, C. and Case, A. (2010), 'Interactions Between Mental Health and Socio-Economic Status in the South African National Income Dynamics Study', *Journal for Studies in Economics and Econometrics* **34**(3), 69–85.

Arrow, K. (1973), The Theory of Discrimination, *in* O. Ashenfelter and A. Reeds, eds, 'Discrimination in Labor Markets', Princeton University Press, Princeton, NJ, pp. 3–33.

Ashraf, N., Bohnet, I. and Piankov, N. (2006), 'Decomposing trust and trustworthiness', *Experimental Economics* **9**(3), 193–208.

Baron, E. C., Davies, T. and Lund, C. (2017), 'Validation of the 10-item Centre for Epidemiological Studies Depression Scale (CES-D-10) in Zulu , Xhosa and Afrikaans populations in South Africa', *BMC Psychiatry* **6**(17), 1–14.

Berg, J., Dickhaut, J. and McCabe, K. (1995), 'Trust, reciprocity, and social history', *Games and Economic Behavior* **10**(1), 122–142.

Blount, S. (1995), 'When Social Outcomes Arent Fair: The Effect of Causal Attributions on Preferences', *Organizational Behavior and Human Decision Processes* **63**(2), 131–144.

Bohnet, I. and Baytelman, Y. (2007), 'Institutions and trust: Implications for preferences, beliefs and behavior', *Rationality and Society* **19**(1), 99–135.

Breuer, J. B. and McDermott, J. H. (2010), 'Trustworthiness and Economic Performance', *MPRA Paper* pp. 1–27.

Brook, J. S., Morojele, N. K., Pahl, K. and Brook, D. W. (2006), 'Predictors of drug use among South African adolescents.', *The Journal of Adolescent Health* **38**(1), 26–34.

Brownhill, S., Wilhelm, K., Barclay, L. and Parker, G. (2002), 'Detecting depression in men: a matter of guesswork.', *International Journal of Men's Health* **1**(3), 259–280.

Brülhart, M. and Usunier, J. C. (2012), 'Does the trust game measure trust?', *Economics Letters* **115**(1), 20–23.

Caceda, R., Moskovciak, T., Prendes-Alvarez, S., Wojas, J., Engel, A., Wilker, S. H., Gamboa, J. L. and Stowe, Z. N. (2014), 'Gender-specific effects of depression and suicidal ideation in prosocial behaviors.', *PloS one* **9**(9), 1–8.

Clark, C. B., Thorne, C. B., Hardy, S. and Cropsey, K. L. (2013), 'Cooperation and depressive symptoms', *Journal of Affective Disorders* **150**(3), 1184–1187.

Coleman, J. (1990), *Foundations of social theory*, The Belknap Press of Harvard University Press.

Cox, J. C. (2004), 'How to identify trust and reciprocity', *Games and Economic Behavior* **46**(2), 260–281.

Cox, J. C., Kerschbamer, R. and Neururer, D. (2016), 'What is trustworthiniess and what drives it?', *Games and Economic Behavior* **98**(1), 197–218.

Das, J., Do, Q. T., Friedman, J., McKenzie, D. and Scott, K. (2007), 'Mental health and poverty in developing countries: Revisiting the relationship', *Social Science and Medicine* **65**(3), 467–480.

De Silva, M. J., McKenzie, K., Harpham, T. and Huttly, S. R. (2005), 'Social capital and mental illness: A systematic review', *Journal of Epidemiology and Community Health* **59**(8), 619–627.

Destoop, M., Schrijvers, D., De Grave, C., Sabbe, B. and De Bruijn, E. R. (2012), 'Better to give than to take? Interactive social decision-making in severe major depressive disorder', *Journal of Affective Disorders* **137**(1-3), 98–105.

Duflo, E., Glennerster, R. and Kremer, M. (2007), Chapter 61 Using Randomization in Development Economics Research: A Toolkit, *in* 'Handbook of Development Economics', Vol. 4, Elsevier, pp. 3895–3962.

Eyal, K. and Burns, J. (2017), 'Up or Down? Intergenerational Mental Health Transmission and Cash Transfers in South Africa', *Saldru Working Paper Series* **200**.

Fehr, E. (2009), 'On the Economics and Biology of Trust', *Journal of the European Economic Association* **7**(2-3), 235–266.

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J. and Wagner, G. G. (2003), 'A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys', *CESifo Working Paper* **866**.

Fujiwara, T. (2007), 'The role of altruistic behavior in generalized anxiety disorder and major depression among adults in the United States', *Journal of Affective Disorders* **101**(1-3), 219–225.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A. and Soutter, C. L. (2000), 'Measuring Trust', *Quarterly Journal of Economics* **115**(3), 811–846.

