THE BEHAVIOURAL ECONOMICS OF INTERVENTIONS FOR HEALTHIER BELIEFS, CHOICES AND OUTCOMES

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2022

presented

by

Sofia Monteiro

from

Cape Town

First reviewer: Prof. Dr Matthias Sutter

Second reviewer: Prof. Dr Bettina Rockenbach

Date of oral defence:

ACKNOWLEDGEMENTS

The past four years as a Research Fellow at the Max Planck Institute for Research on Collective Goods (MPI) and PhD Candidate at the University of Cologne have been a privilege. This thesis is the culmination of the many opportunities, experiences and support extended to me by the remarkable individuals whom I wish to thank.

Firstly, my sincerest thanks to my supervisor, Prof. Dr **Matthias Sutter**, Director of MPI. I met Matthias when he hired me to join the newly formed Experimental Economics Group (EEG) in 2017. Matthias challenged me to develop novel research proposals, some of which we designed together and which he generously funded. Thanks to him, I had the opportunity to learn about designing and running a randomised controlled trial that improved the lives of patients with diabetes. During the most challenging times of data collection in the Covid-19 lockdown, Matthias gave me accountability through progress meetings and the moral support to see my projects to the finish line. In working on meaningful research with Matthias, I have been granted academic freedom and responsibility along with generous resources. It is a pleasure to learn about refining and conveying ideas from someone passionate about generating relevant, rigorous research. My research was funded by the MPI and the Deutsche Forschungsgemeinschaft.

Many thanks to my second supervisor Prof. Dr **Bettina Rockenbach**, Vice Rector of the University of Cologne and Director of the Reinhard Selten Institute. She has guided me and been an admirable presence in my academic life. Thanks, too, to my co-author and mentor, Prof. Dr **Daniel Wiesen**. His BEHnet Summer School with Prof. **Matteo Galizzi** (LSE) was the catalyst in my pursuing the topic of behavioural economics of health. Daniel has been generous with his time, thoughtful feedback and emotional support when things were tough.

I wish to express my gratitude to my co-authors. At the University of Cape Town (UCT), Dr **Kate Larmuth** for her seemingly boundless energy during our long days writing up proposals for ethics approval at UCT, then capturing and analysing data when we finally got them through. But most of all for her friendship and mentoring. Her dual passions for reversing diabetes and rock climbing have rubbed off on me. Kate also roped in her wonderful **Douw Steyn** for data cleaning. Thanks to my UCT co-authors Dr **James Smith**, **Georgina Pujol-Busquets Guillén**, and Dr **Jacolene Kroff**. At MPI, I am grateful that Dr **Matthias Praxmarer** invited me to develop a project together with himself and Matthias Sutter, which became an exciting lab experiment on team decision making.

Working with the Eat Better South Africa (EBSA) programme and the remarkable women in the communities it serves made for a rewarding project and learning experience. I wish to thank Prof. **Tim Noakes** and members of the Noakes Foundation including **Jana Retief**. At Kenilworth Medicross Medical Centre, **Lauren Angelil**, Dr **Neville Wellington**, Dr **Carol Bosch**, Sister **Nthabeleng Khatiti**, and **Jenny Hughes**. Lauren, I am glad we got talking about health coaching as a behaviour change tool at Amy's dinner party. Look how far we got! Thanks to Neville, for being open to undertaking such an ambitious project at his busy practice. Thabi, you always kept a sense of humour.

At MPI, I benefitted from conversations with Prof. Dr **Susann Fiedler**, Dr **Sebastian Tonke**, Dr **Sebastian Schneider**, and my officemates, **Daniel Salicath**, **Shambhavi Priyam** and **Nathan Maddix**. I have been inspired by my chats with formidable female academics, Prof. **Rachel Kranton** (Duke University), Prof. **Erin Krupka** (U Michigan), and Assistant Professor **Camille Terrier** (HEC Lausanne). I much appreciated the administrative support of the MPI, including **Zita Green** and **Lara Gröbelbauer**, as too the excellent research assistance I received, notably from **Michael Kuelper**, **Vyoma Muralidhara**, and **Laris Gomez**.

At UCT, thanks to Associate Prof. **Andre Hofmeyr**, Prof. **Justine Burns** and Prof. **Patrizio Piraino**, who got me excited about economics in the first place. To my UCT friends **Lauren**, **Amy**, **Aimee**, **Johanna**, **Madeleine**, **Melissa,** and **Sean**, and my friend, Dr **Anna Pegels** (German Development Institute/DIE) along with my tango crew in Bonn, thank you.

My thanks to my proofreaders: my aunt **Michele** and my cousin **Lucie**, for keeping a close eye on my grammar and helping to convey my ideas cogently. My mother **Amanda**, for getting me to where I am with her love and support, and my father **Pedro** for giving me the confidence to undertake this PhD, and for telling me that the most important question is "why?" when I was three years old. My sisters **Catie** and **Tina**, for being there for me; sometimes laughing, sometimes crying, occasionally both.

Thank you to **Tristan** for taking my mind out of work and into nature, excellent taste in films and for being there for me at the end.

TABLE OF CONTENTS

Introduction

The burden of non-communicable diseases such as diabetes is a growing global problem, not only for patients and families, but also for health insurance providers and the wider economy. It is a serious threat to human health that respects neither socioeconomic status nor national boundaries. Type 2 diabetes is largely lifestyle driven. However, health-related behaviours are difficult to shift, and measuring and tracking behaviour in the field is often a challenge. Societies have to consider designing health-promoting environments that support people in making 'the healthy choice'. Additional challenges are introduced when there is low health literacy. There may be unintended consequences when the adoption of a new self-monitoring technology is encouraged in certain populations. The mass of information of varying quality that people have access to through our mobile phones has to compete for our limited attention. In addition, we commonly face problems of selected information, and the way it is presented can distort our perception of reality. How we form beliefs and habits regarding nutrition, for example, can impact health outcomes. These observations motivate my thesis. I draw on the behavioural literature to design and test interventions aimed to debias individual belief formation and enable healthier decision-making on food choices, thus improving health outcomes. Beginning with the methodology foundational to behavioural economics, I explore cognitive biases in belief formation in the conventional experimental economist's laboratory as well as online (Chapter 1 and Chapter 2). Later, I use insights from behavioural economics to test empirical hypotheses on behaviour change in a cross-sectional non-randomised field study and in a randomised controlled trial in a developing-country context (Chapter 3 and Chapter 4). This thesis sheds light on the following interconnected questions: How can individuals be supported to shift their mental model and update their beliefs? Can people update their food choices when they receive nutrition education, and is there a valid methodology for measuring food preferences? What measures of support are needed to help individuals with different economic preferences make healthier decisions?

Biased beliefs can be costly to both the individual and society, whether they concern the economy or health. Many pieces of information we process quickly on a daily basis are correlated, meaning they are not independent voices from which to draw a conclusion. One example is traditional news media outlets, which share information sources, such as press agencies, so that the contents of different news reports tend to be correlated. Individuals have conceptual problems in identifying and thinking through the correlation problem in the first place. **Chapter 1 "Correlation neglect in belief formation with team versus individual decision-making"**, together with Matthias Sutter and Matthias Praxmarer, tests whether team decision-making can debias beliefs. This intervention has straightforward policy implications (i.e. when to implement team decision-making) for settings and organisations where individuals tend to exhibit correlation neglect (e.g. social network learning, forecasting, and newsgathering). This chapter contributes to the growing literature on team decision-making, in particular addressing a gap in the literature on teams and cognitive biases. It examines whether team decision-making generates more rational beliefs in settings where individuals tend to exhibit substantial correlation neglect. In the lab, team members can chat with their partners via text message to reach an agreement. The results show that team decision-making is closer to rational beliefs than individual decision-making. We use a single belief formation task as a proxy for naïveté of individuals to account for heterogeneity within teams. Pairs with at least one rational team member form significantly more rational beliefs than teams with two naïve members. Which is to say, having only one rational team member is sufficient to shift beliefs. For this project, my contribution was the data analysis and supporting data collection in the lab. The Covid-19 lockdown and resulting lab closure affected the initial planned data collection.

In news and social media, information is curated into "click-bait" headlines and posts that support a particular interest group. The fact that we face selected data, which is framed to tell a particular story – often one that agrees with our prior expressed beliefs – is not immediately apparent to the consumer. Given our limited attention, many people develop an erroneous mental model prompted by cognitive overload. A majority of individuals can be characterised into two types: someone that neglects the problem of selected information, or the Bayesian updater: someone that updates their prior belief to reflect new information provided. Finding a practicable means of debiasing individuals motivates **Chapter 2 "Visual representation training to debias selection neglect in belief formation"**. This chapter continues the theme of designing interventions to make beliefs more rational in settings where individuals suffer from cognitive biases. Visual representation training has been shown to be both intuitive and to improve Bayesian reasoning in basic problems of conditional probability. However, previous research on this type of representation training has been limited to non-incentivised psychological studies.

In an online experiment, I test whether representation training increases the proportion of Bayesian updaters in a setting in which individuals typically exhibit selection neglect. I designed a visual representation training exercise called "The Dice Game". The treatment significantly attenuates the bias, but there is only weak evidence for this. Chapter 2 provides the first evidence on using the psychological strategy of visual representation training in a behavioural economics experiment with incentivised beliefs, that is, where individuals are rewarded for accuracy with greater monetary payoffs. These first two chapters set the stage to take interventions to the field.

Much of the burden of non-communicable diseases in developing and developed countries is driven by what people eat and drink. Influencing dietary behaviour for prevention and management is therefore a key challenge for policymakers. However, measuring dietary changes can be a challenge with self-report food surveys possibly open to bias. Chapters 3 and 4 evaluate the impact and sustainability of dietary behaviour change interventions and critically examine the survey instrument used in nutrition studies from a behavioural economist's perspective. For example, how often did you eat a piece of white bread in the past four weeks? Four slices per week? It is an open question whether reported memories of food intake correlate with actual food behaviour. There is no gold standard with which to evaluate the Food Frequency Questionnaire, the standard instrument used in nutrition intervention studies. For other health behaviours, namely physical activity, there is no correlation between reported and biometric measures. **Chapter 3 "Improving food preferences through a nutrition education programme: An evaluation comparing survey evidence with a behavioural measure"**, together with Kate Larmuth, Georgina Pujol-Busquets Guillén and James Smith, explores whether measuring what people do, in addition to what they say they do, would provide a more credible channel to understand whether a program affects dietary preferences. One purpose of validating the Food Frequency Questionnaire in Chapter 3 is that we need a credible measure to capture food preferences later in Chapter 4's diabetes reversal intervention in the Western Cape, South Africa.

Chapter 3 investigates the food choice behaviour of women taking part in a Low Carbohydrate, High Fat (LCHF, alternatively called Ketogenic) lifestyle intervention in under-resourced communities in the Western Cape. I had the opportunity to work with the Noakes' Foundation's Eat Better South Africa! (EBSA) programme. EBSA targets health and socioeconomic issues faced by underprivileged South Africans, particularly women. Its programme includes education on nutrition, non-communicable diseases, shopping on a budget, cooking, and how to access healthier nutrient-dense foods.

Measuring behaviours of interest (e.g. what people choose to eat after receiving some advice) is arguably critical to demonstrate the mechanism through which such a program may succeed or fail, and requires a valid measurement tool. Previously, the effectiveness of the EBSA program was evaluated qualitatively through focus group discussions. This is in alignment with much of the scientific literature, where epidemiological nutrition studies, such as the Nurses' Health Study, use diet assessment tools that require subjective responses, such as the Food Frequency Questionnaire, 24-hour food recall or food diaries. It is understood that these subjective measures may suffer from bias and noise due to people's inattention to what they eat, inability to recall fully, social desirability bias and a lack of incentive for accuracy. Food choice behaviour is typically not observed, and the question is whether ours was a more objective measure, or at least, how it may compare with subjective surveys.

In Chapter 3, a behavioural measure of food preferences is designed and compared to a standard survey in the nutrition literature about what participants had eaten over the past four weeks. The behavioural component is a grocery shopping activity in which participants spend a voucher at a major supermarket. Covid-19 introduced additional complexity to field research in July/August 2020. The project was redesigned and resubmitted for ethics approval to remove the risks of interpersonal contact, so all my interviews were conducted over the phone. Remote data collection has its own challenges – for example, load-shedding electricity interruptions to Wi-Fi and potential ambivalence towards interacting with an unknown researcher over the phone. However, what made my experience of remote data collection encouraging was the generosity of the women I spoke to and the support from the Noakes' Foundation when we reached out to communities for recruitment.

The quality of the women's purchases is organised according to the programme's traffic lights lists. The same approach is used to categorise food choices later in Chapter 4. RED is foods to be avoided, ORANGE for occasional consumption, and GREEN to be eaten liberally.[1] The results show that the treatment group made fewer RED choices and more GREEN choices in their grocery shopping compared to the control group of similar women who had not yet been on the EBSA programme. Moreover, the behaviour we observe is reflected in their survey responses, validating the Food Frequency Questionnaire in our sample. Data were collected during one of the strictest lockdowns globally, which impacted participants' employment and food security. Despite the adverse conditions, there is a marked difference in food choice behaviour by the EBSA group in line with the LCHF/Ketogenic program advice. This difference speaks to the sustainability of the nutrition education model with women who have taken part, and who, by inference, are attempting to follow an LCHF diet. However, it does not link food choices to objective health outcomes, such as blood tests. These outcomes are explored later in Chapter 4.

While only an approximation of usual eating habits, the study showed that the inexpensive self-report food survey is representative of women's revealed food preferences. While behavioural economists will likely continue to take a sceptical view of survey measures, the results from this research suggest that when it comes to food: what she says is what she does. The results of Chapter 3 on the food survey lend support for its use as an instrument to measure dietary intake in **Chapter 4 "Technology-assisted behavioural interventions in type 2 diabetes"**, a randomised controlled trial together with Matthias Sutter, Kate Larmuth, Daniel Wiesen and Jacolene Kroff. One finding from Chapter 4 that strengthens Chapter 3 comes from the analysis of blood test Haemoglobin A1c (HbA1c, an objective health marker used to diagnose diabetes). Among a sample of type 2 diabetic patients in a primary care practice, better glucose control is positively associated with consuming more food from the GREEN list and negatively associated with eating more food from the RED list in the preceding four weeks.

Chapter 4 asks: What level of personalised feedback and health practitioner support is most helpful for diabetes patients to implement their ketogenic diet and achieve glucose control goals? The field

[1] RED comprises foods high in sugar/carbohydrates and processed foods containing vegetable oil. ORANGE lists fruit and beans. GREEN includes green leafy vegetables, meat, fish, eggs, full-cream dairy products and traditional fats (butter, lard, olive oil). The full lists are available in the appendix and online.
https://thenoakesfoundation.org/news/media/the-noakes-foundation-food-lists

experiment in Chapter 4 tests the impact of providing a minimally invasive wearable technology called Continuous Glucose Monitoring (CGM) with or without real-time feedback through a mobile phone app. The technology is provided as part of a physician-monitored programme with ketogenic nutrition education and a goal to reach at least 70% time in target glucose range in the CGM report, ultimately reversing diabetes. There are three treatment groups: CONTROL, INFO and COACH. The CONTROL group wears a CGM that cannot be viewed until the data are downloaded intermittently at doctor's visits, where the report is shared with the patient in a supervised way. All patients review the CGM report with the doctor. Real-time feedback on personal glucose levels via the CGM potentially allows the wearer to fine-tune their diet faster but may not be sufficient to overcome cognitive barriers such as present bias or confirmed habits. The effect of real-time CGM feedback is tested in treatment INFO. There is the potential to be overwhelmed with too much unsupervised information when new technology is adopted by patients in outpatient practices (as opposed to clinic stays), especially if they are not necessarily familiar with smartphone apps and/or have limited health literacy. In a third treatment, COACH, we attempt to address this by adding remote health coaching via video chat to help patients identify their goals, what it would mean to achieve them, the obstacles in the way and if/then plans to overcome them. Dietary questionnaires are used to assess the proportion of foods eaten by participants from the same RED, ORANGE, and GREEN lists as Chapter 3.

In the first month of the intervention, all groups show significant improvement in mean glucose levels and achieve the 70% time in target goal, indicating they are implementing the dietary advice. The significant reductions in RED list foods from the survey support this. There is no significant difference between groups in average glucose levels after one month of wearing the CGM with or without real-time feedback. All treatment groups begin with time in target glucose range of 46-52%, which is below the clinical goal of 70% in the range 3.9-8 mmol/L.[2] After a month of wearing the CGM, the best performance on time in target range is the COACH group with 83%, followed by the CONTROL group and INFO group at 74%. Greater time in target range by COACH is not due to more engagement with the CGM than INFO. The rich report of the CGM closes the information gap between patient and doctor, and offers an accountability mechanism. We do not,

[2] We set a substantially more ambitious target range than that of the international guidelines for adult diabetes patients which is 3.9-10 mmol/L.

however, find evidence of the value added by providing access to the relatively more expensive real-time CGM on its own.[3] In fact, the adoption of the real-time devices may be said to have adversely impacted glucose control in a population familiar with fingerprick testing only. At the endline blood test at 3-months, after a sustainability period of two months without real-time CGM to allow the effect of the 1-month CGM intervention on blood markers to be washed out, the real-time group has significantly worse outcomes than the CONTROL group, though an improvement compared to baseline. By the end of the sustainability period, all treatment groups continue to see a significant reduction in average glucose (HbA1c blood test) compared to baseline. However, the real-time with coaching group achieves the greatest reduction in HbA1c of 19%, while control achieves 14% and real-time alone 11%. The treatment effect of coaching is robust and clinically significant. The COACH group is the only one to improve other measured blood markers of metabolic health (HDL and triglycerides).

A notable impact of the present technology-assisted behavioural intervention programme with ketogenic nutrition education is that 43% of the pooled sample of patients reverse their diabetes (HbA1c<6.5%). A similar rate of reversal at 3- and 6-months has been achieved in a non-randomised study of a continuous care clinic using a ketogenic diet and coaching in the United States. The low-carb/ketogenic diets are popular among clinicians and patients but still debated controversially in the literature. In contrast, the rate of diabetes reversal with best practice standard of care has been less than 1%, which explains why type 2 diabetes has been framed as an incurable chronic disease with a focus on management of symptoms with pharmaceutical interventions. About 50% of patients with type 2 diabetes need insulin within ten years of diagnosis. Each case of diabetes reversal is cost-saving for the patient and their health insurance provider if they have one. The possibility for diabetes reversal through dietary interventions has only very recently been acknowledged by the World Health Organisation. This chapter provides convincing evidence that a high rate of diabetes reversal is an attainable goal in 3-months in South African outpatient primary care facility.

[3] The cost of the Libre Pro CGM (no feedback) is ZAR 650 (38 EUR) and the FreeStyle Libre CGM (real-time scans) is ZAR 990 (57 EUR).

Chapter 4 also finds that survey measures of risk aversion and grit are significantly positively associated with better glucose control (lower HbA1c) at endline. Future programmes could benefit from identifying subgroups of patients that may benefit from particular technology-assisted mechanisms of behavioural change by, for instance, measuring economic preferences and psychological scales, as done here. Such patient identification can be used to allocate limited resources more effectively, especially in developing countries. This affirms the relevance of designing interventions to support rational belief formation and healthier decision-making introduced in Chapters 1, 2 and 3. This thesis contributes to the behavioural economic literature, and offers a perspective on the nutrition and medical literatures for diabetes management by designing interventions that address chronic diseases of lifestyle informed by behavioural insights.