Hamad, R., Fernald, L. C. H., Karlan, D. S. and Zinman, J. (2008), 'Social and economic correlates of depressive symptoms and perceived stress in South African adults', *Journal of Epidemiology & Community Health* **62**(6), 538–544.

Harlé, K. M., Allen, J. J. B., Sanfey, A. G., Harlé, K. M. and Sanfey, A. G. (2010), 'The impact of depression on social economic decision making', *Journal of Abnormal Psychology* **119**(2), 440–446.

Henriques, J. B. and Davidson, R. J. (2000), 'Decreased responsiveness to reward in depression', *Cognition & Emotion* **14**(5), 711–724.

Herman, A. A., Stein, D. J., Seedat, S., Heeringa, S. G., Moomal, H. and Williams, D. R. (2009), 'The South African Stress and Health (SASH) study: 12- month and lifetime prevalence of common mental disorders', *South African Medical Journal* **99**(5), 339–344.

Johnson, L. F., Rehle, T. M., Jooste, S. and Bekker, L.-G. (2015), 'Rates of HIV testing and diagnosis in South Africa', *AIDS* **29**(11), 1401–1409.

Johnson, N. D. and Mislin, A. (2012), 'How much should we trust the World Values Survey trust question?', *Economics Letters* **116**(2), 210–212.

Karim, J., Weisz, R., Bibi, Z. and ur Rehman, S. (2015), 'Validation of the Eight-Item Center for Epidemiologic Studies Depression Scale (CES-D) Among Older Adults', *Current Psychology* **34**(4), 681–692.

Kendler, K., Neale, M., MacLean, C., Heath, A., Eaves, L. and Kessler, R. (1993), 'Smoking and major depression: A causal analysis', *Archives of General Psychiatry* **50**(1), 36–43.

King-Casas, B. and Chiu, P. H. (2012), 'Understanding Interpersonal Function in Psychiatric Illness Through Multiplayer Economic Games', *Biological Psychiatry* **72**(2), 119–125.

Knack, S. and Keefer, P. (1997), 'Does Social Capital Have an Economic Payoff ? A Cross-Country Investigation', *Quarterly Journal of Economics* **112**(4), 1251–1288.

Kupferberg, A., Hager, O. M., Fischbacher, U., Brändle, L. S., Haynes, M. and Hasler, G. (2016), 'Testing the social competition hypothesis of depression using a simple economic game.', *BJPsych Open* **2**(2), 163–169.

Lazzarini, S. G. S. G., Madalozzo, R., Artes, R. and de Oliveira Siqueira, J. (2003), 'Measuring trust: An experiment in Brazil', *IBMEC Working Paper* **049**, 1–27.

Lerner, D., Adler, D. A., Chang, H., Lapitsky, L., Hood, M. Y., Perissinotto, C., Reed, J., McLaughlin, T. J., Berndt, E. R. and Rogers, W. H. (2004), 'Job Retention and Productivity Loss Among Employees With Depression', *Psychiatric Services* **55**(12), 1371–1378.

Leventhal, T. and Brooks-Gunn, J. (2003), 'Moving to Opportunity: an Experimental Study of Neighborhood Effects on Mental Health', *American Journal of Public Health* **93**(9), 1576–1582.

Lotem, A., Fishman, M. A. and Stone, L. (2003), 'From reciprocity to unconditional altruism through signalling benefits', *Proceedings of the Royal Society B: Biological Sciences* **270**(1511), 199–205.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R. and Sanbonmatsu, L. (2012), 'Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults', *Science* **337**(September), 1505–1510.

Mellick, W. (2014), The Relation Between Adolescent Depression and Interpersonal Trust, PhD thesis, University of Houston.

Mendelsohn, C. (2012), 'Smoking and Depression: A Review', *Australian Family Physician* **41**(5), 304–307.

Miller, A. S. and Mitamura, T. (2003), 'Are Surveys on Trust Trustworthy?', *Social Psychology Quarterly* **66**(1), 62–70.

North, C., McCullagh, P. and Vu Tran, Z. (1990), 'Effect of Exercise on Depression', *Exercise and Sport Sciences* **18**(1), 379–416.

Ong, D. C., Zaki, J. and Gruber, J. (2017), 'Increased cooperative behavior across remitted bipolar i disorder and major depression: Insights utilizing a behavioral economic trust game', *Journal of Abnormal Psychology* **126**(1), 1–7.

Patel, V. and Kleinman, A. (2003), 'Poverty and common mental disorders in developing countries', *Bulletin of the World Health Organization* **81**(8), 609–615.

Primack, J. M., Addis, M. E., Syzdek, M. and Miller, I. W. (2010), 'The Men's Stress Workshop: A Gender-Sensitive Treatment for Depressed Men', *Cognitive and Behavioral Practice* **17**(1), 77–87.

Quidt, J. D. and Haushofer, J. (2016), 'Depression for Economists', *NBER Working Paper Series* **22973**, 1–44.