Chapter 1. Correlation Neglect in Belief Formation: Team versus Individual Decision-Making

Sofia Monteiro*[1,2] Matthias Praxmarer Matthias Sutter[1, 2, 3]

[1]Max Planck Institute for Research on Collective Goods, Bonn

[2]University of Cologne

[3]University of Innsbruck

Abstract. Recent research explores how individual beliefs respond to information sources that are correlated with each other rather than independent signals. Correlated signals characterise much of the information we process quickly on a daily basis, and the resulting biased beliefs can be costly. We show that team decision-making ameliorates this bias, generating more rational beliefs in a setting where individuals tend to exhibit substantial correlation neglect. In a 2x2 between-subjects design, participants are randomly assigned to either INDIVIDUAL treatment, with *Uncorrelated* or *Correlated* information structures and individual beliefs, or TEAM treatment, with *Uncorrelated* or *Correlated* information structures and team beliefs. Teams of two use a chatbox to reach an agreement. To account for the heterogeneity within teams, we add a single incentivised belief-formation task as a proxy for the naïveté of individual members. Pairs with at least one rational member form significantly more rational beliefs than teams with two naïve members and individuals working alone. Just having one rational team member is sufficient to shift beliefs. In the chat data, teams with at least one rational member shared more communications coded as computational logic and fewer proposals outside the rational interval.

JEL Codes: C91, D03, D83, D84, D40

Keywords: Beliefs, Correlation neglect, Bounded rationality, Team decision-making, Experiment

* Corresponding author: sofia.g.monteiro@gmail.com

1. INTRODUCTION

Biased beliefs can be costly to both the individual and society, whether they concern the economy, an upcoming election, or health. Correlated signals characterise much of the information we process quickly on a daily basis. One example is traditional news media outlets, which share information sources such as press agencies, so that the contents of different news reports tend to be correlated (Enke and Zimmermann, 2019). As a case in point, CNN, Fox News and Sky News all share a common source, Associated Press (AP), and reach their conclusions by taking into account an additional raw news source, *i.e.* a social media platform; among them Twitter, Facebook and YouTube.[4] The digested estimates of the traditional news media outlets will tend to be correlated with their common source, AP, and should not be considered independent voices from which to form a belief about the state of the world. A second example is that of learning in social networks. Individual opinions of different network members or friends are often partially based on information from a shared acquaintance. This means that when we speak to these network members, we are presented with informational redundancies. The same stories are being retold multiple times (Akerlof and Shiller, 2009). Information structures (channels that digest and transmit communications to us) are widely characterised by correlated signals. Enke and Zimmermann's (2019) "*Correlation Neglect in Belief Formation*" (EZ19) investigates how individuals respond to information structures that generate correlated rather than independent signals. To demonstrate a practicable debiasing strategy would have policy relevance, given the substantial correlation neglect observed, the pervasiveness of correlated information structures, and the apparent costliness of biased beliefs resulting from correlation neglect. Finding a means to debias beliefs is the motivation for this paper. Our novel contribution examines whether team decision-making, where members exchange text messages via free-form chat to reach an agreement, generates more rational beliefs in settings where individuals tend to exhibit substantial correlation neglect. Our results are highly relevant given that virtual teams are increasingly common in today's organisations, and that there is a gap in the literature on teams and cognitive biases.

[4] E.g. CNN's estimate = (Twitter's estimate + AP's estimate)/2. Therefore, given only CNN and AP's estimates, Twitter's estimate can be backed out by inverting the average.

EZ19 provide evidence of individuals' correlation neglect when forming beliefs from information sources in a transparent setting. The distribution of beliefs is strongly heterogeneous and bimodal, suggesting different belief updating types, *i.e.* rational and "full neglect". Since subjects are paid for accuracy, biased beliefs lead to significantly lower earnings.[5] The experiment's incentive-compatible payoff rule reflects the real-world consequence that biased beliefs can be costly to the individual. Their findings are consistent with boundedly rational social learning models. In the EZ19 setup, individual beliefs are swayed by (transparently) biased information sources, meaning that they consistently overshoot the true state of the world. This observation speaks to a recent paper by Charness et al. (2021), which empirically demonstrates that – when choosing between sources of biased information – decision-makers are vulnerable to reasoning errors such as confirmation-seeking.

EZ19 explore several mechanisms and extensions to try to eliminate this bias. They present evidence that naïve beliefs are not driven by inadequate mathematical and computational skills necessary to process correlated information. Rather, that subjects show conceptual problems in identifying and thinking through the correlation in the first place when the informational environment becomes complex enough. When subjects are exogenously made to focus on the correlation and underlying independent signals, this "nudge" strongly improves the rationality of beliefs. However, explicitly pointing out the problem is arguably not a practicable strategy; a gap remains for a straightforward policy to debias beliefs. The purpose of the present experiment is to replicate EZ19 and to determine whether team decision-making debiases beliefs. Such a finding would have straightforward policy implications (*i.e.* when to implement team decision-making) for settings and organisations where individuals tend to exhibit correlation neglect (e.g. social network learning, forecasting, and newsgathering).

Our paper contributes to the growing literatures on team decision-making (for a review, see Kocher, Praxmarer and Sutter, 2020) and errors in reasoning (e.g. Charness et al., 2021; Enke, 2019; Monteiro, forthcoming), respectively. There is a gap in the literature on teams and cognitive biases in behavioural economics, which is where this paper is located. However, there are a few

[5] Because of biased beliefs, participants in *Correlated* earned about 2.70 Euro less than those in the *Uncorrelated* benchmark group, which translates to almost 50% of subjects' average variable earnings (EZ19: 320).

informative examples outside of the economics literature. Houghton et al. (2000) note the persistence of biases (law of small numbers and illusion of control) on teams' risk perceptions compared to individuals. This evidence does not suggest that team decision-making is preferable to the individual level. Mancuso et al. (2014) explore confirmation bias and distributed (remote) teams. They find that when incorrect information is presented early, teams appear to focus more on that information in their proceeding discussions, and report incorrect answers more often. This finding is suggestive of confirmation bias in the team's deliberation process. Lehner et al. (1997) find that as time stress increases, teams begin to use a decision-processing strategy that is less effective than the strategy they are trained to use in the experiment. Recent research highlights the attenuation of extrapolation bias (a behavioural bias) as a potential benefit of team-based asset management compared to individual management (Barahona et al., 2021). Their evidence suggests that the reduction in behavioural bias may be driven by eliciting cognitive reflection in team members. There remains a need for further research on team cognition and team cognitive biases.

Team decision-making increasingly characterises modern organisations, and some evidence suggests that teams can significantly enhance productivity (e.g. Bandiera et al., 2013). Economic decisions, both in firms and households, are often made by groups. Research on teams is thus highly relevant for understanding financial choices. Virtual teams that interact using text-based collaboration technology are also increasingly common. Collaboration technology makes unique demands on individual cognitive resources that may influence how team members process information in virtual settings compared to face-to-face environments (Minas et al., 2014). Text-based communication, such as instant messaging and e-mail, possesses some benefits over face-to-face communication yet also introduces potential problems. Previous research indicates that teams using text-based collaboration typically exchange more information than when they perform the same task in a face-to-face setting, but do not necessarily reach better decisions than face-to-face teams (Dennis, 1996). One reason is that team members engage in "multiple monologues" (Dennis and Valacich, 1994), in which they talk about what they know instead of considering information from other team members. More recently, the results of an experiment on online team discussions using electroencephalography, electrodermal activity and facial electromyography show that information that contradicts a person's pre-discussion decision preference is processed similarly to irrelevant information, while information that confirms an individual's pre-discussion decision preference is processed more thoroughly (Minas et al., 2014).

It goes without saying that the advantage of team decision-making is apparent only if a naïve player defers to their sophisticated partner and not the other way around. Some individuals assign a significantly positive intrinsic value to having the power to make a decision. This distorts the choice between delegating a decision or not, and produces conflict and inefficiency (Buffat, Praxmarer and Sutter, 2020). It is arguably worth understanding the decision process (e.g. decision time and free-form chat messages) within teams of varying sophistication and the extent of naïveté in the distribution of beliefs that follows in the present paper. Previously, Cooper and Kagel (2005) used dialogues between teammates to identify factors promoting strategic play in signalling game experiments. They find that teams consistently play more strategically than individuals. They posit that, intuitively, a team should be no less likely to solve a problem than its most able member would be acting alone. In the psychology literature, teams consistently fall short of this "truth wins" norm (e.g. Davis, 1992), which contrasts with Cooper and Kagel's (2005) findings. The difference is attributed to experimental procedural differences in economics versus psychology experiments, such as incentive-compatible decision-making. We contribute to the existing literature by testing whether the presence of at least one rational team member is sufficient to attenuate correlation neglect compared to individuals (and teams with only naïve members).

Recent work in economics explores some of the differences between the decisions by groups versus individuals (Adams and Ferreira, 2010; Buffat, Praxmarer and Sutter, 2020; Kocher and Sutter, 2005; Kocher, Strauß and Sutter, 2006). Groups should make decisions that are easier for others to predict compared to individual decisions since groups have access to more evidence collected through information pooling. Individuals differ in terms of the information they have and, in settings where a compromise is required, deliberation should lead to information sharing within the group (Adams and Ferreira, 2010). Group members favour choices that are easier to justify, and this leads groups to rely more on tangible evidence (Barber, Heath, and Odean, 2003). Small groups learn comparatively faster than individuals and earn greater payoffs in beauty-contest games (Kocher and Sutter, 2005). In order to reach consensus, groups need to balance individual opinions. It follows, naturally, that groups should make less extreme decisions than individuals.

Empirically, Adams and Ferreira find that the distribution of guesses made by groups of bettors conforms better to historical data[6] than the distribution made by individuals. Their evidence suggests a moderating effect of group decision-making.[7] However, it remains an open question whether team decision-making would debias beliefs in our setting of correlation neglect.

In our design, individuals and teams are asked to estimate the true state of the world in 10 estimation tasks from EZ19. A 2x2 between-treatment experiment tests our hypothesis that individuals are more likely to submit beliefs that can be characterised as correlation neglect than teams. Subjects are randomly assigned to one out of four groups: (1) INDIVIDUAL with *Correlated* or (2) INDIVIDUAL with *Uncorrelated* information structures and individual beliefs[8] or (3) TEAM with *Correlated* or (4) TEAM with *Uncorrelated* information structures and team beliefs. Teams consist of randomly matched pairs within a session, and partners are fixed across tasks. Based on the team decision-making literature from the laboratory and field, we expect teams to be more likely to form rational beliefs than individuals. Building on EZ19, we hypothesise that naïve subjects would be less certain about their beliefs and more likely to defer to their teammate (even without knowledge of the partner being naïve or rational). The consequence would be that mixed teams with at least one rational member would do as well on average as rational individuals forming beliefs on their own. Team decision-making as a policy would then be preferred since forming a shared belief increases the rationality of the naïve team member and does not negatively impact the rational team member.

In our design innovation, the signal about a teammate's naïveté (not known to the players) is measured by an individual's incentivised belief in a single decision task with either a *Correlated* or *Uncorrelated* information structure (consistent with their randomly-assigned treatment group). Our addition of an individual measure of naïveté allows us to classify teams since teams are composed of subjects with varying levels of naïveté. We use this measure to explore in which cases teams of two are more rational than individuals. Moreover, the coding of teams' deliberation

---

[6] On the break-up date of the ice in the Nenana river in Alaska.

[7] Our paper also links to the literature on the accuracy of forecasting by experts and non-experts (e.g. DellaVigna and Pope, 2018; DellaVigna, Pope and Vivalt, 2019). We examine response time as a proxy for cognitive effort. Our sample of university students concerns non-experts.

[8] Treatments 1 and 2 are a replication of EZ19's baseline treatments.

processes using the text chat data provides insights not possible in the original setup of EZ19, which studied individual belief formation.

We establish that (1) there are significant differences between *Correlated* and *Uncorrelated* conditions for individuals and teams, respectively; (2) on average, team beliefs converge more often towards the rational belief than individual beliefs; (3) pairs with at least one rational member form significantly more rational beliefs than teams with two naïve members; and (4) communication plays a role, evident in differences in longer response times when there is a rational team member present. Teams tend to chat more at the beginning of the set tasks, and teams with one rational member exchange more comms coded as computational logic than teams with two naïve members. The rest of the paper proceeds as follows: Section 2 details the experimental design, Section 3 presents the results, and Section 4 concludes.

2. A NOVEL EXTENSION OF THE ENKE-ZIMMERMANN MODEL OF CORRELATION NEGLECT TO TEAM DECISION-MAKING

2.1. Experimental design

We extend on the experimental design developed by EZ19. The design of *Correlated* and *Uncorrelated* conditions allows: (a) the experimenter to perfectly control the signals displayed and the degree to which the information communicated is correlated (Figure 1), (b) subjects full knowledge of the transparent data-generating process; (c) a control group (*Uncorrelated*) that provides a benchmark for beliefs in the absence of correlated information; and (d) incentivised belief elicitation, in which subjects are paid for accuracy.

Our novel extension introduces a team decision-making treatment, in which randomly-assigned pairs can communicate via free-form text messages in order to agree on a team belief. The motivation for team decision-making is a practicable strategy to debias beliefs, and the chatbox provides insights into the belief formation process. It also provides us with another means of assessing whether participants understand (and believe) the transparent data generating process, which was not possible in the setup of EZ19. We analyse the chat data below to comment on the validity of the EZ19 model of correlation neglect. Another novel addition to the original design is an individual-level assessment of naïveté prior to the main 10 tasks. The estimation tasks are jointly completed by teams or individuals, respectively. The motivation for the addition of a "pre-belief" proxy of individual naïveté is to classify the teams. Participants are randomly assigned to teams of two, generating teams with 0 rational members, 1 rational member and 2 rational members. This classification is detailed further below. In the following section, we describe the setup developed by EZ19 in more detail and then present our procedures and treatments.

2.1.1. The Estimation Task

Subjects are asked to estimate an *ex ante* unknown continuous state of the world (μ) and are compensated financially for accuracy. The task is framed as guessing how many items are in an imaginary container. The only information given to subjects is unbiased computer-generated signals about the true state. The between-subjects design is motivated by constructing two sets of signals (one with and one without a known and simple correlation), which are identical with respect to their objective informational content. As shown in Figure 1, sourced from the EZ19, subjects in the *Correlated* condition received correlated information and subjects in the *Uncorrelated* condition uncorrelated information about μ.

The fictitious computers A-D generate four unbiased, independent and identically distributed (i.i.d.) signals about μ, identical across participants and treatment groups. This is implemented by random draws from a truncated discretised normal distribution with mean μ and standard deviation $\sigma = \mu/2$. Truncation is at $\mu \pm 2\sigma = \mu \pm \mu$ in order to exclude negative signals. In the *Uncorrelated* condition (left panel, Figure 1), the intermediaries 1-3, which are also fictitious computers, observe the signals of computers B through D, respectively, and simply transmit these signals to the participant. This means that subjects receive information from both computer A directly and the intermediaries 1-3. For example, in one experimental task, the signals of computers A through D are given by 122, 90, 68 and 5, respectively. As in EZ19, we refer to all numbers that are communicated to participants as "messages".

In the *Correlated* treatment (right panel, Figure 1), the three intermediaries observe the signal of computer A and computers B-D, respectively, and then report the average of these two signals.

Participants are given information from computer A (which EZ19 refer to as "the common source signal") as well as from the intermediaries 1-3. In the example above, each of the three intermediaries takes the average of 122 and the corresponding signal of the other computer it communicated with. Therefore, computer A reports 122, intermediary 1 reports 106, intermediary 2 reports 95, and intermediary 3 reports 64. The information structure transmits communications to the participant that obscure the underlying signals, so the participant has to start by backing out the correct signals before she states a belief. She mostly sees messages larger than 90, but the majority of signals is actually 90 or lower. Since the common source signal of computer A is known, being rational requires that participants invert averages to back out the underlying independent signals from the messages of the intermediaries.

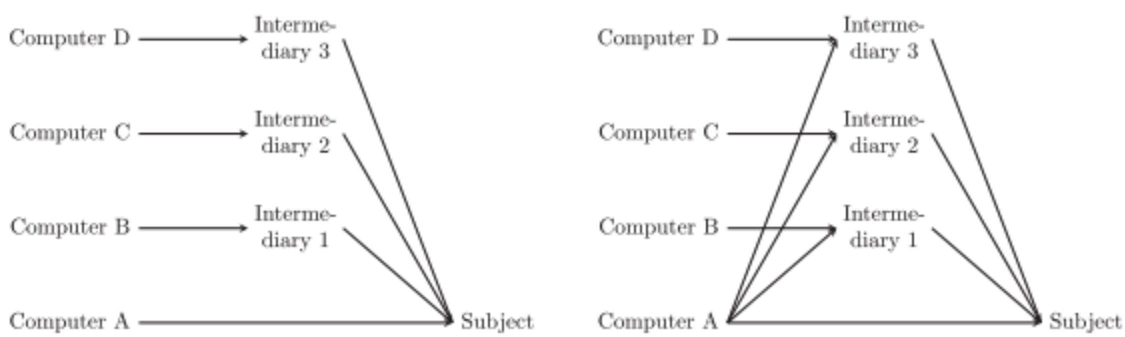


FIGURE 1

Uncorrelated (left panel) and *Correlated* (right panel) information structures

The identification strategy of EZ19 depends entirely on the identical informational content of the two sets of signals. Differences in beliefs between the *Correlated* and *Uncorrelated* conditions can only be explained by variations in the information structure since all other factors are held constant. Thus, comparing beliefs across *Correlated* and *Uncorrelated* allows for the identification of correlation neglect. Further, participants have complete knowledge of how their data are created. In principle, assuming the data-generating process is understood by the participant, this design choice leaves no room for mediating beliefs about the rationality of intermediaries or the precision of particular signals (e.g. not to consider any as outliers and thus less informative about the true state). In the chat analysis in Section 3, we consider the hypothesis that the data-generating process may be systematically misunderstood by certain types of individuals, as measured by mentions of outliers.

In the main experiment, subjects have to complete 10 independent belief formation tasks without feedback between each task. Three different randomised orders of tasks are used (outlined in EZ19 Online Appendix B). At the end of the experiment, participants are paid according to the accuracy of their belief in one randomly chosen task using a quadratic scoring rule (Selten, 1998), where variable earnings in Euro are given by $\pi = max[0, 10 - 160 \times (Belief/True\ state\ - 1)^2]$. Table 1, Part 1 is our novel addition of an individual naïveté parameter task (pre-belief), which is used to classify teams as follows: score = 0 if no rational members, = 1 if one rational member, = 2 if two rational members. Classification of teams is explained further below. Table 1, Part 2 gives an overview of the 10 tasks as designed by EZ19. It also shows the benchmarks of rational beliefs and "full correlation neglect" to indicate the direction and extent of a potential bias. "Full neglect" beliefs are defined as the average of the four messages that subjects received in the *Correlated* condition. EZ19 use the term "belief" to denote the mean of the belief distribution.

TABLE 1

Overview of the belief formation tasks

	True State	Computer A	Intermed. 1 uncorr.	Intermed. 2 uncorr.	Intermed. 3 uncorr.	Intermed. 1 corr.	Intermed. 2 corr.	Intermed. 3 corr.	Rational belief	Correlation neglect belief
Part 1	500	390	460	680	810	425	535	600	585	487.5
Part 2	10	12	9	10	0	10.5	11	6	7.75	9.88
	88	122	90	68	5	106	95	64	71.25	96.63
	250	179	295	288	277	237	234	228	259.75	219.38
	732	565	847	650	1351	706	608	958	853.25	709.13
	1000	1110	1060	629	1100	1085	870	1105	974.75	1042.38
	4698	1608	7240	4866	5526	4424	3237	3567	4810	3209
	7338	9950	1203	11322	11943	5577	10636	10947	8604.5	9277.25
	10000	2543	10780	6898	8708	6662	4721	5626	7232.25	4887.63
	23112	15160	21806	20607	47751	18483	17884	31456	26331	20745.5
	46422	12340	32168	49841	61293	22254	31091	36817	38910.5	25625.25

Notes. The reports of intermediaries 1-3 in the *Uncorrelated* condition directly reflect the draws of computers B–D. Signals from B-D are not directly observed but pass through intermediaries 1-3. There is no intermediary for computer A, which is directly observed. The rational belief is computed by taking the average of computers A–Ds' signals. The correlation neglect belief is given by the average of the signal of computer A and the reports of intermediaries 1-3 in the *Correlated* condition. Participants faced the 10 rounds in randomised order, identical across treatments. Part 1 is our innovation appended to the original design by Enke & Zimmermann (2019) shown in Part 2 and was completed individually by all participants. Part 2 was completed individually or as a team depending on the treatment assignment.