Radloff, L. S. (1977), 'The CES-D Scale : A Self-Report Depression Scale for Research in the General Population', *Applied Psychological Measurement* **1**(50), 385–401.

Reddy, P., Zuma, K., Shisana, O., Kim, J. and Sewpaul, R. (2015), 'Prevalence of tobacco use among adults in South Africa: Results from the first South African National Health and Nutrition Examination Survey', *South African Medical Journal* **105**(8), 648.

Rosenberg, M. (1956), 'Misanthropy and Political Ideology', *American Sociological Review* **21**(6), 690–695.

Ross, C. E. (2000), 'Neighborhood Disadvantage and Adult Depression*', *Journal of Health and Social Behavior* **41**(2), 177–187.

Salmon, P. (2001), 'Effects of Physical Exercise on Anxiety, Depression, and Sensitivity to Stress: A Unifying Theory', *Clinical Psychology Review* **21**(1), 33–61.

Sapienza, P., Toldra-Simats, A. and Zingales, L. (2013), 'Understanding trust', *Economic Journal* **123**(573), 1313–1332.

Seidler, Z. E., Dawes, A. J., Rice, S. M., Oliffe, J. L. and Dhillon, H. M. (2016), 'The role of masculinity in men's help-seeking for depression: A systematic review', *Clinical Psychology Review* **49**, 106–118.

Slemrod, J. and Katuscak, P. (2005), 'Do Trust and Trustworthiness Pay Off?', *The Journal of Human Resources* **40**(3), 621–646.

Sobocki, P., Jönsson, B., Angst, J. and Rehnberg, C. (2006), 'Cost of Depression in Europe', *The journal of mental health policy and economics* **9**(2), 87–98.

Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R. and Morganstein, D. (2003), 'Cost of Lost Productive Work Time Among US Workers With Depression Walter', *Journal of the American Medical Association* **289**(23), 3135–3144.

Teychenne, M., Ball, K. and Salmon, J. (2008), 'Physical activity and likelihood of depression in adults: A review', *Preventive Medicine* **46**(5), 397–411.

Triant, V. (2002), The Recognition and Determinants of Depression at a South African Primary Care Clinic, PhD thesis, Yale University.

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V., Abraham, J., Ackerman, I., Aggarwal, R. and Ahn, S. Y. (2014), 'Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990 2010 : a systematic analysis for the Global Burden of Disease Study 2010', *Lancet* **380**(9859), 2163–2196.

Wang, Y., Zhou, Y., Li, S., Wang, P., Wu, G. W. and Liu, Z. N. (2014), 'Impaired social decision making in patients with major depressive disorder', *BMC Psychiatry* **14**(18), 2–11.

Warren, L. W. (1983), 'Male intolerance of depression: A review with implications for psychotherapy', *Clinical Psychology Review* **3**(2), 147–156.

Wooldridge, J. M. (2009), *Introductory Econometrics: A Modern Approach*, South-Western Cengage Learning.

World Health Organization (2017), Depression and Other Common Mental Disorders: Global Health Estimates, Technical report, World Health Organization, Geneva.

Zak, P. J. and Fakhar, A. (2006), 'Neuroactive hormones and interpersonal trust: International evidence', *Economics & Human Biology* **4**(3), 412–429.

Zak, P. J. and Knack, S. (2001), 'Trust and Growth', *The Economic Journal* **111**(470), 295–321.

A Appendices: Tables

Table A.1: Regression of various traits from survey on baseline depression (Base CES-D)

	(1) Base CES-D	(2) Base CES-D	(3) Base CES-D	(4) Base CES-D	(5) Base CES-D	(6) Base CES-D	(7) Base CES-D	(8) Base CES-D
Treatment	-0.0252	-0.0106	-0.00803	-0.00783	-0.00700	-0.0326	-0.0583	-0.0465
	(0.203)	(0.203)	(0.203)	(0.202)	(0.200)	(0.200)	(0.199)	(0.199)
Gender (1=Male)		-0.255	-0.253	-0.187	-0.113	-0.166	-0.108	-0.0739
		(0.205)	(0.205)	(0.207)	(0.206)	(0.206)	(0.207)	(0.208)
Unemployed (1=yes)			0.388	0.415	0.454	0.441	0.420	0.329
			(0.283)	(0.283)	(0.280)	(0.279)	(0.278)	(0.281)
Exercise (1= 3+ times pw)				-0.405**	-0.303	-0.305	-0.327	-0.340*
				(0.204)	(0.204)	(0.203)	(0.203)	(0.203)
Health (1=Excellent)					-0.700***	-0.693***	-0.691***	-0.645***
					(0.204)	(0.203)	(0.203)	(0.203)
Smoke cigarettes (1=yes)						0.737**	0.771**	0.801***
						(0.306)	(0.305)	(0.305)
Stay in Neighbourhood (1=yes)							-0.451**	-0.456**
							(0.208)	(0.208)
Computer Literacy (1=yes)								-0.423**
								(0.210)
Race (1=black African)								0.172
								(0.572)
Constant	4.335***	4.475***	4.414***	4.570***	4.787***	4.742***	4.894***	4.967***
	(0.149)	(0.187)	(0.192)	(0.207)	(0.214)	(0.214)	(0.225)	(0.622)
Observations	542	542	542	542	542	542	542	542

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

Table A.2: Regressions showing the Impact of the Activate! Programme on Endline Depression

	(1) CES-D	(2) CES-D	(3) CES-D	(4) CES-D	(5) CES-D	(6) CES-D	(7) CES-D	(8) CES-D
Treatment	-0.224	-0.134	-0.141	0.792	0.810	0.816	0.810	0.811
	(0.331)	(0.347)	(0.326)	(0.497)	(0.496)	(0.497)	(0.498)	(0.499)
Gender (1=Male)		-0.231	0.0307	0.892*	0.966**	0.970**	0.950*	0.952*
		(0.350)	(0.329)	(0.477)	(0.480)	(0.480)	(0.485)	(0.488)
Base CES-D			0.576***	0.570***	0.555***	0.554***	0.555***	0.556***
			(0.0687)	(0.0684)	(0.0691)	(0.0692)	(0.0694)	(0.0699)
Treatment x Gender				-1.626**	-1.657**	-1.663**	-1.654**	-1.655**
				(0.656)	(0.655)	(0.656)	(0.658)	(0.658)
Health (1=Excellent)					-0.477	-0.475	-0.488	-0.488
					(0.330)	(0.330)	(0.334)	(0.334)
Race (1=black African)						0.318	0.313	0.314
						(0.972)	(0.973)	(0.975)
Exercise (1= 3+ times pw)							0.0981	0.0981
							(0.332)	(0.332)
Smoke cigarettes (1=yes)								-0.0145
								(0.502)
Constant	6.255***	6.403***	3.744***	3.303***	3.540***	3.231***	3.199***	3.198***
	(0.242)	(0.318)	(0.428)	(0.462)	(0.490)	(1.066)	(1.072)	(1.074)
Observations	593	547	529	529	529	529	529	529

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

Table A.3: Regressions showing the Impact of the Activate! Programme on the WVS Trust Question (Q6)

	(1) Q6	(2) Q6	(3) Q6	(4) Q6	(5) Q6	(6) Q6	(7) Q6	(8) Q6
Treatment	0.110**	0.127***	0.119**	0.127*	0.134***	0.139**	0.119*	0.136*
	(0.0425)	(0.0444)	(0.0462)	(0.0665)	(0.0473)	(0.0691)	(0.0703)	(0.0733)
Base CES-D	-0.0177**	-0.0190**	-0.0133	-0.0207	-0.0129	-0.0204	-0.0174	-0.0169
	(0.00901)	(0.00932)	(0.00968)	(0.0143)	(0.00986)	(0.0144)	(0.0152)	(0.0153)
Gender (1=Male)	0.116***	0.124***	0.106**	0.102	0.113**	0.0960	0.110	0.103
	(0.0429)	(0.0445)	(0.0466)	(0.0676)	(0.0471)	(0.0681)	(0.0717)	(0.0723)
Race (1=black African)	-0.235*	-0.227*	-0.187	-0.165	-0.190	-0.160	-0.154	-0.149
	(0.121)	(0.121)	(0.134)	(0.161)	(0.133)	(0.162)	(0.172)	(0.172)
Mean Row (Risk)		-0.00193			-0.0119	-0.0171		-0.0215
		(0.0124)			(0.0131)	(0.0204)		(0.0217)
Dictator Game Offer			-0.215*		-0.236**		-0.0883	-0.124
			(0.117)		(0.120)		(0.173)	(0.175)
Expected Return				0.182		0.164	0.250	0.229
				(0.151)		(0.153)	(0.163)	(0.164)
Constant	0.694***	0.693***	0.700***	0.517***	0.747***	0.593***	0.478**	0.582**
	(0.130)	(0.142)	(0.147)	(0.195)	(0.159)	(0.212)	(0.220)	(0.240)
Observations	542	501	457	230	447	227	204	201

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Q6. Would you say most people can be trusted or that most people can't be trusted?