2.1.2. Treatments

We implement two main treatments, INDIVIDUAL and TEAM. Subjects are randomly assigned to one in four groups (2x2): (1) INDIVIDUAL with *Correlated*- or (2) INDIVIDUAL with *Uncorrelated* information structures and individual beliefs[9] or (3) TEAM with *Correlated* or (4) TEAM with *Uncorrelated* information structures and team beliefs. The INDIVIDUAL treatment is a replication of EZ19. The treatment closely follows the steps described above, and subjects are exposed to Enke and Zimmermann's original instructions (for Part 2, the main experiment, note that we added a Part 1 pre-belief individual assessment that preceded the main experiment, see Table 1 for details), software (only for INDIVIDUAL, adapted for TEAM) and experimental procedure (for which we thank Enke and Zimmermann). There is no feedback between Part 1 (individual belief) and Part 2 (individual or team belief, depending on treatment assigned).

In the TEAM treatment, a team consists of two participants. Partners remain fixed for the entire duration of the experiment. The TEAM pairs face the same set of decisions as the participants in INDIVIDUAL, but teams have to reach a joint decision before moving to the next estimation task. When a new estimation task is presented to the team, they have to make proposals and then each enter a belief for the team's decision. They have to agree. That is, the beliefs of the two members of the team have to match in order to be submitted successfully.

To facilitate reaching a unanimous decision, team members can exchange messages in free-form, real-time chat. Participants can chat even before entering the first proposal for the team decision. Once a proposal has been entered on the screen, each team member can confirm the decision she has proposed. If both team members confirm the same belief, they can move on to the next estimation task. If there is still disagreement, team members can adjust their choices (typically after further discussing what to agree upon). After receiving the information structures, a participant has seven minutes to state a belief. We extend EZ19's five-minute limit to seven minutes in the present experiment to allow for subjects in the TEAM treatment to chat with their partner without inducing time pressure. This time limit is constant across treatments.

[9] Treatments 1 and 2 are a replication of EZ19's baseline treatments.

2.1.3. *Innovation: Classifying teams (and individuals) using a proxy measure of individual naïveté*

Before subjects complete the 10 belief formation tasks (see Table 1, Part 2), we run a single incentivised individual-level task to measure naïveté (see Table 1, Part 1). This is simply an additional estimation task that we generated identically to the ones designed by Enke and Zimmermann, framed according to whether a participant was assigned to *Correlated* or *Uncorrelated*. We decided to measure individual naïveté to account for potential differences between teams and individuals since teams are composed of individuals of varying naïveté. Both individuals and teams completed Part 1 to keep treatments comparable.

2.1.4. *Team decision-making process in the chatbox*

The perception of the overall decision-making process in a team may affect responses. For example, a subject may feel they gave up in the decision-making process at some point in the experiment. A naïve team member, who may be less certain about her beliefs, may delegate to her partner. Such delegation may be advantageous if her partner is rational. However, a subject receives no information about the extent of her partner's naïveté. We collect data on the time taken to reach agreement, the beliefs submitted by each member of the team, and the chat over text messages within each team to provide insight into the decision-making process of teams. For example, the chatbox allows team members to signal their lack of confidence in their math ability to their partner and for this data to be coded as such.

2.2. *Hypotheses*

INDIVIDUAL (Replication)

Below, we provide Enke and Zimmermann's derivation of beliefs formed by rational and naïve individuals (EZ19, p. 319) before extending the framework to teams.

In the information structure presented in Figure 1, the fictitious computers generate four i.i.d. signals $s_h \sim \aleph(\mu, (\mu/2)^2)$ (truncated at $(0, 2\mu)$) for $h \in \{1, \dots, 4\}$. In the *Correlated* condition, participants are shown messages s_1 and $\tilde{s}_h = (s_1 + s_h)/2$ for $h \in \{2, 3, 4\}$. When asked to estimate μ, a rational decision-maker would back out the underlying independent signals from the messages $\tilde{s}_h$ and calculate the mean rational belief as $b_B = \sum_{h=1}^{4} s_h/4$, which by design is also equal to the rational belief in the *Uncorrelated* condition.

Suppose the decision-maker suffers from correlation neglect, *i.e.* she does not fully take into account the degree to which $\tilde{s}_h$ reflects s_1, but instead treats $\tilde{s}_h$ as more or less independent. EZ19 label such a decision naïve and let her degree of naïveté be characterised by the parameter $\chi_{indiv} \in [0, 1]$ such that $\chi_{indiv} = 1$ implies full neglect. A naïve agent extracts s_h from $\tilde{s}_h$ according to the following rule:

$$\hat{s}_h = \chi\tilde{s}_h + (1 - \chi_{indiv})s_h = s_h + \frac{1}{2}\chi_{indiv}(s_1 - s_h) \tag{1}$$

where $\hat{s}_h$ for $h \in \{2,3,4\}$ denotes the agent's (potentially biased) inference of s_h. She consequently forms mean beliefs according to

$$b_{CN,indiv} = \frac{s_1 + \sum_{h=1}^{3} \hat{s}_h}{4} = \bar{s} + \frac{3}{8}\chi_{indiv}(s_1 - \bar{s}_{-1}) \tag{2}$$

where $\bar{s} = (\sum_{h=1}^{4} s_h)/4$ and $\bar{s}_{-1} = (\sum_{h=2}^{4} s_h)/3$. Therefore, a (potentially partially) naïve belief is given by the rational belief $\bar{s}$ plus a belief bias component that depends on the extent of naïveté parameter χ_{indiv} and the magnitude of the common source signal (Computer A) relative to the other three signals (Intermediaries 1-3).

Hypothesis 1. *Assuming that $\chi > 0$, beliefs in the* Correlated *treatment exhibit an overshooting pattern. Given a high common source signal, i.e. $s_1 > \bar{s}_{-1}$, beliefs in the* Correlated *treatment are biased upward compared to the* Uncorrelated *treatment. Conversely, if $s_1 < \bar{s}_{-1}$, beliefs in the* Correlated *condition are biased downward.*

By partly neglecting the redundancies among the signals, the subject double counts the first signal (Computer A) so that beliefs are biased in the corresponding direction. At the same time, the beliefs of a naïve agent remain statistically unbiased with a zero expected error. This means that the signal from the common source may swing beliefs of a naïve decision-maker either too high or too low on an individual task, but on average naïve agents are correct.

TEAM (Novel extension)

Hypothesis 2. *Teams will form beliefs that are less biased towards the common source signal compared to individual decision-makers. If a team consists of at least one rational member with naïveté parameter* $0 \leq \chi_{indiv} < 0.5$*, the average of the two members' belief bias components, which form the team belief bias component and reflect the joint naïveté parameter* χ_{team}*, will be lower than that of a naïve individual forming beliefs alone,* $b_{CN,indiv} > b_{CN,team} \geq b_B$*. If the naïve member defers to the rational partner (because they are not confident in their ability to identify/solve the problem), then heterogeneous teams (with at least one rational member) will form beliefs at least as good as rational individuals.*

Teams consequently form mean beliefs according to

$$b_{CN,team} = \frac{s_1 + \sum_{h=1}^{3} \hat{s}_h}{4} = \bar{s} + \frac{3}{8}\chi_{team}(s_1 - \bar{s}_{-1}) \tag{3}$$

where $\chi_{team} \leq \chi_{indiv}$ on average.

It is an open question whether randomly assigning individuals to teams (of two players) – in a population with a bimodal distribution of rational and naïve subjects – would produce more sophisticated play in settings where individuals exhibit substantial correlation neglect. There are three simple combinations worth considering: (1) partnering two rational players, (2) partnering a rational and a naïve player, (3) partnering two naïve players. Partnering two rational players is unlikely to swing beliefs away from the distribution of individual sophisticated players. Such teams should reach an agreement quickly without conflict. A more interesting case is the partnering of a naïve subject with a rational subject, since, in a setting where a team must reach a compromise, this may debias beliefs. However, *ex ante*, agents have no information about the sophistication of their teammate. Moreover, a lack of feedback would preclude Bayesian updating by a subject about her partner's perceived sophistication. It is also not clear, *ex ante*, whether the naïve subject would quickly defer to her more sophisticated teammate. It is possible that naïve agents are less certain of their beliefs and may be more willing to delegate to their partner when a compromise must be reached. If not, heterogeneous teams would likely experience conflict and take longer to reach a shared belief than individual agents and homogenous teams.

2.3. Procedural details

Participants receive detailed written instructions about the task and the payment incentive structure.[10] At the beginning of each session, instructions are handed out individually together with control questions on the computer that tested the understanding of the experimental instructions. Specifically, they are told about the signals of the four computers, how these signals map onto the reports of the intermediaries, and the fact that the four computers are of identical quality. For example, the instructions include the applicable panel (*Correlated* or *Uncorrelated*) from Figure 1. The instructions also include an example that consists of four exemplary computer signals as well as the respective messages of the three intermediaries, given a particular state of the world. Participants are given a visual representation of an exemplary distribution function, and the concept of unbiasedness is explained in intuitive language. General instructions about the experiment are read out loud. Additionally, participants complete control questions with a special focus on the information structure (*Correlated* or *Uncorrelated*). In all treatments, participants have to calculate the reports of intermediaries 1 and 3 given exemplary signals of the four computers to ensure that they understand the correlated (or uncorrelated) nature of the messages. Subjects only participate in the experiment once they have answered all control questions correctly. They can raise their hand if they have questions. Whenever a subject appears to have trouble solving the control questions, an experimenter approaches them and clarifies open questions.[11] At the end of the experiment, we conduct a questionnaire on socio-demographics.

The experiment was conducted at the Decision Lab of the Max Planck Institute for Research on Collective Goods in Bonn. Subjects were mostly students from the University of Bonn and were

[10] See Appendix for a translation of the instructions and control questions.
[11] No subjects had to be excluded due to inadequate understanding of the task.

recruited using the online recruitment system ORSEE by Greiner (2004). No subject participated in more than one session. The experiment was run using the experimental software z-Tree (Fischbacher, 2007). Table 2 provides an overview of treatments. For the replication of EZ19 baseline treatments with individuals, we conducted three sessions with between 22 to 29 subjects per session, yielding a total of 76 participants in treatment INDIVIDUAL. For treatment TEAM, we conducted five sessions with between 22 to 30 participants per session yielding 68 teams (136 subjects). A total of 212 subjects participated in our study. A session was run as INDIVIDUAL or TEAM, and, within the session, participants were randomised to either *Correlated* or *Uncorrelated.* Sessions lasted about 1.5 hours, and average (total) payment equalled 16.94 Euro. To determine variable earnings, one decision out of the 10 tasks was randomly chosen to be paid, with a maximum of 10 Euro. The show-up fee was 5 Euro. The demographic questionnaire was 5 Euro. The additional payment of maximum 2 Euro for completing the individual naïveté parameter task (pre-belief) rendered the average payment consistent with the laboratory's guidelines for participant compensation.[12]

On average, teams earned more than individuals. Of particular interest is the significantly greater average earnings of teams compared to individuals in the *Correlated* condition (two-sample Wilcoxon rank-sum test, p=0.0013). While the differences in earnings within INDIVIDUAL (*Correlated* vs *Uncorrelated*) and within TEAM (*Correlated* vs *Uncorrelated*) are in the expected direction (suggestive of the cost of correlation neglect), there is no significant difference between *Correlated* and *Uncorrelated* in the actual earnings within TEAM and within INDIVIDUAL, respectively. This result is in contrast to EZ19 and may be due to lack of power because of small sample sizes.[13]

[12] This motivated the higher average payment per participant compared to €11.60 by Enke and Zimmermann at UniBonn Lab a couple of years prior.
[13] Data collection was limited by the start of Covid-19 lockdown which closed labs in Germany.

TABLE 2

Treatment overview

Treatment	# of observations	Ave. earnings (Euro)
INDIVIDUAL *Correlated*	37	4.59
INDIVIDUAL *Uncorrelated*	39	5.23
TEAM *Correlated*	33	6.89
TEAM *Uncorrelated*	35	7.13

Notes. Horizontal lines indicate which treatments were randomised within the same experimental sessions. Average earnings is the profit from the round randomly selected for payment in the *main estimation task only*. i.e., these earnings illustrated exclude show-up fee, questionnaire fee, and pre-belief estimation task which are also paid out.

3. RESULTS

3.1. Beliefs across Correlated and Uncorrelated conditions

INDIVIDUAL. Table 3 Panel A provides summary statistics for all tasks and shows that in four out of 10 estimation tasks, beliefs differ significantly at the 5% level between *Uncorrelated* and *Correlated* conditions in the INDIVIDUAL treatment, in the direction predicted by correlation neglect. In eight out of 10 tasks, beliefs differ significantly at the 10% level. This result replicates the EZ19 finding of nine out of 10 tasks but at a lower significance threshold.[14] The bias appears to be stable across tasks and appears not to depend on the magnitude of the true state. *Correlated* individuals earned 4.59 Euro on average compared to 5.23 Euro by *Uncorrelated* individuals (Table 2). In contrast to EZ19, the earnings difference within INDIVIDUAL *Correlated* versus *Uncorrelated* was not significant (p=0.3432, Wilcoxon rank-sum test). This is one example of where increasing the sample size would be valuable.

TEAM. Table 3 Panel A shows that in two out of 10 estimation tasks, beliefs differ significantly at the 5% level between *Uncorrelated* and *Correlated* conditions in the TEAM treatment. In four out of 10 tasks, beliefs differ significantly at the 10% level. Since teams are about half as likely to hold beliefs in the direction of correlation neglect compared to individuals, this suggests that teams may be more rational in this setting. In Table 2, average earnings accrued to *Correlated* teams is 6.89

[14] EZ19 had larger numbers of observations. They also pooled the baseline and high-stakes treatments.

Euro compared to 7.13 Euro to *Uncorrelated* teams. The earnings difference within TEAM *Correlated* versus *Uncorrelated* is not significant (p=0.9657, Wilcoxon rank-sum test). Consistent with EZ19, we observe that beliefs in the *Correlated* condition exhibit larger within-condition belief variance than those in the *Uncorrelated* condition. Specifically, for INDIVIDUAL, in seven out of 10 tasks, there is a higher variance in beliefs in *Correlated*. For TEAM, in nine out of 10 tasks, there is a higher variance in beliefs in *Correlated*.

INDIVIDUAL vs TEAM. If it were random whether individuals or teams had a larger absolute difference in median belief (per task) between *Correlated* and *Uncorrelated* conditions, then one would expect for the 10 tasks that in five cases, the absolute difference in median belief between *Correlated* and *Uncorrelated* would be greater for the INDIVIDUAL treatment and in five cases, the absolute difference in median beliefs would be greater for the TEAM treatment. Table 3, Panel B, reports the difference-in-difference analysis and Table 3, Panel C, provides a robustness check of Panel B to show that the results do not change when the normalised absolute difference is considered.

TABLE 3

Panel A: Correlation neglect by belief formation task

True state	Rational belief	Corr. neglect belief	Median belief				Rank sum test (p-value)	
			TEAM		INDIVIDUAL		*Correlated vs. Uncorrelated*	
			Uncorrelated	*Correlated*	*Uncorrelated*	*Correlated*	TEAM	INDIVIDUAL
10	7,75	9,88	9	9,25	8	9	0,7339	0,2967
88	71,25	96,63	71,25	80	71,25	80	0.0306**	0.0001***
250	259,75	219,38	260	260	260	258,75	0,5748	0.0725*
732	835,15	709,13	829	800	853	780	0,2795	0.0143**
1000	974,75	1042,38	1000	977,5	979	1010	0,413	0.0698*
4698	4810	3209	4810	4810	4810	4560	0.0758*	0.0196**
7338	8604,5	9277,25	9500	9064	8899	9277,5	0,3765	0,8434
10000	7232,25	4887,63	7232,25	6530	7232,25	6658	0.0045***	0.0173**
23112	26331	20745,5	25000	22000	25800	20995,75	0.0947*	0.0517*
46422	38910,5	25625	38910,5	38886	38910,5	35300	0,1597	0.079*

Notes. Panel A of the table presents an overview of beliefs in the *Uncorrelated* and *Correlated* information conditions across the 10 estimation tasks for the INDIVIDUAL and TEAM treatments, respectively. The p-values refer to a Wilcoxon rank-sum test between beliefs in the *Correlated* and *Uncorrelated* conditions. For reference, we provide the benchmarks of rational and fully naïve beliefs. See Table 1 for details of the computation of the rational and correlation neglect benchmarks. Note that participants encountered the 10 tasks in a randomised order.

Panel B: Difference-in-difference

True state	Rational belief	TEAM Median of abs(belief - rational belief) *Uncorrelated*	*Correlated*	Diff *C-U*	INDIVIDUAL Median of abs(belief - rational belief) *Uncorrelated*	*Correlated*	Diff *C-U*	Diff-in-Diff	INDIV vs TEAM
10	7.75	1.25	1.55	0.3	1.11	1.75	0.64	0.34	I > T
88	71.25	3.75	8.75	5	3.25	8.75	5.5	0.5	I > T
250	259.75	8.25	24.5	16.25	9.75	26.45	16.7	0.45	I > T
732	835.15	35.15	47	11.85	34.85	78.15	43.3	31.45	I > T
1000	974.75	49.25	52.25	3	39.75	67.75	28	25	I > T
4698	4810	190	375	185	190	610	420	235	I > T
7338	8604.5	1276	895.5	-380.5	1341.5	947.5	-394	-13.5	
10000	7232.25	117.75	635	517.25	267.75	980.75	713	195.75	I > T
23112	26331	2919	4331	1412	2169	5716.47	3547.47	2135.47	I > T
46422	38910.5	863.5	6089.5	5226	1089.5	8732.25	7642.75	2416.75	I > T

Notes. Panel B of the table presents an overview of the absolute difference in belief and rational belief in the *Uncorrelated* and *Correlated* conditions across the 10 estimation tasks for the INDIVIDUAL and TEAM treatments, respectively. We examine whether individuals or teams converged more towards rational belief.

Panel C: robustness check of Panel B with normalised absolute difference

True state	Rational belief	TEAM Median norm abs(belief - rational belief) *Uncorrelated*	*Correlated*	Diff *C-U*	INDIVIDUAL Median norm abs(belief - rational belief) *Uncorrelated*	*Correlated*	Diff *C-U*	Diff-in-Diff	INDIV vs TEAM
10	7.75	0.161	0.200	0.039	0.143	0.226	0.083	0.044	I > T
88	71.25	0.053	0.123	0.070	0.046	0.123	0.077	0.007	I > T
250	259.75	0.032	0.094	0.063	0.038	0.102	0.064	0.002	I > T
732	835.15	0.042	0.056	0.014	0.042	0.094	0.052	0.038	I > T
1000	974.75	0.051	0.054	0.003	0.041	0.070	0.029	0.026	I > T
4698	4810	0.040	0.078	0.038	0.040	0.127	0.087	0.049	I > T
7338	8604.5	0.148	0.104	-0.044	0.156	0.110	-0.046	-0.002	
10000	7232.25	0.016	0.088	0.072	0.037	0.136	0.099	0.027	I > T
23112	26331	0.111	0.164	0.054	0.082	0.217	0.135	0.081	I > T
46422	38910.5	0.022	0.157	0.134	0.028	0.224	0.196	0.062	I > T

Notes. Panel C of the table presents an overview of the normalised absolute difference in belief and rational belief in the *Uncorrelated* and *Correlated* conditions across the 10 estimation tasks for the INDIVIDUAL and TEAM treatments, respectively. This table is a robustness check of the above panel.