CES-D: a short, self reported scale used as an indicator of depression

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

Table A.4: Regressions showing the Impact of the Activate! Programme on % Sent of Player A (trust game)

	(1) % Sent	(2) % Sent	(3) % Sent	(4) % Sent	(5) % Sent	(6) % Sent	(7) % Sent	(8) % Sent
Treatment	0.0348	0.0327	0.0443	0.0515*	0.0414	0.0438	0.0606*	0.0529
	(0.0310)	(0.0324)	(0.0331)	(0.0307)	(0.0348)	(0.0320)	(0.0326)	(0.0342)
Base CES-D	0.00460	0.00490	0.00627	0.00604	0.00655	0.00687	0.00835	0.00904
	(0.00680)	(0.00691)	(0.00731)	(0.00660)	(0.00742)	(0.00669)	(0.00703)	(0.00714)
Gender (1=Male)	0.0677**	0.0676**	0.0753**	0.0611*	0.0753**	0.0626**	0.0676**	0.0685**
	(0.0313)	(0.0317)	(0.0334)	(0.0312)	(0.0339)	(0.0316)	(0.0333)	(0.0337)
Race (1=black African)	0.0956	0.0943	0.0759	0.0885	0.0743	0.0841	0.0548	0.0502
	(0.0754)	(0.0760)	(0.0805)	(0.0743)	(0.0813)	(0.0749)	(0.0796)	(0.0804)
Mean Row (Risk)		0.00357			0.00412	0.0102		0.00980
		(0.00950)			(0.0103)	(0.00944)		(0.0101)
Dictator Game Offer			0.0598		0.0642		0.000361	0.00967
			(0.0802)		(0.0823)		(0.0802)	(0.0818)
Expected Return				-0.0779		-0.0775	-0.0351	-0.0346
				(0.0698)		(0.0708)	(0.0754)	(0.0767)
Constant	0.277***	0.262***	0.268***	0.341***	0.249**	0.300***	0.336***	0.294***
	(0.0836)	(0.0930)	(0.0934)	(0.0899)	(0.105)	(0.0983)	(0.102)	(0.112)
Observations	256	253	228	230	225	227	204	201

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

% Sent: the % that Player A sends in the trust game

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

Table A.5: Regressions showing the Impact of the Activate! Programme on Return of Player B (trust game)

	(1)	(2)	(3)	(4)	(5)
	% Returned	% Returned	% Returned	% Returned	% Returned
Treatment	-0.0295	-0.0307	-0.0307	0.0350	0.0370
	(0.0233)	(0.0235)	(0.0236)	(0.0349)	(0.0349)
Base CES-D		0.00159	0.00166	0.00110	0.000872
		(0.00470)	(0.00473)	(0.00469)	(0.00469)
Gender (1=Male)			0.00401	0.0582*	0.0590*
			(0.0237)	(0.0317)	(0.0317)
Treatment x Gender				-0.119**	-0.122***
				(0.0470)	(0.0470)
Race (1=black African)					0.0866
					(0.0767)
Constant	0.259***	0.249***	0.247***	0.219***	0.135*
	(0.0158)	(0.0263)	(0.0303)	(0.0318)	(0.0811)
Observations	276	254	254	254	254

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

% Returned: the fraction returned by Player B in the trust game

Table A.6: Regressions showing the Impact of the Activate! on various trust, risk and dictator games

	(1) Q1	(2) Expected Return	(3) Mean Row (Risk)	(4) Dictator Game Offer
Treatment	0.197	0.0242	0.545***	-0.0265
	(0.121)	(0.0293)	(0.159)	(0.0185)
Base CES-D	-0.0751***	0.00107	-0.0206	0.000806
	(0.0256)	(0.00630)	(0.0338)	(0.00388)
Gender (1=Male)	0.354***	-0.0207	-0.122	-0.0139
	(0.122)	(0.0298)	(0.161)	(0.0187)
Race (1=black African)	-0.241	0.0786	-0.204	-0.0210
	(0.345)	(0.0708)	(0.440)	(0.0536)
Constant	4.250***	0.531***	4.520***	0.315***
	(0.369)	(0.0782)	(0.473)	(0.0569)
Observations	542	230	501	457

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

Q1. On a scale of 1 (not at all) to 7 (trust completely), how much can other people be trusted?

Expected Return: the % Player A expects to be sent back in the trust game

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Table A.7: Regressions showing the relationship between endline depression and WVS trust (Q6)