We observe that in nine cases, the difference between *Correlated* and *Uncorrelated* is smaller for teams than individuals. In one case, it is smaller for individuals (task 7, true state = 7,338). The interpretation would be 9/10 smaller, 1/10 larger. We conduct a binomial test to investigate whether these differences in medians are significantly different from random. A two-sided binomial test of the probability of observing at least 9/10 or 1/10 or less, conditional on expecting 5/10, is significant at the 5% level; $\Pr(k \leq 1 \text{ or } k \geq 9) = 0.0215$. This result supports the hypothesis that, on average, *Correlated* and *Uncorrelated* beliefs converge more towards rational for teams than individuals.

3.2. Heterogeneity

We have established suggestive evidence of correlation neglect *on average* for both individuals and teams. The average patterns may hide heterogeneity. To investigate this, we use Enke and Zimmermann's measure of an individual's belief type that allows for the estimation of naïveté parameter χ and extend this to joint naïveté for team beliefs. The experimental design, in combination with the simple model of belief formation, introduced above in equations (1)-(3), allows us to estimate an individual's or team's naïveté χ, respectively. For each belief, the naïveté parameter χ is calculated by solving for χ from equation (2) for individuals or equation (3) for teams. The median of the naïveté values forms an estimator for the subject-level naïveté parameter:

$$\hat{\chi}_i \equiv med\left(\tilde{b}_i^j\right) = med\left(\frac{8\left(b_i^j - \bar{s}^j\right)}{3\left(s_1^j - \bar{s}_{-1}^j\right)}\right)$$

As in EZ19, we analyse a sample of censored data in which the absolute value of naïveté chi was ≤ 3. In our experiment, 5.28% of the pooled data are classified as outliers using the same rule as EZ19.[15]

Replication of EZ19, Table 4 Column (1). Table 4 provides the regression analysis of correlation neglect with individual and team decision-making, respectively. Column (1) demonstrates a successful replication of EZ19. The coefficient on the *Correlated* dummy variable is nearly identical to EZ19 and highly significant.

Prechi (Part 1) proxy for individual team members' naïveté *is a good predictor of performance, Table 4 Column (2).* In column (2), we demonstrate that for individuals, the single pre-task individual naïveté measure that we added to the experimental design (see Table 1, Part 1 for details) significantly predicts naïveté in the original main 10 tasks of EZ19 (see Table 1, Part 2). This supports using our innovation of the pre-task (Part 1) as an individual proxy measure of naïveté to classify the composition of teams.

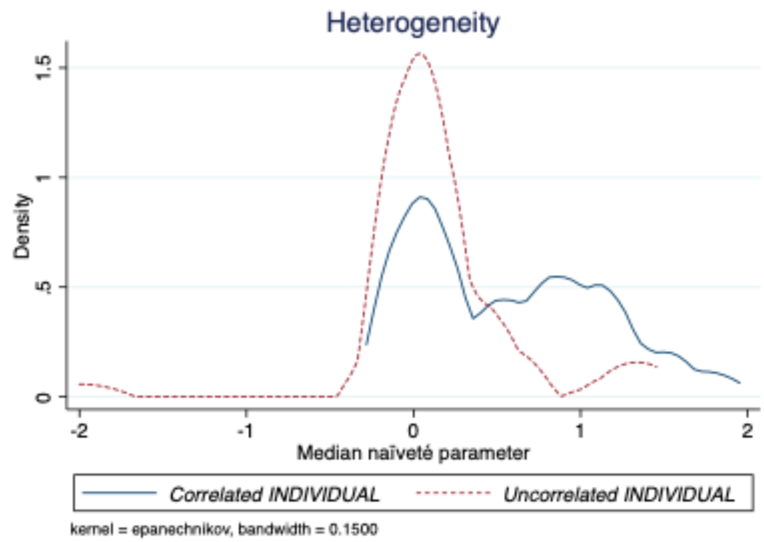
Team results, Table 4 Column (3). Column (3) demonstrates the existence of correlation neglect amongst teams. The magnitude of the coefficient is half as large as it is for individuals indicating

[15] About 4% of the EZ19 sample was excluded from their analysis as outlier beliefs.

an attenuated bias, however the difference remains statistically significant at the 5% level. Next, we explore the underlying distributions of individuals and teams to better understand the differences in belief formation.

Distributions of INDIVIDUAL and TEAM, Figure 2 and Figure 3. Figure 2 shows kernel density estimates of the distribution of naïveté parameters for both the *Correlated* and the *Uncorrelated* conditions in the INDIVIDUAL treatment. Figure 3 illustrates the same comparison for the TEAM treatment. Consistent with EZ19, there is a density spike around zero in Figure 2 and Figure 3 for the *Uncorrelated* condition, indicating that the majority of subjects behave approximately rationally. This pattern holds for both the INDIVIDUAL and TEAM treatments. The *Correlated* condition, however, exhibits substantial heterogeneity. The *Correlated* INDIVIDUAL treatment has two main peaks around the rational benchmark (χ=0) and around full neglect (i.e. χ=1). This suggests the presence of different updating types – in particular, one of full correlation neglect. A two-sample Kolmogorov-Smirnov test for equality of distributions between *Correlated* and *Uncorrelated*, conditional on INDIVIDUAL, rejects the hypothesis that the two distributions are the same (*Uncorrelated*<*Correlated*, D=0.4096, p=0.002), and is suggestive of correlation neglect among individuals. Team beliefs in the *Correlated* condition exhibit a similar bimodal pattern to individuals, suggestive of rational and naïve updating types. However, there are differences worth noting. The second peak of the *Correlated* TEAM distribution (Figure 3) appears to be centred around χ=0.75. Visually, it is left-shifted when compared to the second peak of *Correlated* INDIVIDUAL, which is centred around χ=1 (full neglect). A two-sample Kolmogorov-Smirnov test for equality of distributions between *Correlated* and *Uncorrelated*, conditional on TEAM, fails to reject the hypothesis that the distributions are the same at the 5% level (*Uncorrelated*<*Correlated*, D=0.2797, p=0.070). Finally, we conduct two-sample Kolmogorov-Smirnov tests for equality of distributions between INDIVIDUAL and TEAM, conditional of *Correlated* or *Uncorrelated*. For *Correlated*, the distribution of TEAM is closer to rational (zero) than INDIVIDUAL, though not significant at the 5% level (TEAM<INDIVIDUAL, D=-0.2703, p=0.078).[16] Reassuringly, for *Uncorrelated*, the distribution of TEAM is indistinguishable from the distribution of INDIVIDUAL (TEAM<INDIVIDUAL, D=-0.0769, p=0.804).

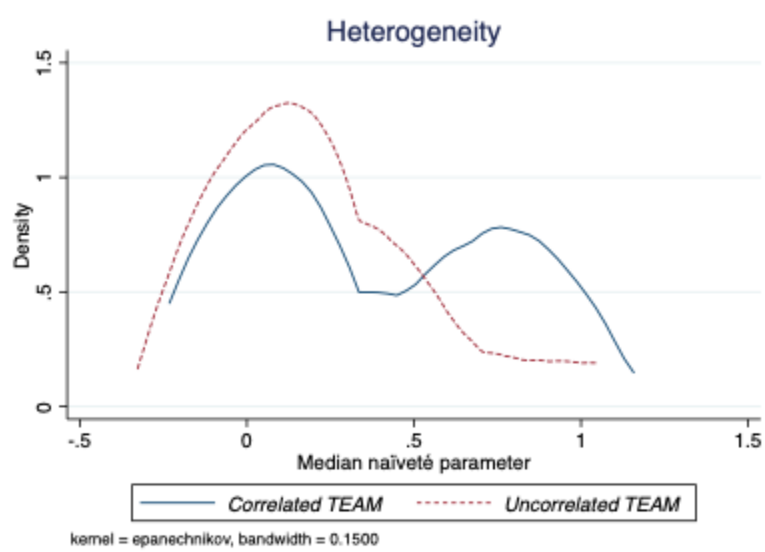
[16] Potentially due to small sample size.



Notes. The two kernels show the distributions of naïveté in the *Correlated* and *Uncorrelated* conditions for the INDIVIDUAL treatment.

FIGURE 2

Kernel density estimates of median naïveté parameters for individuals



Notes. The two kernels show the distributions of naïveté in the *Correlated* and *Uncorrelated* conditions for the TEAM treatment.

FIGURE 3

Kernel density estimates of median naïveté parameters for teams

Result 1: *Team beliefs exhibit a similar bimodal pattern to individuals, suggestive of rational and (to a lesser extent) naïve belief updating types.*

Team composition matters, Table 4 Column (4) and Figure 4. In Column (4), the *Correlated* dummy variable loses statistical significance and is close to zero after we control for the team composition. We expected substantial heterogeneity in the composition of teams. We examine team composition according to our classification specified above in Section 2.4. Using the Part 1 individual naïveté measure, which we call prechi, the first individual member of a team (Player 1) is classified as rational if the abs(prechi) < 0.5 and naïve otherwise. Likewise, the second member of a team (Player 2) is classified as rational if abs(other_prechi) < 0.5, and naïve otherwise. We take the average prechi for the team of two and compare this to prechi for individuals. The distribution of prechi is indistinguishable between individuals and teams at the 5% level (Two-sample Wilcoxon rank-sum test, z=-1.785, p=0.0743). We generated the variable team score, which takes on the following values: 0 if the team has two naïve subjects (19.12% of teams), 1 if a single subject is rational (50% of teams), and 2 if both subjects are rational (30.88% of teams). Table 5 shows the percentage of individuals classified as rational or naïve according to this Part 1 assessment used to classify our teams. Compared to teams with two naïve members, teams with at least one rational member form significantly more rational beliefs.

This heterogeneity is apparent in the density plots of Figure 4, which show that teams with two naïve members are demonstrably more likely to suffer correlation neglect than teams with at least one rational member (Two-sample Kolmogorov-Smirnov test, D=-0.7619, p<0.001). Column (4) in Table 4 suggests that one rational team member is sufficient to make beliefs significantly more rational, compared to teams with two naïve members. Having at least one rational team member reduces naïveté by about 0.5 units of χ. Assuming that group interactions are neutral, the group should be able to reach a correct answer if at least one member proposes it. Failure to reach the Lorge-Solomon baseline indicates that interactions actually make teams perform worse than their members would do acting on their own (Cooper and Kagel, 2005). Figure 4 and Table 4 Column (4) shows that having just 1 rational team member is sufficient to reduce naïveté. This suggests an advantage of team decision-making in a heterogeneous population since there is a high chance that a naïve individual will be paired with a rational teammate, improving the joint rationality of beliefs.

INDIVIDUAL versus TEAM, Table 4 Column (5). In Column (5), the treatment dummy for TEAM is small and not significant. This indicates *in aggregate* there was no significant statistical difference between the naïveté of individuals and teams in the full sample.

TABLE 4

Heterogeneity in correlation neglect

	Dependent variable:							
	Naïveté χ							*Response time*
	INDIVIDUAL		TEAM		All	*Correlated* INDIV.		*Correlated* All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0 if *Uncorrelated*, 1 if *Correlated*	0.37***	0.28***	0.16**	-0.04	0.27***			
	(0.12)	(0.09)	(0.08)	(0.09)	(0.07)			
Response time (in minutes)						-0.02***	-0.01***	
						(0.005)	(0.004)	
Pre-task individual naivete		0.44***					0.38***	
		(0.09)					(0.08))	
0 if INDIVIDUAL, 1 if TEAM					-0.06			
					(0.07)			
TEAM = naïve + naïve								1.15
								(3.96)
TEAM = rational + naïve				-0.51***				12.28***
				(0.10)				(2.95)
TEAM = rational + rational				-0.54***				21.13***
				(0.13)				(3.12)
Constant	0.19**	0.37	0.23***	0.74***	0.24***	0.86	1.09**	27.23***
	(0.08)	(0.40)	(0.04)	(0.13)	(0.07)	(0.62)	(0.48)	(2.06)
Additional controls	No	Yes	No	No	No	Yes	Yes	No
Observations	726	726	638	638	1364	354	354	660
R-squared	0.05	0.26	0.02	0.09	0.03	0.16	0.26	0.23

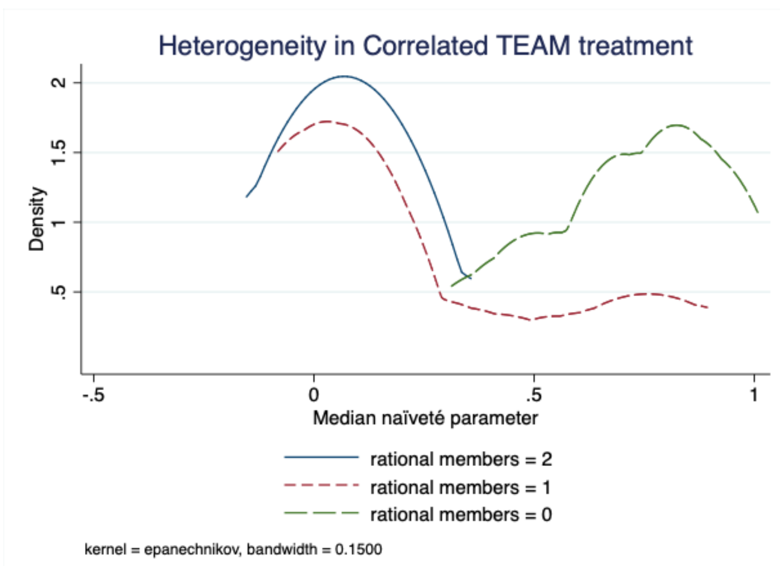
Notes. OLS estimates, robust standard errors (clustered at subject level) in parentheses. The table analyses the determinants and correlates of subjects' naïveté as implied in each of the 10 beliefs. In columns (1)-(2), observations include all subjects from the *Correlated* and *Uncorrelated* conditions in the INDIVIDUAL treatment. In columns (3)-(4), observations include all subjects from the *Correlated* and *Uncorrelated* conditions in the TEAM treatment. In column (5), the sample includes all subjects from the full sample. In columns (6)-(8), observations include subjects from *Correlated* conditions only. Additional controls include age, gender, monthly income, GPA, last math grade, marital status fixed effects, and task fixed effects. All regressions exclude extreme outliers with the absolute value of $\chi > 3$. Significance is starred, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 5

Distribution of (Part 1) prechi naïveté and response times

	% of sample rational (Part 1)	% sample naïve (Part 1)	% of sample	Ave. response time (min)
INDIVIDUAL *Correlated*	51.43	48.57		27,49
INDIVIDUAL *Uncorrelated*	69.23	30.77		16,67
TEAM *Correlated*	23.33	76.67		
TEAM *Uncorrelated*	87.10	12.90		
TEAM *2 naïve subjects*			19,12	29,1
TEAM *1 naïve + 1 rational*			50	34,61
TEAM *2 rational subjects*			30,88	31,81
TEAM *Correlated, 2 naïve subjects*			44,29	28,69
TEAM *Correlated 1 naïve + 1 rational*			51,43	39,25
TEAM *Correlated 2 rational subjects*			4,29	47,97

Notes. This table shows the percentage distribution of the sample that can be classified as "prechi rational", *i.e.* absolute value of prechi<0.5, according to our innovation of the pre-task assessment used to examine heterogeneity in the composition of teams. For teams, we take the average prechi of the individual team members. We examine the same prechi (Part 1) measure amongst individuals to check that the randomisation was successful. The table also shows average response times in minutes as a proxy for cognitive effort and breaks it down for the three different types of teams: naïve+naïve, naïve+rational, and rational+rational.



Notes. The three kernels densities show the distributions of naïveté in the *Correlated* condition for the TEAM treatment. Before the 10 joint tasks, each subject completed an individual belief formation task, and the implied naïveté of the belief was computed. Rational = 1 if abs(prechi) < 0.5, 0 otherwise. Teams were given a team score: = 0 if both members were naïve, = 1 if one member was naïve and the other rational, and = 2 if both were rational.

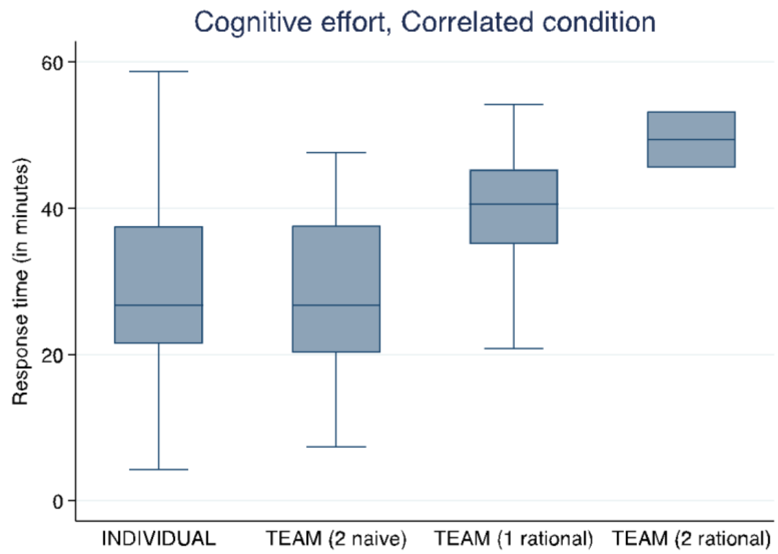
FIGURE 4

Kernel density estimates of median naïveté parameters for teams

Result 2: *Team decision-making with one or two rational members leads to significantly more rational beliefs than teams with only naïve members, and individual decision-making, respectively.*

3.3. Cognitive effort

Table 4 Columns (6)-(7) show that response time, a proxy for cognitive effort, is significantly associated with individuals forming more rational beliefs in the *Correlated* treatment. This is consistent with EZ19. We explore cognitive effort as a mechanism for why teams with at least one rational member perform better than individuals. Column (8) shows that in the *Correlated* condition teams with 2 naïve members do not spend more cognitive effort than individuals on average, but teams with at least one rational member do expend more effort. In Column (8), the three coefficients on "TEAM =…" are jointly significantly different from zero, $F(3, 69)=17.85$, $p<0.001$. The coefficient on "TEAM=naïve+naïve" is not significantly different from zero, $F(1, 69) = 0.08$, $p=0.77$. The coefficients for "TEAM=naïve+naïve" and "TEAM=rational+naïve" are not equal, $F(1, 69)=7.78$, $p=0.007$. The coefficients for "TEAM=rational+naïve" and "TEAM=rational+rational" are likewise not equal, $F(1, 69)=7.86$, $p=0.007$. This heterogeneity in cognitive effort between the different groups can also be seen visually in Figure 5. Table 5 shows the average response times. If we consider the *Correlated* condition only, individuals spent about 27 minutes on task while teams spent about 36 minutes, considerably longer on average. In the *Correlated* condition, compared to teams with two naïve members who spent about 29 minutes on task, teams with one rational member spent about 10 extra minutes, and teams with two rational members spent about 20 extra minutes to reach a decision. Thus, having at least one rational member on the team is associated with greater cognitive effort.



Notes. The four box plots show the distribution of response time (minutes) in the *Correlated* condition – which faced the problem of correlated information – for the INDIVIDUAL and TEAM treatment; in particular, teams with 2 naïve members, 1 rational and 1 naïve member, and 2 rational members, respectively. The Part 1 proxy for individual naïveté was used to classify team composition.

FIGURE 5

Box plots of response time for individuals and teams of varying composition

Result 3: *Teams with at least one rational member spend significantly more cognitive effort on the estimation tasks than teams with only naïve members, and individuals, respectively.*

3.4. Decision process: Analysis of the chat text messages

So far, we have established that there are significant differences between *Correlated* and *Uncorrelated* conditions for individuals and teams, respectively; secondly, that team beliefs converge more towards rational than individual beliefs; thirdly, that beliefs are bimodally distributed suggestive of different types, *i.e.* rational and full neglect belief types; fourthly, that communication plays a role, evident in differences in longer response times when there is a rational team member present. These findings prompt identifying in the chat content what kind of communication influences the naïveté of beliefs formed by teams.

We analysed the chat text between teammates to understand the deliberation process. We read through parts of the dialogues and identified the 12 categories of interest in Table 6. An independent research assistant coded the chats, and the researchers double-checked and resolved ambiguous cases. Table 6 provides descriptive statistics of the categories that communications tended to be characterised by. We refer to a single message or communication of text between teammates in the chatbox as a "comm". In our sample of 68 teams, 41% signalled poor math ability[17] at least once in 10 rounds. All teams asked at least one question, with an average of 13 questions.[18] All teams made statements of agreement; 15 of such comms on average.[19] Disagreement characterised only 26% of teams, with an average of less than one disagreement per team. This lack of disagreement is surprising and may suggest that, in the case of mixed teams with only one rational player, the naïve partner tended to defer to her rational counterpart. 96% of teams made comms of computational logic, with an average of 18 such comms, varying widely from 0 to 67 comms.[20] In addition, 62% of teams mentioned an outlier. Such findings suggest that more than half of teams were uncertain of the description of the true state of the world and unconvinced about the independence of the four signals, despite a transparent data-generating process provided in the instructions. More teams mentioned an outlier in *Uncorrelated* (66%) than *Correlated* (58%). 54% of teams with two naïve members mentioned an outlier, while 64% of teams with at least one rational member mentioned an outlier. Such impressions could not have been captured in the original experiment and imply that biased beliefs observed may represent something besides correlation neglect. One explanation for this may be collective statistical illiteracy. Future research could examine the validity of the Enke-Zimmermann model of correlation neglect.

46.15% of teams with two naïve members signalled poor math ability or lack of confidence, while 44% of teams with one rational member, and 33% of teams with two rational members, signalled lack of confidence, respectively. This is aligned with our expectation that naïve individuals are less confident about their ability to identify and solve the problem of correlation neglect than rational individuals. Teams with one rational member made 23 comms characterised as computational logic statements, while teams with two naïve members and teams with two rational members made 13 such comms on average.

[17] E.g. *"ich vertraue meinem Mathe-Denken nicht"* (I don't trust my math).
[18] E.g. "*oki sollen wir das eingeben? odeer hast du n vorschlag?"* (Should we enter that? Or do you have a suggestion?)
[19] E.g. "yes", "ok", "sounds good", "perfect!"
[20] E.g. "hello, I would like to say we calculate B, C, and D and then make an estimate".

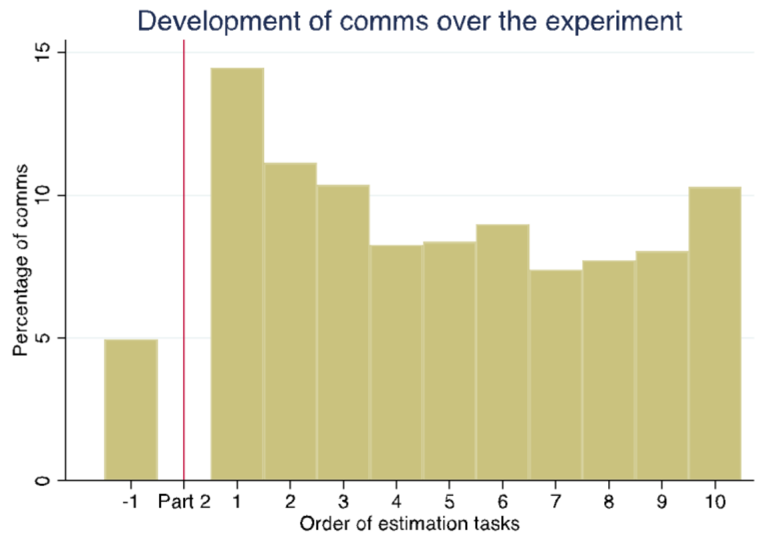
TABLE 6

Average frequency of types of communications between teams over 10 rounds

Categories of comms	Mean	Std. Dev.	Min	Max	% of teams (N=68)
Signal poor math ability (not confident)	0.69	1.14	0	6	41.18
Asked a question	13.08	7.06	2	31	100.00
Agreement	15.01	5.52	5	33	100.00
Disagreement	0.43	0.89	0	4	26.47
Mention of an outlier	2.54	3.33	0	16	61.76
Irritation/bored with task	0.62	1.34	0	8	29.41
Computational logic	18.03	18.84	0	67	95.59
Proprosal outside rational interval	12.18	8.38	0	42	92.65
Proposal inside rational interval	8.59	6.65	0	27	95.59
Proposal in wrong direction	4.57	4.26	0	20	94.12
Proposal in correct direction	6.04	4.75	0	21	88.24
Stressed about time	0.35	0.79	0	4	22.06

Notes. This table considers the types of communications (comms) between teams. The data have been pooled across the 10 tasks. We report the average, standard deviation, and minimum and maximum number of types of statements sent from one teammate to another. The rational interval considered is the rational belief +/- 5%. The last column on the right shows the percentage of teams that made at least one comm in this category over the 10 tasks.

Figure 6 shows a histogram of the distribution of all pooled communications over the experiment for all teams. The highest percentage of comms occurred at the beginning of the tasks. “-1” in the bar chart refers to the 1-minute introduction period that we provided before the 10 estimation tasks where teams could introduce themselves and familiarize using the freeform instant message chat. They could not exchange messages during the Part 1 individual proxy of naïveté.

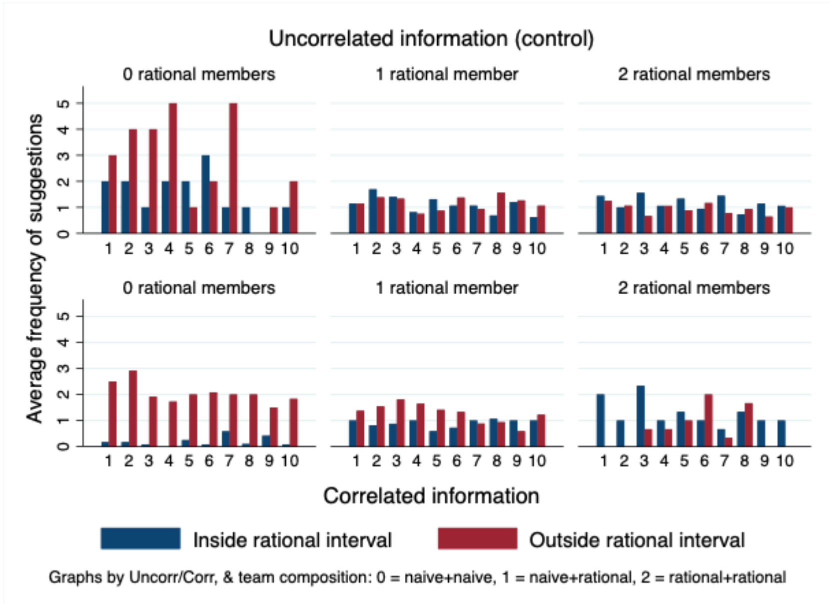


Notes. The histogram shows the distribution of the full sample of chat data chronologically over the 10 tasks (Part 2). Chat only began after Part 1 was completed individually. Part 2 was preceded by the 1-minute introduction period (labelled -1). Teams faced the 10 tasks in randomised order, so not all teams completed the same estimation task in Task 1, for example.

FIGURE 6

The distribution of communications sent over the experiment

Figure 7 shows the distribution of proposals made to teammates that were inside or outside a rational interval of +/- 5% from the rational belief. The top three graphs show how teams with different compositions approach *Uncorrelated*, while the bottom three graphs show how teams approach *Correlated*. Teams with two naïve members make more than twice as many proposals outside the rational interval as inside the rational interval when faced with *Correlated* information. Interestingly, teams with two naïve members also make many more proposals outside the rational interval, even without the problem of correlated information. This suggests that collective statistical illiteracy may drive some of the results. Teams with at least one rational member do not seem to systematically misunderstand the experimental instructions for *Uncorrelated* in the same manner (see top row). In *Correlated* (bottom row), as the number of rational team members increases, so the number of proposals inside the rational interval increases, and proposals outside the rational interval tend to decrease.

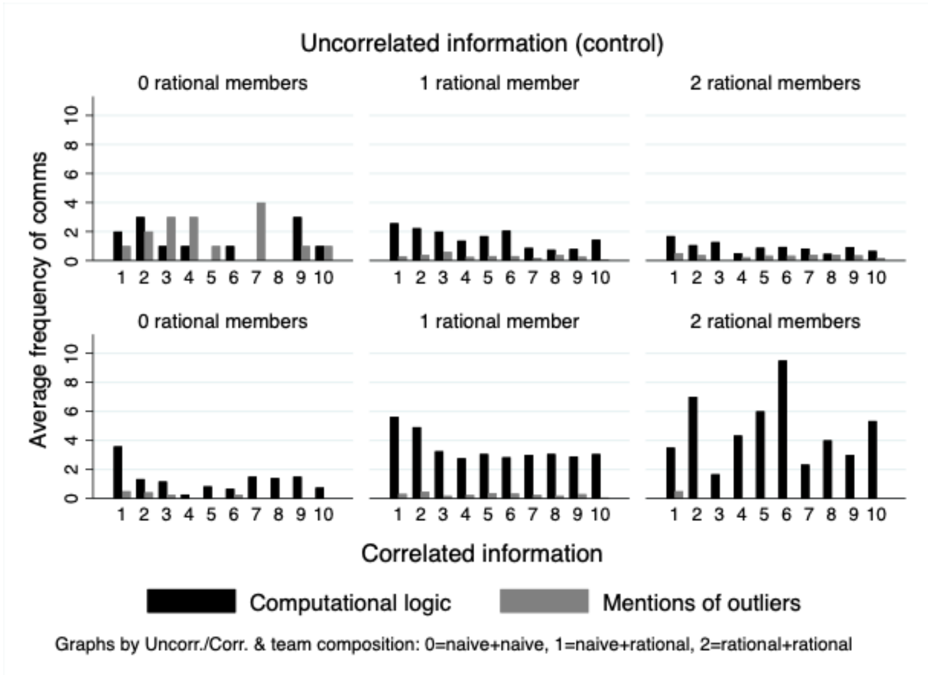


Note. The bar chart shows the frequency of comms characterised by proposals outside and inside an interval around the rational belief, mapped chronologically over the 10 tasks.

FIGURE 7

Proposals inside/outside a rational interval of +/-5% by Correlated/Uncorrelated and team composition

Figure 8 shows bar graphs of two categories: comms that were coded as computational logic or comms that mentioned an outlier. In *Correlated* (bottom row), teams with at least one rational member send more comms of computational logic than teams with only naïve members and are compared to similar teams in *Uncorrelated* (top row). We further explore mentions of outliers in *Uncorrelated* to see which teams systematically misunderstood the data generating process. We find that teams with two naïve members in the *Uncorrelated* condition make the most of such comms. This supports the hypothesis of collective statistical illiteracy driving some of the results rather than correlation neglect.



Notes. The bar chart shows the frequency of communications characterised by computational logic and mention of an outlier chronologically over the 10 tasks, where 1 refers to the first task, and 10 refers to the last task completed.

FIGURE 8

Distribution of communications: computational logic and questions to teammate

Table 7 shows OLS models of the influence of different kinds of comms on team performance, as measured by the dependent variable naïveté χ. Model 1 considers each of the 10 tasks separately. For example, in Column (1), we consider only the naïveté χ of the belief formed in Task 1 and the comms from Task 1. Similarly, Column (5) considers only the naïveté χ of Task 5 and the comms during Task 5. In Table 7 Panel A, the comms explanatory variables are binary, *i.e.* =1 if a type of comm was coded to be present, =0 otherwise. In Panel A, the correlation neglect coefficient is not significant in any of the 10 tasks. Having at least one rational team member is associated with forming significantly more rational beliefs in 4/10 tasks (rational benchmark is χ=0). Of interest is the binary explanatory variables for the comms. A subject signalling lack of confidence in the math was significantly associated with the team forming a more rational belief in 2/10 tasks. Agreement comms were significantly associated with more rational beliefs in 2/10 tasks. However, disagreement comms significantly predicted greater rationality of team belief in 4/10 tasks. Mention of an outlier was significantly associated with greater naïveté in 1/10 tasks. Proposals inside the rational interval were significantly associated with teams forming more rational beliefs in 4/10 tasks.

TABLE 7

Panel A. Model 1: Task-by-task influence of team comms on naïveté of team belief

Binary dummy variables for types of comms

	Dependent variable: Naïveté χ per task									
Task order	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0 if Uncorrelated, 1 if Correlated	0.062	-0.067	0.12	0.015	0.023	0.0085	-0.21	-0.16	0.38	-0.20
	(0.10)	(0.15)	(0.18)	(0.27)	(0.11)	(0.13)	(0.22)	(0.12)	(0.26)	(0.13)
At least 1 rational team member	-0.41**	0.29	-0.14	-0.19	-0.39*	-0.36	-0.74*	-0.33	0.060	-0.37*
	(0.15)	(0.28)	(0.26)	(0.35)	(0.21)	(0.24)	(0.39)	(0.25)	(0.36)	(0.22)
Signal bad at math (not confident)	-0.25*	-0.0100	-0.20	-0.56	0.20	-0.32	-0.13	-0.31	-0.12	-0.63***
	(0.13)	(0.36)	(0.28)	(0.42)	(0.16)	(0.21)	(0.41)	(0.25)	(0.26)	(0.23)
Asked question of partner	-0.0052	-0.24	-0.087	0.23	-0.053	0.18	-0.23	-0.23	0.33	-0.18
	(0.17)	(0.23)	(0.15)	(0.28)	(0.13)	(0.17)	(0.21)	(0.18)	(0.21)	(0.13)
Agreement	0	0.056	-0.23	-0.21	-0.16*	0.10	0.22	0.13	-0.96*	-0.049
	(.)	(0.30)	(0.20)	(0.24)	(0.09)	(0.19)	(0.33)	(0.37)	(0.51)	(0.10)
Disagreement	0.13	0	-0.43*	-0.077	-0.67	-0.87***	0	-0.72**	0.73**	0.24
	(0.20)	(.)	(0.23)	(0.56)	(0.57)	(0.27)	(.)	(0.28)	(0.34)	(0.27)
Mentioned an outlier	-0.23**	0.24	-0.010	-0.081	-0.078	0.36	0.25	0.032	0.55	0.046
	(0.11)	(0.21)	(0.26)	(0.36)	(0.12)	(0.24)	(0.36)	(0.26)	(0.37)	(0.23)
Irritated/bored with experiment	0.011	-0.20	0.64*	0.18	-0.27	0.31	-0.27	0.077	-0.82**	-0.015
	(0.17)	(0.36)	(0.36)	(0.52)	(0.21)	(0.20)	(0.35)	(0.16)	(0.31)	(0.14)
Computational logic	0.075	-0.028	0.076	0.051	-0.21*	-0.17	0.074	0.085	-0.28	0.13
	(0.10)	(0.19)	(0.17)	(0.20)	(0.12)	(0.15)	(0.23)	(0.16)	(0.26)	(0.13)
Suggestion outside rational interval	-0.22**	-0.17	0.24*	0.54*	-0.16	0.20	0.44	0.092	0.56	0.0098
	(0.10)	(0.25)	(0.12)	(0.30)	(0.14)	(0.19)	(0.31)	(0.21)	(0.41)	(0.21)
Suggestion inside rational interval	-0.60***	-0.82***	-0.058	-0.0063	-0.15	-0.16	-0.058	-0.75***	-0.59	-0.50***
	(0.16)	(0.25)	(0.24)	(0.38)	(0.15)	(0.21)	(0.28)	(0.26)	(0.46)	(0.17)
Suggestion right direction	0.083	-0.10	-0.23	-0.14	0.028	0.059	-0.38	0.044	-0.52*	0.19
	(0.10)	(0.12)	(0.19)	(0.28)	(0.13)	(0.13)	(0.24)	(0.13)	(0.30)	(0.22)
Suggestion wrong direction	0.34***	0.24	0.12	-0.20	0.026	-0.18	0.21	0.28	0.10	-0.23
	(0.10)	(0.20)	(0.23)	(0.35)	(0.14)	(0.18)	(0.26)	(0.26)	(0.28)	(0.24)
Stressed about time	-0.026	0.15	-0.059	-0.28	0.27	-0.045	0	0.091	-0.23	-0.77***
	(0.17)	(0.21)	(0.31)	(0.80)	(0.58)	(0.17)	(.)	(0.69)	(0.46)	(0.25)
Controls for session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	61	55	65	65	67	67	67	61	61	68
R-sq	0.70	0.45	0.27	0.24	0.40	0.48	0.36	0.58	0.54	0.47

Notes. Each of the comms explanatory variables is binary = 1 if the category was coded in the text, 0 otherwise. OLS estimates, robust standard errors (clustered at team level) in parentheses. The table analyses the sample of teams only. In particular, the influence of different types of communications on naïveté as implied in each of the 10 beliefs. All regressions exclude extreme outliers with the absolute value of $\chi > 3$. Significance is starred. * $p < 0.10$, **$p < 0.05$, ***$p < 0.01$

In Table 7 Panel B, the comms explanatory variables are continuous and count the number of types of comms made in a particular task. For example, in Task 2 (column 2), asking one or more questions of a partner is associated with a reduction in naïveté of 0.11 units of χ. Having at least one rational member in the team is associated with significantly more rational beliefs in 6/10 tasks.

Making more suggestions inside the rational interval led to significantly more rational beliefs being formed in 5/10 tasks.