	(1) Q6	(2) Q6	(3) Q6	(4) Q6	(5) Q6	(6) Q6	(7) Q6	(8) Q6	(9) Q6	(10) Q6
End CES-D	-0.0228***	-0.0200***	-0.0197***	-0.0177***	-0.0169***	-0.0256***	-0.0167***	-0.0273***	-0.0248***	-0.0261***
	(0.00504)	(0.00578)	(0.00575)	(0.00592)	(0.00610)	(0.00850)	(0.00616)	(0.00860)	(0.00887)	(0.00901)
Base CES-D		-0.00777	-0.00761	-0.00810	-0.00292	-0.00287	-0.00265	-0.000833	-0.00115	0.000960
		(0.00968)	(0.00963)	(0.00989)	(0.0102)	(0.0153)	(0.0104)	(0.0155)	(0.0161)	(0.0163)
Treatment			0.113***	0.120***	0.118**	0.121*	0.131***	0.129*	0.115	0.125*
			(0.0428)	(0.0444)	(0.0462)	(0.0657)	(0.0472)	(0.0681)	(0.0696)	(0.0724)
Gender (1=Male)				0.133***	0.115**	0.116*	0.122***	0.111	0.130*	0.124*
				(0.0445)	(0.0466)	(0.0669)	(0.0472)	(0.0673)	(0.0713)	(0.0718)
Race (1=black African)				-0.331***	-0.318**	-0.331*	-0.323**	-0.328*	-0.344*	-0.344*
				(0.127)	(0.142)	(0.174)	(0.142)	(0.175)	(0.190)	(0.190)
Mean Row (Risk)				-0.00416			-0.0141	-0.0163		-0.0197
				(0.0124)			(0.0131)	(0.0201)		(0.0216)
Dictator Game Offer					-0.218*		-0.242**		-0.0678	-0.104
					(0.117)		(0.120)		(0.174)	(0.177)
Expected Return						0.103		0.0734	0.162	0.127
						(0.152)		(0.153)	(0.165)	(0.167)
Constant	0.662***	0.673***	0.610***	0.865***	0.887***	0.808***	0.945***	0.896***	0.789***	0.905***
	(0.0369)	(0.0497)	(0.0549)	(0.148)	(0.155)	(0.214)	(0.168)	(0.229)	(0.247)	(0.264)
Observations	593	529	529	488	446	224	436	221	198	195

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Q6. Generally speaking, would you say most people can be trusted or that most people can't be trusted?

CES-D: a short, self reported scale used as an indicator of depression

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

Table A.8: Regressions showing the relationship between endline depression and Likert trust (Q1)

	(1) Q1	(2) Q1	(3) Q1	(4) Q1	(5) Q1	(6) Q1	(7) Q1	(8) Q1	(9) Q1	(10) Q1
End CES-D	-0.0643***	-0.0561***	-0.0557***	-0.0623***	-0.0612***	-0.0855***	-0.0633***	-0.0847***	-0.0926***	-0.0902***
	(0.0144)	(0.0163)	(0.0163)	(0.0170)	(0.0172)	(0.0246)	(0.0174)	(0.0250)	(0.0250)	(0.0255)
Base CES-D		-0.0507*	-0.0505*	-0.0477*	-0.0488*	-0.0611	-0.0488*	-0.0677	-0.0600	-0.0680
		(0.0274)	(0.0273)	(0.0284)	(0.0288)	(0.0441)	(0.0293)	(0.0449)	(0.0453)	(0.0460)
Treatment			0.180	0.191	0.119	0.133	0.167	0.190	0.0519	0.133
			(0.122)	(0.127)	(0.130)	(0.190)	(0.133)	(0.198)	(0.196)	(0.205)
Gender (1=Male)				0.442***	0.380***	0.471**	0.398***	0.460**	0.442**	0.428**
				(0.128)	(0.132)	(0.193)	(0.133)	(0.195)	(0.201)	(0.203)
Race (1=black African)				-0.329	-0.479	-0.340	-0.493	-0.296	-0.382	-0.325
				(0.365)	(0.401)	(0.503)	(0.401)	(0.507)	(0.536)	(0.538)
Mean Row (Risk)				-0.0608*			-0.0761**	-0.0812		-0.106*
				(0.0355)			(0.0370)	(0.0584)		(0.0612)
Dictator Game Offer					-0.154		-0.255		-0.160	-0.292
					(0.331)		(0.340)		(0.490)	(0.500)
Expected Return						0.921**		0.909**	1.026**	1.010**
						(0.439)		(0.445)	(0.466)	(0.472)
Constant	4.408***	4.576***	4.475***	4.836***	4.821***	4.240***	5.174***	4.567***	4.342***	4.795***
	(0.106)	(0.140)	(0.156)	(0.425)	(0.438)	(0.618)	(0.474)	(0.663)	(0.696)	(0.746)
Observations	593	529	529	488	446	224	436	221	198	195

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Q1. On a scale of 1 (not at all) to 7 (trust completely), how much can other people be trusted?