TABLE 7

Panel B. Model 1: Task-by-task influence of team comms on naïveté of team belief

Continuous variables for types of comms

	Dependent variable: Naïveté χ per task									
Task order	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0 if Uncorrelated, 1 if Correlated	0.055	-0.078	0.20	0.012	0.026	-0.11	-0.23	-0.17	0.27	-0.10
	(0.14)	(0.19)	(0.17)	(0.26)	(0.13)	(0.16)	(0.22)	(0.16)	(0.33)	(0.16)
At least 1 rational team member	-0.64***	0.11	-0.13	-0.37	-0.39*	-0.59**	-0.77**	-0.42*	-0.21	-0.44*
	(0.17)	(0.36)	(0.25)	(0.35)	(0.22)	(0.26)	(0.35)	(0.22)	(0.36)	(0.23)
Signal bad at math (not confident)	-0.14	-0.042	-0.20	-0.46	0.13	-0.047	0.13	-0.28	-0.45	-0.55***
	(0.10)	(0.37)	(0.26)	(0.29)	(0.08)	(0.14)	(1.00)	(0.32)	(0.35)	(0.20)
Asked question of partner	-0.021	-0.11**	0.068	0.32*	-0.093	0.045	-0.16	-0.041	0.15	0.036
	(0.04)	(0.05)	(0.08)	(0.17)	(0.07)	(0.08)	(0.11)	(0.12)	(0.14)	(0.07)
Agreement	-0.0023	0.21*	-0.065	-0.046	-0.086	-0.050	-0.13	0.021	0.084	0.0089
	(0.05)	(0.12)	(0.10)	(0.13)	(0.07)	(0.09)	(0.15)	(0.16)	(0.12)	(0.07)
Disagreement	0.089	0	-0.30	0.17	-0.39	-0.67***	-0.21	-0.17	0.0061	0.50*
	(0.19)	(.)	(0.32)	(0.54)	(0.41)	(0.22)	(0.40)	(0.49)	(0.37)	(0.27)
Mentioned an outlier	-0.10	-0.014	-0.10	0.051	-0.078	0.22	0.0038	0.20	0.73**	0.16
	(0.07)	(0.12)	(0.08)	(0.19)	(0.12)	(0.14)	(0.22)	(0.22)	(0.36)	(0.19)
Irritated/bored with experiment	-0.0083	-0.35	0.20**	0.19	-0.24	0.092	-0.47	-0.024	-0.23	-0.026
	(0.08)	(0.39)	(0.10)	(0.28)	(0.18)	(0.25)	(0.34)	(0.14)	(0.35)	(0.14)
Computational logic	0.018	-0.019	-0.0020	0.035	-0.028	-0.0088	0.038	0.010	-0.087	0.042
	(0.02)	(0.04)	(0.03)	(0.05)	(0.02)	(0.03)	(0.07)	(0.06)	(0.06)	(0.03)
Suggestion outside rational interval	-0.10	0.13	-0.082	-0.21	-0.036	-0.13	0.011	0.085	-0.061	0.025
	(0.07)	(0.14)	(0.10)	(0.23)	(0.11)	(0.12)	(0.18)	(0.16)	(0.24)	(0.11)
Suggestion inside rational interval	-0.23***	-0.098	-0.16	-0.34*	-0.0022	-0.20*	-0.045	-0.30**	-0.51**	-0.14
	(0.07)	(0.11)	(0.14)	(0.18)	(0.10)	(0.11)	(0.15)	(0.14)	(0.24)	(0.11)
Suggestion right direction	0.084	-0.14	-0.14	0.093	0.080	0.083	-0.13	0.088	-0.0032	0.0052
	(0.07)	(0.09)	(0.16)	(0.27)	(0.12)	(0.11)	(0.15)	(0.16)	(0.27)	(0.15)
Suggestion wrong direction	0.23***	0.075	0.17	-0.042	0.046	0.14	0.25	-0.057	0.37	-0.10
	(0.08)	(0.14)	(0.19)	(0.17)	(0.11)	(0.15)	(0.16)	(0.19)	(0.25)	(0.13)
Stressed about time	0.070	0.13	0.28	-0.52	0.75	0.021	0	-0.058	-1.85***	-0.57**
	(0.20)	(0.41)	(0.37)	(0.94)	(0.73)	(0.30)	(.)	(0.71)	(0.69)	(0.22)
Controls for session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	61	55	65	66	67	67	67	61	61	68
R-sq	0.66	0.46	0.32	0.27	0.37	0.44	0.33	0.50	0.44	0.35

Notes. Each of the comms explanatory variables is continuous, taking into account the number of times that a particular category (e.g. signalled bad at math) was made by a teammate in that task. OLS estimates, robust standard errors (clustered at team level) in parentheses. The table analyses the sample of teams only. In particular, the influence of different types of communications on naïveté as implied in each of the 10 beliefs. All regressions exclude extreme outliers with the absolute value of $\chi > 3$. Significance is starred. * $p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 8 shows the results of Model 2, an alternative approach to analysing the influence of types of comms. For Model 2, we generate 10 dependent variables for naïveté_chi (e.g. naïveté_chi_5 for Task 5 etc.), that take on the value of naïveté_chi in each task, and missing otherwise (e.g. naïveté_chi_5 took on a missing value for Tasks 1-4 and Tasks 6-10). In Model 2, OLS regressions model the influence of cumulative comms in all tasks up to that point on naïveté. For example, the model of Task 5 considered the influence of comms in Tasks 1-5 on naïveté_chi_5. An increase in cumulative disagreement comms leads to a reduction in naïveté of 0.3 units of χ in Task 6 and Task 8. Both cumulative disagreement and having at least one rational member on the team are associated with forming more rational beliefs. Types of communications such as disagreement provide insight into the advantage of team decision-making given the random assignment to teams and heterogeneous distribution of rational and naïve individuals.

TABLE 8

Model 2: Influence of accumulated team comms on naïveté of team belief

Continuous cumulative variables for types of comms

	Dependent variable: Naïveté χ per task									
Task order	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0 if Uncorrelated, 1 if Correlated	0.055	0.084	0.22	-0.077	-0.13	0.087	-0.42**	-0.31*	0.38	-0.20
	(0.14)	(0.19)	(0.19)	(0.24)	(0.12)	(0.18)	(0.21)	(0.18)	(0.37)	(0.21)
At least 1 rational team member	-0.64***	0.15	-0.032	-0.57**	-0.63**	-0.33	-1.04**	-0.28	0.16	-0.55*
	(0.17)	(0.31)	(0.32)	(0.28)	(0.25)	(0.28)	(0.41)	(0.27)	(0.44)	(0.32)
C. Signal bad at math (not confident)	-0.14	0.062	-0.090	-0.22**	-0.085	-0.078	-0.14	-0.12*	0.014	-0.065
	(0.10)	(0.11)	(0.13)	(0.09)	(0.07)	(0.07)	(0.09)	(0.07)	(0.11)	(0.05)
C. Asked question of partner	-0.021	-0.069	-0.039	0.0037	-0.026	0.026	-0.022	-0.029	0.011	0.0078
	(0.04)	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)
C. Agreement	-0.0023	0.053	0.026	0.0038	0.0066	-0.022	0.042	0.051	0.023	0.0049
	(0.05)	(0.07)	(0.03)	(0.05)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
C. Disagreement	0.089	0.10	0.40	-0.099	-0.14	-0.32***	0.23	0.047	-0.30*	-0.095
	(0.19)	(0.36)	(0.33)	(0.28)	(0.09)	(0.09)	(0.15)	(0.16)	(0.16)	(0.10)
C. Mentioned an outlier	-0.10	0.11	-0.054	0.048	-0.046	0.082**	-0.030	0.051	0.13**	0.022
	(0.07)	(0.08)	(0.05)	(0.05)	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)	(0.03)
C. Irritated/bored with experiment	-0.0083	-0.17	0.21**	-0.024	-0.067*	0.032	0.21*	0.10*	-0.063	-0.045
	(0.08)	(0.14)	(0.10)	(0.10)	(0.04)	(0.10)	(0.11)	(0.05)	(0.09)	(0.04)
C. Computational logic	0.018	-0.016	-0.018	0.020	0.0033	-0.0053	-0.012	-0.0082	-0.0084	0.0037
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
C. Suggestion outside rational interval	-0.10	0.010	0.039	-0.13**	0.036	0.024	-0.067	0.055	0.17***	0.020
	(0.07)	(0.08)	(0.05)	(0.06)	(0.04)	(0.05)	(0.05)	(0.03)	(0.06)	(0.05)
C. Suggestion inside rational interval	-0.23***	-0.10	-0.039	-0.17***	0.041*	-0.0035	-0.061	-0.0053	0.092	-0.0015
	(0.07)	(0.07)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)	(0.03)	(0.07)	(0.04)
C. Suggestion right direction	0.084	-0.043	-0.070	0.061	-0.035	-0.029	0.024	-0.059*	-0.10	-0.036
	(0.07)	(0.08)	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.07)	(0.04)
C. Suggestion wrong direction	0.23***	0.13*	0.097**	0.14**	-0.036	-0.029	0.071	-0.00017	-0.21***	-0.029
	(0.08)	(0.07)	(0.04)	(0.06)	(0.03)	(0.05)	(0.05)	(0.04)	(0.06)	(0.04)
C. Stressed about time	0.070	0.25	-0.024	-0.27	0.028	0.0023	0.41*	-0.014	0.080	-0.024
	(0.20)	(0.23)	(0.16)	(0.19)	(0.13)	(0.18)	(0.21)	(0.11)	(0.20)	(0.14)
Controls for session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	61	55	65	66	67	67	67	61	61	68
R-sq	0.66	0.41	0.39	0.39	0.40	0.42	0.38	0.51	0.43	0.31

Notes. OLS estimates, robust standard errors (clustered at subject level) in parentheses. The table analyses the sample of teams only. In particular, the influence of all accumulated communications in previous tasks on the dependent variable: naïveté of a team belief. For example, the model of Task 5 naïveté considers the influence of comms accumulated in Task 1-5. All regressions exclude extreme outliers with the absolute value of $\chi > 3$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4. CONCLUSION

Previous research considers how individuals respond to information structures that generate correlated rather than independent signals (Enke and Zimmermann, 2019). Our contribution shows that team decision-making ameliorates this bias, generating more rational beliefs in a setting where individuals tend to exhibit substantial correlation neglect. We established that (1) there are significant differences between *Correlated* and *Uncorrelated* for individuals and teams, respectively; (2) on average, team beliefs converged more often towards the rational belief than individual beliefs; (3) pairs with at least one rational member formed significantly more rational beliefs than teams with two naïve members; and (4) communication plays a role, evident in differences in longer response times when there is a rational team member present. These findings prompt identifying in the chat content that factors teams' performance. In particular, what kind of communication influences the naïveté of beliefs formed by teams? Teams tend to chat the most at the beginning of the ten tasks with the majority of communications characterised as computational logic. Teams with at least one rational member exchange significantly more communication coded as computational logic, which explains their longer response times compared to teams with two naïve members. An insight into the decision-making process is that naïve team members are more likely to signal a lack of confidence in their ability to solve the correlation problem. They are also more likely to misunderstand the transparent data generating process, illustrated by mentions of outliers in the control treatment *Uncorrelated.* Since having at least one rational member is sufficient to attenuate the bias, this shows that the naïve team member tends to defer to her more rational partner. We find evidence that team decision-making can be a practicable policy to reduce the cognitive bias of correlation neglect in belief formation.

Acknowledgements. We would like to thank Vyoma Muralidhara for oTree programming and Michael Kuelper for coding the chat data. Thanks to Michael Kuelper, Celina Kullmann and Alina Sowa for assisting in running the experiment. Financial support from the Max Planck Institute for Research on Collective Goods and the University of Cologne is gratefully acknowledged.

Data. Data are available upon request from the authors.

5. APPENDICES

A. Order of Belief Formation Tasks in Main Treatments

This is kept the same as EZ19 (Online Appendix) for the purposes of replication. All belief elicitation treatments are implemented in three different randomised orders of rounds. Ordered by true state, they are as follows:

1. 10000, 88, 46422, 4698, 250, 23112, 1000, 10, 7338, 732
2. 732, 23112, 88, 1000, 250, 4698, 10, 7338, 10000, 46422
3. 250, 7338, 10000, 10, 4698, 88, 46422, 732, 1000, 23112

B. Instructions for TEAM Correlated Treatment

Instructions (to be handed out to participants in hard copy)

You will now take part in an economic experiment. You will receive an amount of 5 euros for your participation in the experiment. This will be paid to you at the end of the experiment. You can earn additional money, which is also paid out at the end of the experiment. How much you earn depends on your decisions. In this experiment we will talk about points. The points you collect during the experiment are exchanged for euros and paid out at the end of the experiment. Talking to other participants is not allowed during the experiment. If you have any questions, you can raise your arm outside the booth and the experimenter will attempt to answer your questions.

Your task:
In this experiment you have to solve eleven guessing tasks. In **Part 1** you have to solve one **guessing task alone** and in **part 2** you have to solve ten guessing tasks **together with a randomly selected partner in the team**. These tasks require you to estimate how many objects are in a given imaginary container. These objects can be, for example, peas in a vase or stones in a jar. All estimation tasks are different and completely independent of each other. Your earnings depend on how accurately you estimate, that is, how close your estimate is to the actual number of objects in the container. At the end of the experiment, the guess from part 1 and a randomly chosen guess from part 2 will be payable and you will be paid based on the accuracy of your guesses in the tasks. This is explained in more detail in the following pages.

Your decision:
See the final page for a screenshot of the decision screen. You can enter a number by typing the number in the blue box. **A decision is not valid until you have confirmed the estimate entered.**

Procedure for the individual estimation tasks:

You are given a total of 7 minutes for each of the 11 estimation tasks. After the individual estimation, the start of the team part will be announced. In the team section, after 5 minutes you will receive a message on the screen that you only have 2 minutes to make your decision. For both team members, this means that a decision must be confirmed (click confirmation button) after a maximum of 7 minutes in order to have a valid estimate. However, you can also move on to the next estimation task at any time beforehand, provided that you have submitted a valid estimation.

Team decisions in part 2:
For these estimation tasks, you are randomly paired with another participant (always the same person). You and your team partner can communicate with each other (anonymous team chat). **Both team members see the same screen and numbers (see last page) but have to enter estimates individually**. You can enter a number by typing the number in the blue field (your guess: the jar contained the following number of items) and clicking the "OK" button. You can change the number by entering a new number and clicking OK. At the end of each estimation task, your team must have made a decision (= both team partners have entered the same estimation individually). To finally confirm a number as a team decision, one of you must click the "**Confirm Estimation Task**" button. **A number cannot be confirmed until both team members have entered the same number with "OK".** If your team's entered estimates do not match, you will receive information on the screen.

Team chat:
In each task, you and your partner can talk via **chat, which is only visible to your team**. The chat is displayed in the lower half of the screen for all tasks. You send a message by typing the text in the blue line and then confirming with the "Enter" key. The message is then immediately visible to you and your partner in the chat history. **The chat history is emptied after each guess task.**
No threats or insults may be uttered in the chat and no information may be disclosed that allows conclusions to be drawn about your identity (e.g. name or seat number).

Your earnings:
In addition to your show up fee payment, you will be paid based on the accuracy of your estimates. In this experiment we will talk about points. 100 points correspond to 2 euros in part 1 (individual estimation) and 10 euros in part 2 (team estimations). The points you collect during the experiment are exchanged for euros and are also paid out at the end of the experiment. You get more money the closer your estimate is to the actual number of items in the bin. Your individual estimate and a randomly selected estimation task (from a total of 10 estimates) from part 2 (team estimates) are relevant for the payment. This means each team estimate is potentially relevant to your payment, so you should think carefully about each task. You can earn a maximum of 100 points per task. In the **team part, the 100 points are to be understood per person, which means that the payout for each team member is calculated according to the formula below**. You get these 100 points if you guess the number of objects correctly. The further your estimate is from the actual number, the less you will earn. This is determined using the following formula:

Payment = 100-0.16×(percentage deviation of appraised value from real value)^2

This means that the percentage deviation (in percent) between your estimate and the true value is squared and multiplied by 0.16. This number is then subtracted from the maximum earning of 100 points. Although this formula looks complicated, the underlying principle is very simple: the smaller the difference between your estimate and the true value, the larger your profit. However, your earnings can never be less than zero, i.e., you cannot lose money.

Consider the following example: Suppose said container contains 650 objects. You will then get the following number of points depending on your guess:

Estimate 0 → deviation 100% → earns 0 points
Estimate 520 → deviation 20% → earns 36 points
Estimate 585 → deviation 10% → earns 84 points
Estimate 650 → deviation 0% → earns 100 points
Estimate 715 → deviation 10% → earns 84 points
Estimate 780 → deviation 20% → earns 36 points
Estimate 1300 → deviation 100% → earns 0 points

You see that your earnings depend only on the percentage deviation. For example, it doesn't matter if you overestimate or underestimate the true contents of the container by 10 items.

Information regarding the estimation tasks: You will not see any of the estimation tasks, i.e. You won't see any of these imaginary containers. Instead, you receive various computer-generated information about the correct estimation result for each estimation task. For each task, you see this information and then enter your own estimate. The information you will receive is explained in detail below.
On computers, we simulate devices that solve exactly the same guessing tasks as you. There are two different types of devices. On the one hand there are devices that themselves give an estimate of the contents of the containers (these devices are called measuring devices and are identified by letters). On the other hand, there are devices that observe the estimates of the measuring devices and calculate their own estimate based on these reports (these devices are called communication devices and are identified by numbers). For each estimation task, 10,000 computers (measuring devices) made their own estimates; the estimates of these devices are completely independent of each other. The measuring devices all have the same quality, i.e., they are equally good at estimating. Note that these devices are good at solving these guessing tasks. So if these devices determined many estimates, then the average of those estimates would be correct.

So while almost every single guess contains an error, the average of those 10,000 guesses is correct (or very accurate). In addition, many estimates are relatively close to the true value. Below is an illustrative example in the summary.

Summary:

(1) The measuring devices make mistakes, but the estimate is much more likely to be close to the true value than to be far from it.

(2) The average of these 10,000 estimates is correct. This means that the devices are correct (or very accurate) on average. Below we present an example.

Suppose the container contains 650 objects. The following figure illustrates how often the 10,000 measuring devices give which estimate. You can see that the frequency (frequency) increases as you approach the centre of the bell curve (the true value is 650). Therefore, it is much more likely that an estimate approaches the true value rather than deviating greatly from it. Also, you can deduce that the average of all estimates is 650, which is correct.

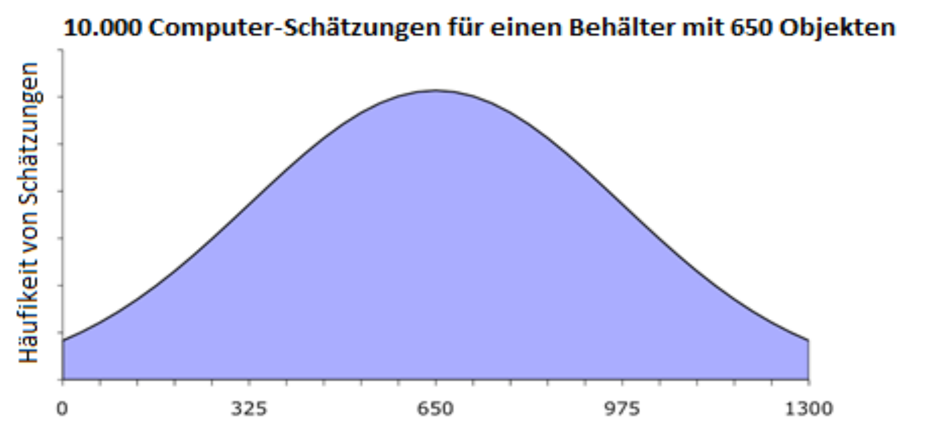


Figure 1: Sample computer estimates for a bin containing 650 items

For each estimation task, a computer randomly selects the estimates from exactly four measuring devices (A, B, C, D) from these 10,000 estimates. These four estimates are completely independent of each other. In addition to the measuring devices, there are three other communication devices (1, 2, 3). These communication devices do not determine their own estimation. Rather, they always observe two measuring devices and calculate an estimate from these estimates. More specifically, they do this by averaging these two estimates. You will get the following information as described in Figure 2.

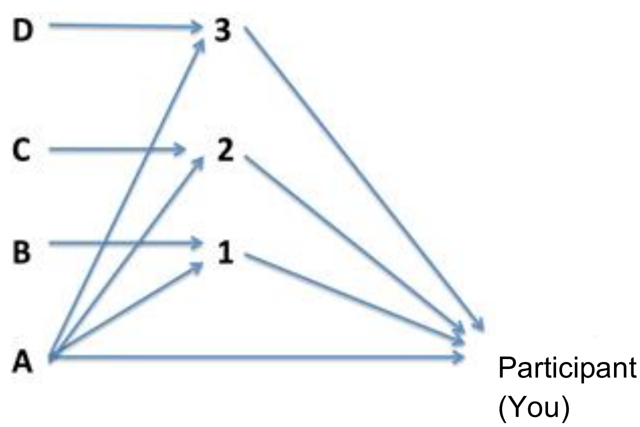


Figure 2: The four measuring devices determine their own estimates. The three communication devices receive the estimate of measuring device A and another measuring device as described by the arrows. They calculate their own estimate by averaging the two estimates. You can see the estimate of measuring device A as well as the three communication devices.

This means that you get the following information: You get the estimate of measuring device A as well as the estimates of all communication devices. As shown in Figure 2, each communication device sees the estimates of measuring device A and another measuring device. As you can see, communicator 1 receives the estimates from measuring devices A and B. Communicator 2 sees the estimates from measuring devices A and C. Communicator 3 sees the estimates from measuring devices A and D. As discussed earlier, the communicators average of these two estimates and report that average as their estimate.

An example:
You get the estimate of measuring device A, as well as the estimates of communicators 1, 2, and 3. The following simple example illustrates this. Above, we showed you the 10,000 estimates for the 650-object

container. Now imagine that from these estimates, a computer randomly selects the following estimates from the estimators:

Measuring device A: 810
Measuring device B: 126
Measuring device C: 1078
Measuring device D: 937

As described above, the communication devices take the average of two of these estimates. You then get the following estimates:

Communication device 1: 468
Communication device 2: 944
Communication device 3: 873.5

So for this guessing task, you would see the following information on your computer screen (see Figure 3):

Figure 3: Example screenshot

IMPORTANT!
Confirm chat entries with the **Enter key on the keyboard.**

Both team members see the same numbers.

Each round, they enter estimates individually by clicking the "OK" button. A new number can be entered by clicking "OK" again.

If you and your team partner agree, confirm the number with the "confirmation button". Confirmation is only possible if the individual estimates are identical.

Please read these instructions carefully again. The experimenter will shortly read a summary of the instructions. You will then answer a series of control questions on the computer to ensure that you have understood the instructions.

C. Instructions for TEAM Uncorrelated

Notes: The instructions for TEAM Uncorrelated proceed exactly as for TEAM Correlated (see above Appendix B) with the exception of the following section below:

For each estimation task, a computer randomly selects the estimates from exactly four measuring devices (A, B, C, D) from these 10,000 estimates. These four estimates are completely independent of each other. In addition to the measuring devices, there are three other communication devices (1, 2, 3). These communication devices do not determine their own estimation. Rather, they are always watching a measuring device and reporting that device's estimate. You will get the following information as described in Figure 2.

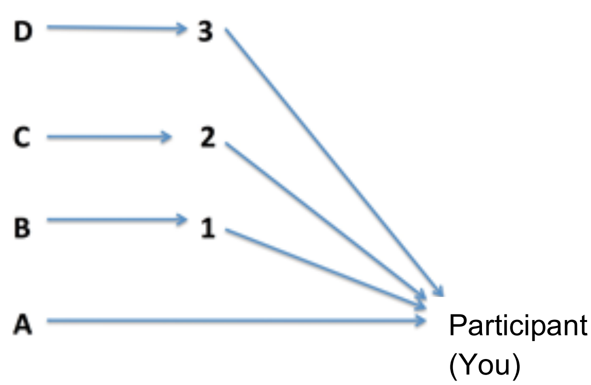


Figure 2: The four measuring devices determine their own estimates. The three communication devices receive a measuring devices estimate as described by the arrows. They calculate their own estimate by taking the measuring device's estimate. You can see the estimate of measuring device A as well as the three communication devices 1, 2, and 3.

This means that you get the following information: You get the estimate of measuring device A as well as the estimates of all communication devices. As shown in Figure 2, each communication device sees a measuring device's estimates. As you can see, communicator 1 receives measuring device B's estimate. Communicator 2 sees measuring device C's estimate. Communicator 3 sees measuring device D's estimate. As previously discussed, communicator reports the estimates of each measuring device.

An example:
You get the estimate of measuring device A, as well as the estimates of communicators 1, 2, and 3. The following simple example illustrates this. Above, we showed you the 10,000 estimates for the 650-object container. Now imagine that from these estimates, a computer randomly selects the following estimates from the estimators:

Measuring device A: 810
Measuring device B: 126
Measuring device C: 1078
Measuring device D: 937

As described above, each communication device reports an estimate of a meter. So they give the following estimates:

Communication device 1: 126
Communication device 2: 1078
Communication device 3: 937

So for this guessing task, you would see the following information on your computer screen (see Figure 3):

Figure 3: Example screenshot

D. *Summary of instructions for TEAM treatments only*

Summary (to be read out loud by an experimenter)

- In this experiment you will be presented with 11 different estimation tasks.
- Part 1 is an individual decision and Part 2 consists of 10 team decisions.
- Both team members see the same numbers in the team decisions.
- At the end of the experiment, the task from Part 1 and one of the 10 rounds from Part 2 will be randomly selected to determine your earnings. Your earnings per task depend on how precise your

answer was in the guessing task, that is, how close it was to the right value. Since all estimation tasks are potentially payout relevant, you should answer carefully in all rounds.

- You won't actually see the guessing tasks. Instead, you receive computer-generated information about the solution to the estimation tasks.
- The structure of the information is as follows: In each round, four measuring devices that try to solve your guessing task are randomly selected. These devices are always equally good at solving the task. Each device provides its own estimate.
- In addition, there are 3 communication devices that process the estimates they observe from the estimators and then return an estimate derived from those estimates.
- Please look again at Figure 2 in your instructions. There you can see which estimates each communication device is observing, how they are processed, and how they are incorporated into the estimates reported by communication devices.
- For each estimation task you get the following information: You see the estimation of the measuring device A, as well as the estimations of communication devices 1, 2 and 3.
- After observing this information, you have 7 minutes to think about your own estimate and enter it into your computer.

E. Instructions for INDIVIDUAL Correlated treatment

Instructions (to be handed out to participants)

You will now take part in an economic experiment. You will receive an amount of 5 euros for your participation in the experiment. This will be paid to you at the end of the experiment. You can earn additional money, which is also paid out at the end of the experiment. How much you earn depends on your decisions. In this experiment we will talk about points. The points you collect during the experiment are exchanged for euros and paid out at the end of the experiment. Talking to other participants is not allowed during the experiment. If you have any questions, you can raise your arm outside the booth and the experimenter will attempt to answer your questions.

Your task:
In this experiment you have to solve eleven guessing tasks. In **part 1** you have to solve one guessing task and then in **part 2** ten guessing tasks. These tasks require you to estimate how many objects are in a given imaginary container. These objects can be, for example, peas in a vase or stones in a jar. All estimation tasks are different and completely independent of each other. Your earnings depend on how accurately you estimate, e.g. how close your estimate is to the actual number of objects in the container. At the end of the experiment, the guess from part 1 and a randomly chosen guess from part 2 will be payable and you will be paid based on the accuracy of your guesses in the tasks. This is explained in more detail on the next page.

Your decision:
See the final page for a screenshot of the decision screen. You can enter a number by typing the number in the blue box and clicking the "Next" button. Only click on "Continue" if you really want to send the estimate, since no changes can be made afterwards. **A decision is only valid once you have confirmed the entered estimate with "Next".**

Procedure for the individual estimation tasks:
You are given a total of **7 minutes** for each estimation task, and after 5 minutes you will receive information about the remaining time in the upper right part of the screen. For you, this means that a decision must be confirmed after a maximum of 7 minutes in order to have a valid (payout-relevant) estimate. However, you can also move on to the next estimation task at any time beforehand, provided that you have submitted a valid estimation.

Your earnings:
In addition to your show up fee earnings, you will be paid based on the accuracy of your estimates. In this experiment we will talk about points. 100 points correspond to 2 euros in part 1 and 10 euros in part 2. The points that you collect in the course of the experiment are exchanged for euros and are also paid out at the end of the experiment. You get more money the closer your estimate is to the actual number of items in the bin. Your estimate from Part 1 and a randomly selected estimation task (from a total of 10 estimates) from Part 2 are relevant for the payout. This means each estimate is potentially relevant to your payment, so you should think carefully about each task. You can earn a maximum of 100 points with one task. You get these 100 points if you guess the number of objects correctly. The further your estimate is from the actual number, the less you will earn.
This is determined using the following formula (per task):

Payment = 100-0.16×(percentage deviation of estimated value from real value)^2

This means that the percentage deviation (in percent) between your estimate and the true value is squared and multiplied by 0.16. This number is then subtracted from the maximum earning of 100 points. Although this formula looks complicated, the underlying principle is very simple: the smaller the difference between your estimate and the true value, the larger your profit. However, your earnings can never be less than zero, i.e., You cannot lose money.

Consider the following example: Suppose said container contains 650 objects. You will then get the following number of points depending on your guess:

Estimate 0 → deviation 100% → earns 0 points
Estimate 520 → deviation 20% → earns 36 points
Estimate 585 → deviation 10% → earns 84 points
Estimate 650 → deviation 0% → earns 100 points
Estimate 715 → deviation 10% → earns 84 points
Estimate 780 → deviation 20% → earns 36 points
Estimate 1300 → deviation 100% → earns 0 points

You see that your earnings depend only on the percentage deviation. For example, it doesn't matter if you overestimate or underestimate the true contents of the container by 10 items.

Information regarding the estimation tasks: You will not see any of the estimation tasks, i.e. You won't see any of these imaginary containers. Instead, you receive various computer-generated information about the correct estimation result for each estimation task. For each task, you see this information and then enter your own estimate. The information you will receive is explained in detail below.
On computers, we simulate devices that solve exactly the same guessing tasks as you. There are two different types of devices. On the one hand there are devices that themselves give an estimate of the contents of the containers (these devices are called measuring devices and are identified by letters). On the other hand, there are devices that observe the estimates of the measuring devices and calculate their own estimate based on these reports (these devices are called communication devices and are identified by numbers). For each estimation task, 10,000 computers (measuring devices) made their own estimates; the estimates of

these devices are completely independent of each other. The measuring devices all have the same quality, i.e., they are equally good at estimating. Note that these measuring devices are good at solving these guessing tasks. So if these devices determined many estimates, then the average of those estimates would be correct.

So while almost every single guess contains an error, the average of those 10,000 guesses is correct (or very accurate). In addition, many estimates are relatively close to the true value. Below is an illustrative example in the summary.

Summary:

(1) The measuring devices make mistakes, but the estimate is much more likely to be close to the true value than to be far from it.
(2) The average of these 10,000 estimates is correct. This means that the devices are correct (or very accurate) on average. Below we present an example.

Suppose the container contains 650 objects. The following figure illustrates how often the 10,000 measuring devices give which estimate. You can see that the frequency (frequency) increases as you approach the centre of the bell curve (the true value is 650). Therefore, it is much more likely that an estimate approaches the true value rather than deviating greatly from it. Also, you can deduce that the average of all estimates is 650, which is correct.

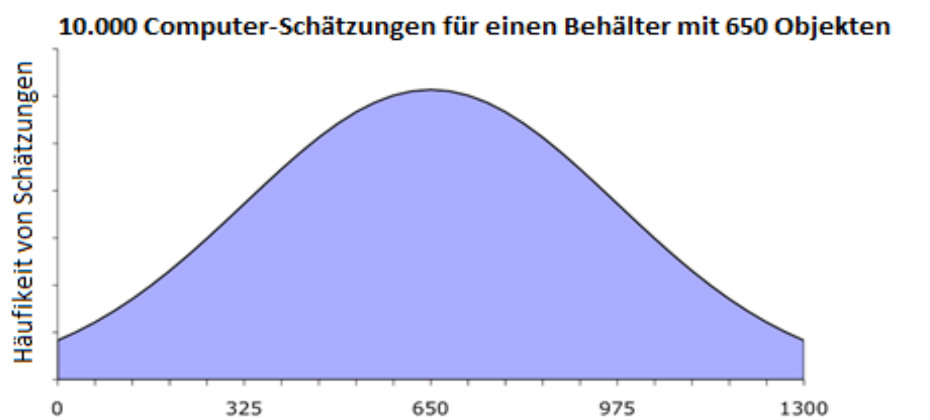


Figure 1: Sample computer estimates for a bin containing 650 items

For each estimation task, a computer randomly selects the estimates from exactly four measuring devices (A, B, C, D) from these 10,000 estimates. These four estimates are completely independent of each other. In addition to the measuring devices, there are three other communication devices (1, 2, 3). These communication devices do not determine their own estimation. Rather, they always observe two measuring devices and calculate an estimate from these estimates. More specifically, they do this by averaging these two estimates. You will get the following information as described in Figure 2.

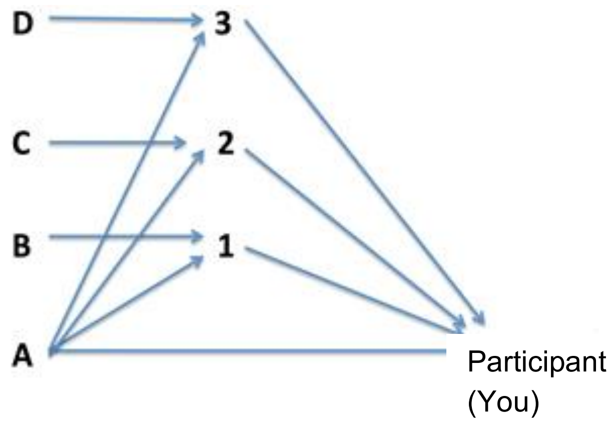


Figure 2: The four measuring devices determine their own estimates. The three communication devices receive the estimate of measuring device A and another measuring device as described by the arrows. They calculate their own estimate by averaging the two estimates. You can see the estimate of measuring device A as well as the three communication devices.

This means that you get the following information: You get the estimate of measuring device A as well as the estimates of all communication devices. As shown in Figure 2, each communication device sees the estimates of measuring device A and another measuring device. As you can see, communicator 1 receives the estimates from measuring devices A and B. Communicator 2 sees the estimates from measuring devices A and C. Communicator 3 sees the estimates from measuring devices A and D. As discussed earlier, the communicators average of these two estimates and report that average as their estimate.

An example:
You get the estimate of measuring device A, as well as the estimates of communicators 1, 2, and 3. The following simple example illustrates this. Above, we showed you the 10,000 estimates for the 650-object container. Now imagine that from these estimates, a computer randomly selects the following estimates from the estimators:

Measuring device A: 810
Measuring device B: 126
Measuring device C: 1078
Measuring device D: 937

As described above, the communication devices take the average of two of these estimates. You then get the following estimates:

Communication device 1: 468
Communication device 2: 944
Communication device 3: 873.5

So for this guessing task, you would see the following information on your computer screen (see Figure 3):

Figure 3: example screenshot

Please read these instructions carefully again. The experimenter will shortly read a summary of the instructions. You will then answer a series of control questions on the computer to ensure that you have understood the instructions.

F. Instructions for INDIVIDUAL Uncorrelated

Notes: The instructions for INDIVIDUAL Uncorrelated proceed exactly as for INDIVIDUAL Correlated (see above Appendix E) besides the following section below:

For each estimation task, a computer randomly selects the estimates from exactly four measuring devices (A, B, C, D) from these 10,000 estimates. These four estimates are completely independent of each other. In addition to the measuring devices, there are three other communication devices (1, 2, 3). These communication devices do not determine their own estimation. Rather, they are always watching a measuring device and reporting that device's estimate. You will get the following information as described in Figure 2.

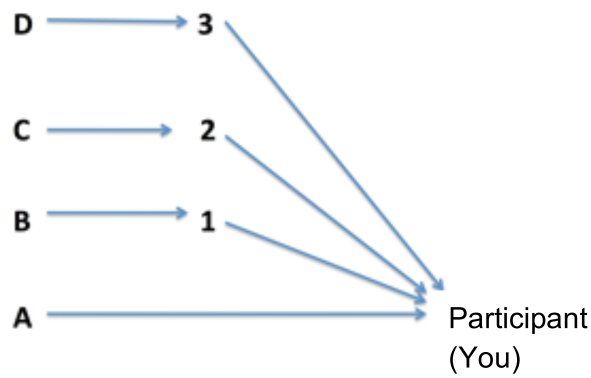


Figure 2: The four measuring devices determine their own estimates. The three communication devices receive a measuring devices estimate as described by the arrows. They calculate their own estimate by taking the measuring device's estimate. You can see the estimate of measuring device A as well as the three communication devices 1, 2, and 3.

This means that you get the following information: You get the estimate of measuring device A as well as the estimates of all communication devices. As shown in Figure 2, each communication device sees a measuring device's estimates. As you can see, communicator 1 receives measuring device B's estimate. Communicator 2 sees measuring device C's estimate. Communicator 3 sees measuring device D's estimate. As previously discussed, communicator reports the estimates of each measuring device.

An example:
You get the estimate of measuring device A, as well as the estimates of communicators 1, 2, and 3. The following simple example illustrates this. Above, we showed you the 10,000 estimates for the 650-object container. Now imagine that from these estimates, a computer randomly selects the following estimates from the estimators:

Measuring device A: 810
Measuring device B: 126
Measuring device C: 1078
Measuring device D: 937

As described above, each communication device reports an estimate of a meter. So they give the following estimates:

Communication device 1: 126
Communication device 2: 1078
Communication device 3: 937

So for this guessing task, you would see the following information on your computer screen (see Figure 3):

Figure 3: Example screenshot

G. Summary of instructions – INDIVIDUAL treatments only

Summary (to be read out loud by an experimenter)

- In this experiment you will be confronted with 11 different estimation tasks.
- At the end of the experiment, the task from Part 1 and one of the 10 rounds from Part 2 will be randomly selected to determine your win. Your earnings per task depend on how accurate your answer was in the guessing task, that is, how close it was to the right value. Since all estimation tasks are potentially payout relevant, you should answer carefully in all rounds.
- You won't actually see the guessing tasks. Instead, you receive computer-generated information about the solution to the estimation tasks.
- The structure of the information is as follows: In each round, four measuring devices that try to solve your guessing task are randomly selected. These devices are always equally good at solving the task. Each device provides its own estimate.
- In addition, there are three communication devices that process the estimates they observe from the estimators and then return an estimate derived from those estimates.
- Please look again at Figure 2 in your instructions. There you can see which estimates each communication device is observing, how they are processed, and how they are incorporated into the estimates reported by communication devices.
- For each estimation task you get the following information: You see the estimation of measuring device A as well as the estimations of communication devices 1, 2 and 3.
- After observing this information, you have 7 minutes to think about your own estimate and enter it into your computer.

6. REFERENCES

Adams, R., & Ferreira, D. (2010). Moderation in groups: Evidence from betting on ice break-ups in Alaska. *Review of Economic Studies*, *77*, 882–913. Retrieved from https://academic.oup.com/restud/article-abstract/77/3/882/1572031

Agranov, M., Caplin, A., & Tergiman, C. (2015). Naïve play and the process of choice in guessing games. *Journal of the Economic Science Association, 1*(2), 146–157. https://doi.org/10.1007/s40881-015-0003-5

Akerlof, G. A. and Shiller, R. J. (2009). *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*. Princeton, NJ: Princeton University Press.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association, 11*(5), 1079–1114. Retrieved from https://academic.oup.com/jeea/article-abstract/11/5/1079/2315903

Barahona, R., Cassella, S., & Jansen, K. A. (2021). Do Teams Alleviate or Exacerbate Cognitive Biases? Evidence from Extrapolation in Mutual Funds (February 10, 2021). *SSRN*

Barber, B. M., Heath, C., & Odean, T. (2003). Good Reasons Sell: Reason-Based Choice Among Group and Individual Investors in the Stock Market. *Management Science, 49*(12), 1636–1652. https://doi.org/10.1287/mnsc.49.12.1636.25109

Bénabou, R. (2013). Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies, 80*(2), 429–462. https://doi.org/10.1093/restud/rds030

Buffat, J., Praxmarer, M., & Sutter, M. (2020). The Intrinsic Value of Decision Rights: A Note on Team vs Individual Decision-Making. *MPI Collective Goods Discussion Paper, (2020/30).*

Charness, G., Oprea, R., & Yuksel, S. (2021). How do people choose between biased information sources? Evidence from a laboratory experiment. *Journal of the European Economic Association, 19*(3), 1656-1691.

Cooper, D. J., & Kagel, J. H. (2005). Are Two Heads Better Than One? Team versus Individual Play in Signaling Games. *American Economic Review, 95*(3), 477–509. https://doi.org/10.1257/0002828054201431

Davis, J. H. (1992). Some Compelling Intuitions about Group Consensus Decisions, Theoretical and Empirical Research, and Interpersonal Aggregation Phenomena: Selected Examples, 1950 –1990. *Organizational Behavior and Human Decision Processes, 52*(1), 3–38.

DellaVigna, S., & Pope, D. (2018). Predicting experimental results: who knows what?. *Journal of Political Economy, 126*(6), 2410-2456.

DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. *Science, 366*(6464), 428-429.

Dennis, A. R., & Valacich, J. S. (1994). Group, sub-group, and nominal group idea generation: New rules for a new media?. *Journal of Management, 20*(4), 723-736.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics, 135*(3), 1363-1398.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171-178.