CES-D: a short, self reported scale used as an indicator of depression

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

Table A.9: Regressions showing the Endline Depression on % Sent of Player A in the trust game

	(1) % Sent	(2) % Sent	(3) % Sent	(4) % Sent	(5) % Sent	(6) % Sent	(7) % Sent	(8) % Sent	(9) % Sent	(10) % Sent
End CES-D	-0.00174	-0.00444	-0.00433	-0.00461	-0.00347	-0.00675*	-0.00353	-0.00704*	-0.00554	-0.00585
	(0.00353)	(0.00414)	(0.00414)	(0.00419)	(0.00434)	(0.00398)	(0.00444)	(0.00406)	(0.00417)	(0.00428)
Base CES-D		0.00853	0.00852	0.00890	0.00967	0.0119*	0.00989	0.0129*	0.0135*	0.0143*
		(0.00742)	(0.00742)	(0.00751)	(0.00785)	(0.00715)	(0.00801)	(0.00729)	(0.00756)	(0.00772)
Treatment			0.0359	0.0317	0.0420	0.0494	0.0408	0.0426	0.0587*	0.0525
			(0.0314)	(0.0327)	(0.0335)	(0.0308)	(0.0352)	(0.0321)	(0.0327)	(0.0344)
Gender (1=Male)				0.0612*	0.0684**	0.0534*	0.0682**	0.0552*	0.0591*	0.0604*
				(0.0319)	(0.0339)	(0.0313)	(0.0345)	(0.0317)	(0.0335)	(0.0341)
Race (1=black African)				0.134	0.111	0.113	0.110	0.108	0.0725	0.0681
				(0.0833)	(0.0896)	(0.0816)	(0.0905)	(0.0823)	(0.0893)	(0.0903)
Mean Row (Risk)				0.00146			0.00174	0.00821		0.00754
				(0.00958)			(0.0105)	(0.00950)		(0.0103)
Dictator Game Offer					0.0503		0.0521		-0.00943	-0.000264
					(0.0818)		(0.0846)		(0.0818)	(0.0840)
Expected Return						-0.0722		-0.0743	-0.0326	-0.0342
						(0.0712)		(0.0723)	(0.0777)	(0.0792)
Constant	0.462***	0.439***	0.418***	0.248**	0.247**	0.336***	0.240**	0.306***	0.339***	0.309**
	(0.0261)	(0.0353)	(0.0399)	(0.0993)	(0.102)	(0.100)	(0.113)	(0.108)	(0.116)	(0.125)
Observations	283	250	250	247	222	224	219	221	198	195

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

% Sent: the % that A sends in the trust game

Mean Row: mean switching row in risk/coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Expected Return: the % Player A expects to be sent back in the trust game

Table A.10: Regressions showing the relationship between endline depression and various measures

	(1)	(2)	(3)	(4)
	% Returned	Expected Return	Mean Row (Risk)	Dictator Game Offer
End CES-D	0.00129	0.000370	-0.0162	0.00429*
	(0.00331)	(0.00379)	(0.0218)	(0.00247)
Base CES-D	0.00115	0.00119	-0.0145	-0.00117
	(0.00512)	(0.00680)	(0.0364)	(0.00415)
Treatment	-0.0288	0.0262	0.511***	-0.0297
	(0.0243)	(0.0292)	(0.162)	(0.0187)
Gender (1=Male)	0.00127	-0.0213	-0.104	-0.0186
	(0.0243)	(0.0297)	(0.164)	(0.0189)
Race (1=black African)	0.0753	-0.0315	-0.217	-0.0271
	(0.0785)	(0.0776)	(0.468)	(0.0577)
Constant	0.169**	0.639***	4.601***	0.307***
	(0.0823)	(0.0849)	(0.503)	(0.0613)
Observations	247	224	488	446

Standard errors in parentheses $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

% Returned: the fraction returned by Player B in the trust game

Expected Return: the % Player A expects to be sent back in the trust game

Mean Row: mean switching row in coin flip game, higher values indicate lower risk preferences

Dictator Game Offer: the % that A offers to B in the dictator game

Table A.11: First Stage IV Regression: Impact of depression on the WVS Trust Question (Q6)

	(1) End CES-D	(2) End CES-D	(3) End CES-D	(4) End CES-D
Treatment	0.792	0.810	0.834*	0.855*
	(0.497)	(0.496)	(0.494)	(0.495)
Base CES-D	0.570***	0.555***	0.569***	0.569***
	(0.0684)	(0.0691)	(0.0691)	(0.0691)
Gender (1=Male)	0.892*	0.966**	0.940**	0.954**
	(0.477)	(0.480)	(0.478)	(0.478)
Treatment x Gender	-1.626**	-1.657**	-1.722***	-1.745***
	(0.656)	(0.655)	(0.654)	(0.654)
Health (1=Excellent)		-0.477	-0.545*	-0.542
		(0.330)	(0.330)	(0.330)
Computer Literacy (1=yes)			0.747**	0.751**
			(0.339)	(0.339)
Unemployed (1=yes)				0.459
				(0.471)
Constant	3.303***	3.540***	3.060***	2.982***
	(0.462)	(0.490)	(0.534)	(0.540)
Observations	529	529	529	529

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Second Stage IV Regression: Impact of depression on the WVS Trust Question (Q6)

	(1) Q6	(2) Q6	(3) Q6	(4) Q6
End CES-D	-0.0309**	-0.0284*	-0.0271*	-0.0272*
	(0.0151)	(0.0156)	(0.0152)	(0.0152)
Health (1=Excellent)		0.0307	0.0241	0.0233
		(0.0451)	(0.0454)	(0.0452)
Computer Literacy (1=yes)			0.0681	0.0673
			(0.0452)	(0.0451)
Unemployed (1=yes)				-0.112*
				(0.0628)
Constant	0.707***	0.678***	0.629***	0.646***
	(0.0960)	(0.106)	(0.104)	(0.103)
Observations	529	529	529	529

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Q6. Generally, would you say most people can be trusted or that most people can't be trusted?

CES-D: a short, self reported scale used as an indicator of depression

Table A.12: First Stage IV Regression: Impact of depression on the Likert Scale Trust Question (Q1)

	(1) End CES-D	(2) End CES-D	(3) End CES-D	(4) End CES-D
Treatment	0.792	0.810	0.834*	0.855*
	(0.497)	(0.496)	(0.494)	(0.495)
Base CES-D	0.570***	0.555***	0.569***	0.569***
	(0.0684)	(0.0691)	(0.0691)	(0.0691)
Gender (1=Male)	0.892*	0.966**	0.940**	0.954**
	(0.477)	(0.480)	(0.478)	(0.478)
Treatment x Gender	-1.626**	-1.657**	-1.722***	-1.745***
	(0.656)	(0.655)	(0.654)	(0.654)
Health (1=Excellent)		-0.477	-0.545*	-0.542
		(0.330)	(0.330)	(0.330)
Computer Literacy (1=yes)			0.747**	0.751**
			(0.339)	(0.339)
Unemployed (1=yes)				0.459
				(0.471)
Constant	3.303***	3.540***	3.060***	2.982***
	(0.462)	(0.490)	(0.534)	(0.540)
Observations	529	529	529	529

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Second Stage IV Regression: Impact of depression on the Likert Scale Trust Question (Q1)

	(1) Q1	(2) Q1	(3) Q1	(4) Q1
End CES-D	-0.126***	-0.116***	-0.120***	-0.121***
	(0.0396)	(0.0407)	(0.0399)	(0.0396)
Health (1=Excellent)		0.159	0.163	0.160
		(0.117)	(0.119)	(0.118)
Computer Literacy (1=yes)			-0.0592	-0.0619
			(0.118)	(0.118)
Unemployed (1=yes)				-0.385**
				(0.164)
Constant	4.731***	4.602***	4.660***	4.720***
	(0.251)	(0.277)	(0.271)	(0.268)
Observations	529	529	529	529

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Q1. On a scale of 1 (not at all) to 7 (trust completely), how much can other people be trusted?

CES-D: a short, self reported scale used as an indicator of depression

Table A.13: First Stage IV Regression: Impact of depression on the % returned by Player B

	(1) End CES-D	(2) End CES-D	(3) End CES-D	(4) End CES-D
Treatment	1.502**	1.464**	1.485**	1.452**
	(0.690)	(0.689)	(0.690)	(0.693)
Base CES-D	0.503***	0.472***	0.479***	0.481***
	(0.0919)	(0.0944)	(0.0948)	(0.0950)
Gender (1=Male)	1.401**	1.493**	1.495**	1.517**
	(0.621)	(0.624)	(0.624)	(0.626)
Treatment x Gender	-2.934***	-2.999***	-3.061***	-3.057***
	(0.929)	(0.928)	(0.932)	(0.934)
Health (1=Excellent)		-0.681	-0.717	-0.733
		(0.487)	(0.490)	(0.491)
Computer Literacy (1=yes)			0.370	0.367
			(0.480)	(0.481)
Unemployed (1=yes)				0.417
				(0.682)
Constant	3.226***	3.621***	3.384***	3.330***
	(0.622)	(0.682)	(0.749)	(0.755)
Observations	247	247	247	247

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Second Stage IV Regression: Impact of depression on the % returned by Player B

	(1) % Returned	(2) % Returned	(3) % Returned	(4) % Returned
End CES-D	0.0139*	0.0170*	0.0167*	0.0164*
	(0.00835)	(0.00890)	(0.00881)	(0.00875)
Health (1=Excellent)		0.0406	0.0414	0.0423
		(0.0275)	(0.0276)	(0.0275)
Computer Literacy (1=yes)			-0.00968	-0.00933
			(0.0258)	(0.0257)
Unemployed (1=yes)				-0.0424
				(0.0366)
Constant	0.157***	0.122**	0.129**	0.136**
	(0.0531)	(0.0613)	(0.0625)	(0.0620)
Observations	247	247	247	247

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CES-D: a short, self reported scale used as an indicator of depression

% Returned: the fraction returned by Player B in the trust game