Greiner, B. (2004). The Online Recruitment System ORSEE. University of Cologne (No. 10). *Working Paper.*

Houghton, S. M., Simon, M., Aquino, K., & Goldberg, C. B. (2000). No safety in numbers:

Persistence of biases and their effects on team risk perception and team decision making. *Group & Organization Management, 25*(4), 325-353.

Kocher, M. G., Praxmarer, M., & Sutter, M. (2020). Team Decision-Making. *Handbook of Labor, Human Resources and Population Economics,* 1-25.

Kocher, M., Strauß, S., & Sutter, M. (2006). Individual or team decision-making—causes and consequences of self-selection. *Games and Economic Behavior, 56*(2), 259-270.

Kocher, M. G., & Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal, 115*(500), 200-223.

Lehner, P., Seyed-Solorforough, M. M., O'Connor, M. F., Sak, S., & Mullin, T. (1997). Cognitive biases and time stress in team decision making. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 27*(5), 698-703.

Luhan, W. J., Kocher, M. G., & Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics, 12*(1), 26–41. https://doi.org/10.1007/s10683-007-9188-7

Maciejovsky, B., & Budescu, D. (2007). Collective induction without cooperation? Learning and knowledge transfer in cooperative groups and competitive auctions. *Journal of Personality and Social Psychology, 92*(5), 854. Retrieved from https://psycnet.apa.org/record/2007-06231-007

Mancuso, V. F., Finomore, V. S., Rahill, K. M., Blair, E. A., & Funke, G. J. (2014, September). Effects of cognitive biases on distributed team decision making. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 405-409). Sage CA: Los Angeles, CA: SAGE Publications.

Minas, R. K., Potter, R. F., Dennis, A. R., Bartelt, V., & Bae, S. (2014). Putting on the thinking cap: using NeuroIS to understand information processing biases in virtual teams. *Journal of Management Information Systems, 30*(4), 49-82.

Monteiro, S. (forthcoming). Visual representation training to debias selection neglect in individual belief formation. *Working paper.*

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. Experimental Economics, 1(1), 43-61.

Sutter, M. (2007). Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. Economics Letters, 97(2), 128–132. Retrieved from https://www.sciencedirect.com/science/article/pii/S0165176507000699

Chapter 2. Visual representation training to debias selection neglect in belief formation

Sofia Monteiro*

Max Planck Institute for Research on Collective Goods, Bonn
University of Cologne

Abstract. This paper tests whether instructing participants in a novel Bayesian reasoning exercise increases the proportion of Bayesian updaters in a setting of selection neglect; a series of estimation tasks in which a substantial proportion of people have been shown to neglect the problem of selected information. The exercise comprises a story of a dice game played between each participant and a fictitious person, and uses visual representation and experience sampling. In an online experiment, I replicate laboratory results demonstrating selection neglect in the median regressions. The results were not robust to Ordinary Least Squares (OLS) winsorised and OLS trimmed models. The treatment significantly attenuates the bias, but there is only weak evidence for this. This is the first test of visual representation training in an incentivised belief formation setting.

JEL classification: D03; D80; D84.
Keywords: Bounded rationality; Bayesian inference; representation of information; visualisation; incentivised beliefs

* Correspondence: sofia.g.monteiro@gmail.com

1. INTRODUCTION

Visual art uses both actual and implied volume.[21] "Positive space" refers to the space of the defined shape, while "negative space" describes the space around shapes and figures. Shape, volume, and space, whether actual or implied, are the foundation of the perception of reality. On social media platforms, such as Instagram, the stories we do not "see" can be just as influential on the picture we form about the world even though they fade into the background as implicit information. Information, whether as art or virtual "stories", has to compete for our limited attention. We commonly face a problem of selected information, and presentation can distort our perception of reality. A social media algorithm – aiming to optimise engagement with the platform – faces a budget constraint on what it can show us in the first few posts that we scroll through. Consequently, it prioritises the types of subjects with which we have previously engaged. To form a rational perception of reality, we have to draw inferences from the "absence" of information or the "negative space".

I use Enke's (2020) paper as a framework with which to further study the problem of selection neglect with an online sample and how one might go about debiasing beliefs with a practical treatment. In the laboratory, he observes that participants often form beliefs that reflect a complete neglect of the selected information problem. His paper provides a theoretical and empirical basis for the notion that some people use a "misspecified mental model" of the environment, namely, "what you see is all there is". A subject's mental model describes the way they set up, rather than solve, the problem of calculating posterior beliefs from the information that is revealed to them after initially having formed a prior belief about the state of the world.

The relevance of misspecified mental models is covered by an active theoretical literature in economics (Bohren & Hauser, 2017; Bordalo, Gennaioli, & Shleifer, 2017; Bushong & Gagnon-Bartsch, 2016; Enke & Zimmermann, 2019; Eyster & Rabin, 2010; Gabaix, 2014; Gagnon-Bartsch, Rabin, & Schwartzstein, 2018; Gennaioli & Shleifer, 2010; Heidhues, Kőszegi, & Strack, 2018; Jehiel, 2001; Schwartzstein, 2014; Spiegler, 2016). Empirical work, however, is scarce

[21] For an overview of elements of art and design see: lumenlearning.com. [Date Accessed: 30 September 2021] Located at: https://courses.lumenlearning.com/boundless-arthistory/chapter/visual-elements/

(Charness, Oprea, & Yuksel, 2021; Enke & Zimmermann, 2019). Charness et al. (2021) find convincing evidence of confirmation-seeking heuristics leading people to choose information sources biased towards their priors. It is suggested that participants follow this type of heuristic because of fundamental errors in reasoning about the relative value of the information from biased sources (Charness et al., 2021). Enke (2020) shows that mental models respond predictably to variations in the environment. At least part of the reason that subjects neglect selection in this setting is that they do not focus on the selection problem in the first place. They seem overly concerned with implementing a solution strategy, jumping to the calculations, and in doing so, miss that their assumptions rest on an erroneous mental model. The bias is quite resistant to various treatment variations designed to eliminate it (detailed below), including explicitly nudging people to pay attention to signals not shown to them on the decision screen (Enke, 2020).

Gigerenzer et al. (2007) advocate for statistical literacy as a necessary precondition for an educated citizenship in a technological democracy. Several papers suggest statistical inference can be taught to almost anyone (even to children) using frequency statements instead of single-event probabilities, absolute risks instead of relative risks, mortality rates instead of survival rates, and natural frequencies instead of conditional probabilities (Hoffrage & Gigerenzer, 1998; Hoffrage, Krauss, Martignon, & Gigerenzer, 2015; Kurzenhäuser & Hoffrage, 2002; Sedlmeier & Gigerenzer, 2001; Zhu & Gigerenzer, 2006). Improved Bayesian reasoning achieved by presenting information in natural frequencies (rather than conditional probabilities) can be further enhanced by teaching people to use frequency tree representations. For example, Sedlmeier and Gigerenzer (2001) teach one of two groups how to represent probabilities in terms of natural frequencies, supported by two visual diagrams – frequency grid and frequency tree (representation training) – while the other group was taught Bayes' rule for probabilities (rule training). Participants are then tested on tasks in which the statistical information is always provided in terms of probabilities. The representation training group improves from 10 to 90% Bayesian answers, while the rule training group improves from 0 to about 65% Bayesian answers. Bayesian answers in the representation training group remain high at 90%, even five weeks after training. In contrast, the rule training group's performance declines to about 20%. These results are achieved for basic Bayesian tasks (*i.e.* two hypotheses, and one dichotomous cue). The findings suggest a comparative advantage of representation training over rule training for robust learning over time.

It is an open question whether this type of training would increase the number of Bayesian updaters in the setting of selection neglect, which is a markedly different set of estimation tasks to previous learning environments. Moreover, the present paper's use of incentivised beliefs distinguishes it from the existing psychology literature on training in Bayesian reasoning with visual representation.

There is some evidence of transfer learning. Hoffrage et al. (2015) investigate whether learning from representation training with a basic Bayesian task transfers to complex Bayesian tasks (for which they do not receive any training). The researchers use simple written instructions on how to solve a basic task (as opposed to a computerised training program above). First, each participant receives a two-page instruction sheet on how to solve the mammography task, which is a basic task with two hypotheses and one dichotomous cue. Participants are then randomly assigned to one of three groups. For Group 1, the mammography task is presented in terms of probabilities, and training includes how to insert probabilities into Bayes' rule to solve the problem. For Group 2, the task is presented in terms of probabilities, and training includes how to translate probabilities into natural frequencies, put the natural frequencies into a frequency tree diagram, and calculate the answer from the tree. For Group 3, the task is presented in terms of natural frequencies only (no probabilities), and training includes instructions on how to solve it using the frequency tree. After studying their instructions, participants have to solve two complex Bayesian tasks. Participants in Groups 1 and 2 receive probability versions, while Group 3 receive natural frequency versions. Group 1 performed the worst (18% and 22%), a significant improvement was observed for Group 2 (40% for both tasks), and the top performer was Group 3 (73% and 81%, respectively).

A problem that goes unnoticed cannot be attempted, let alone solved. Often, we do not understand what health statistics mean or unwittingly draw erroneous conclusions. This inability to understand the meaning of numbers is pervasive (Gigerenzer et al., 2007). *"Collective statistical illiteracy is common to patients, journalists and physicians; it is created by nontransparent framing of information that is sometimes an unintentional result of lack of understanding but can also be a result of intentional efforts to manipulate or persuade people; and can have serious consequences* *for health"* (Gigerenzer et al., 2007: 53).[22] Interestingly, Enke (2020) shows that even when the information-generating process is transparent and known to individuals, selection neglect still occurs in a substantial amount of cases.

In this paper, I test whether instructing participants online in a novel Bayesian reasoning exercise increases the proportion of Bayesian updaters. The short educational activity is completed directly before the series of estimation tasks identical to Enke's setting. It comprises a story of a dice game played between yourself and another person during lockdown and uses visual representation training of Bayesian reasoning in the spirit of Hoffrage and Gigerenzer (1998), Zhu and Gigerenzer (2006), and Hoffrage et al. (2015). The original laboratory experiment by Enke (2020) is successfully replicated with an online student sample in the median regressions but not robust to Ordinary Least Squares (OLS) winsorised or OLS trimmed models. The novel *Training* treatment significantly attenuates the bias, shifting a majority of subjects towards the Bayesian benchmark. However, there is only weak evidence of this treatment effect since it is not robust to alternative regression specifications.

In this paper, at the median, the *Control* group is approximately at the rational benchmark ($\hat{\chi}_i^j$ = 0) as expected, the *Selected* group shows significant neglect (for reference, full neglect is $\hat{\chi}_i^j$= 1) as expected, and my novel *Training* treatment group exhibits significantly lower selection neglect compared to *Selected*. Note that $\hat{\chi}_i^j$ is the naïveté parameter defined by Enke's theoretical framework and is explained further below. However, at the mean, these differences between treatment groups are not significant. The median neglect parameter for *Training* is $\hat{\chi}_i^j$= 0.10, so the treatment reduces neglect by 0.50 units of $\hat{\chi}_i^j$ compared to *Selected*. This level of neglect is comparable to Enke's (2020) *Nudge* treatment in which participants were told explicitly on their decision screens: "*HINT: Also pay attention to those randomly drawn balls that are not shown to you by the information source."* Treatment *Nudge* reduces neglect by about 0.2 - 0.4 units of $\hat{\chi}_i^j$, which corresponds to about half of the difference between *Selected* and *Control* in the original lab experiment. Enke finds that in *Selected*, the median and average neglect are $\hat{\chi}_i^j$ = 0.50 each, while

[22] For example, the statement '1 less woman out of 1000 will die of breast cancer due to mammography screening' can be emotively reframed as 'mammography screening reduces the risk of dying from breast cancer by 25%'.

in *Nudge*, the median is $\hat{\chi}_i^j$ = 0.10 and the average is $\hat{\chi}_i^j$ = 0.30. This explicit nudge is arguably not a practicable policy, however. In a less explicit attentional nudge of endogenous information acquisition[23], Enke's treatment *Endogenous* generates lower levels of neglect than *Selected* (median $\hat{\chi}_i^j$ = 0.14, while the mean is $\hat{\chi}_i^j$ = 0.34), but is only marginally significant at the 10% level. Enke's additional mechanism treatments that reduced complexity in the set of signals – namely, *Simple* and *Few* – caused a reduction in neglect by about 0.2 – 0.3 units of $\hat{\chi}_i^j$. This indicates that the increased cognitive load has a systematic effect on how subjects approach the initial conceptual stage of forming a mental model. Interestingly, Enke (2020:15) finds that feedback on the true state and expected profits from a subject's submitted belief do not lead to less neglect. *Feedback* does not significantly reduce neglect beliefs between treatments, nor does neglect decrease across periods as participants accumulate feedback.

These previous findings provide some context as to what effectively shifts beliefs to be more rational and what does not. This paper builds on these insights to test a more practicable policy of visual representation training to increase the proportion of Bayesian updaters. The rest of the paper proceeds as follows: Section 2 details the experimental design, Section 3 presents the results of the replication and extension, and Section 4 concludes.

2. EXPERIMENTAL DESIGN

2.1. The Dice Game

Before reading the estimation task instructions (detailed below), subjects randomly assigned to treatment *Training* click through a short, interactive "warm-up" exercise called "The Dice Game", which I designed. It consists of a story in which your girl-/boyfriend challenges you to a game and asks you to guess the average of the six dice they threw (hidden to you). You have the chance to click on an interactive die to gain sampling experience with the uniform distribution. In the story,

[23] Subjects were required to deliberately choose between different information structures that provide predominantly high or low signals, respectively (Enke, 2020: 26).

your girl-/boyfriend reveals one randomly chosen die to you, and then reveals a subsample of the five remaining dice aligned with your prior belief (high or low). After you have made your guess, they explain how to solve the task the Bayesian way using visual representation, taking you through the logic step by step with words and pictures. For details of the game, refer to the Appendix.

2.2. The Estimation Tasks

Participants are asked to estimate an *ex ante* unknown state of the world θ. A computer had generated θ by drawing six times; six draws with replacement, from the set {50, 70, 90, 110, 130, 150}. The average of these six draws makes the true state θ, which Enke refers to as the "variable" that subjects need to estimate. I follow Enke's nomenclature and refer to the random draws as "signals". For the purpose of replication, I use the same predetermined true states of the world and original experimental instructions translated from the German.

During the experiment, a participant interacts with a computerised information source that shows her subsets of the signals. Table 1 summarises the multiple stages of the task as devised by Enke (2020). All subjects, no matter which treatment arm they are initially assigned to, follow this pathway:

(i) A subject is shown one randomly selected signal from the set of signals generated by a computer.

(ii) Based on the first signal, participants provide an incentivised guess b_1, which indicates whether they believe θ to be greater or smaller than 100, $b_1 \in$ {low, high}.

(iii) The information source reveals additional signals. This is the only stage in which baseline treatments *Selected* and *Control* differ, as detailed below in Section 2.3.

(iv) After subjects are shown messages from the information source, they enter an incentivised belief about the state $b_2 \in [50, 150]$, with at most two decimals.

TABLE 1

Overview of the experimental design

Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Computer determines state by drawing six signals	Subject receives one signal	First binary guess b_1 based on signal	Subject observes messages of information source	Continuous guess b_2

Note. This table replicates Table 1 in Enke (2020: 7).

In total, participants complete four periods (eight tasks). These are summarised in Table 2. The first guess (high/low) only serves the purpose of imposing a selection problem. The second guess is the subject of interest. The first guess is made incentive-compatible to reduce noise, *i.e.* earnings from the first guess would be maximised in expectation if they follow the first signal, stating a guess below (above) if the signal was below (above). This acts as an attention check for reading the instructions. 72.88% of 2400 observations (*i.e.* 1749 observations) followed the first signal.[24]

TABLE 2

Overview of the experimental tasks

True State	First signal	Observed Signal A	Observed Signal B	Observed Signal C	Observed Signal D	Unobs. Signal E	Unobs. Signal F	Bayesian Belief	Neglect Belief
96.67	130	130	150	70	-	50	50	103.33	122.00
110.00	150	110	150	110	-	50	90	110.00	130.00
93.33	50	90	50	130	-	110	130	96.67	80.00
90.00	110	150	90	50	-	50	90	90.00	100.00
103.33	150	110	130	70	-	70	90	100.00	115.00
116.67	90	90	70	150	-	150	150	110.00	100.00
116.67	110	150	130	150	110	50	-	120.00	130.00
96.67	130	130	90	110	-	70	50	90.00	100.00

Notes. This table replicates Table 2 in Enke (2020: 9). There is a typo in the original paper, first column on the left, bottom row. The table shows an overview of the belief formation tasks in order of appearance. The categorisation into observed and unobserved signals applies to the case in which subjects follow their first signal, *i.e.* guess ≥ 100 if their signal was larger than 100, and $<$ otherwise. Subjects in the *Selected* and *Training* treatment observed only their own signal and the "observed" signals. Subjects in the *Control* condition additionally had access to a coarse version of the "unobserved" signals, *i.e.* if the corresponding signal was less than 100, they saw 70, and if the signal was larger than 100, they saw 130. See derivation of the Bayesian and neglect benchmarks in text.

[24] "Control questions serve a very different purpose in online experiments than in lab experiments because in the former people are on average less attentive, so screening becomes more important" (Enke, 2021: correspondence 13.05.21).

2.3. The Treatments

Enke's between-subject design emphasises transparency and the full knowledge of participants when it comes to the data generating process. The experimenter has full control over the data generating process and can exogenously manipulate the degree of selection. A control group provides the benchmark of rational beliefs without selected information. In addition, there is incentive-compatible belief elicitation, where participants are rewarded for the accuracy of their beliefs. The two baseline treatments, *Control* and *Selected* from the original design, make up two sets of signals which result in the same Bayesian posterior, but only the *Selected* group has the problem of selected information.

Treatment *Selected* differs from *Control* in that it faces a budget constraint; a limitation on how many signals it can display. It has to condition its decision on which out of the remaining five signals to reveal on the subject's first guess, their prior b_1. In particular, if the participant's first guess is higher than 100, the information source displays all signals above 100. If the subject's first guess is smaller than 100, the information source reveals to her all signals below 100. In either case, the participant is always shown at least three signals.[25] The "missing" signals that subjects "do not see" are described as "coarse information" (Enke, 2020: 7). Given the discrete, uniform distribution of the known signal space {50, 70, 90, 110, 130, 150}, it is relatively straightforward to infer which types of signals are missing. Being sophisticated about selection requires that a subject understands that when she first guesses b_1 = *low*, a missing signal is 130, in expectation, while it is 70 if she guesses b_1 = *high*.

Enke designed the *Control* treatment to give the same Bayesian posterior as *Selected* without the issue of a selection problem. The participant faces two types of signals on her decision screens. First, she sees those signals that subjects in the *Selected* treatment are also shown. Second, she observes a coarse version of the signals that subjects in the *Selected* condition do not observe. If a missing signal is in {50, 70, 90}, the information source communicates 70 to her, while if the missing signal is in {110, 130, 150}, the information source communicated 130. These coarse

[25] "For example, if a participant's first guess was above 100 and only two of the remaining five signals were above 100, the information source showed the subject these two signals and one randomly selected signal of those below 100. If four signals were above 100, the subject would be shown (only) these four." (Enke, 2020: 7).

messages equal the expected signal conditional on a subject's first guess in *Selected*. This means that the informational content of the *Selected* and *Control* treatments is the same (Enke, 2020: 8).

Information presented in the third group, *Training*, is the same as *Selected*. For the purposes of evaluating the main treatment effect of interest in this paper, *Training* versus *Selected* is the relevant comparison. For the purposes of replication of the selection effect with an online sample, I also include the *Control*. Table 3 provides an overview of all treatments. Treatments are randomised within the session with equal probability to one of three treatment groups, *Control, Selected* or *Training*.[26]

2.4. Procedural details

The experimental participants were recruited using Prolific (www.prolific.co) [May 2021, pilot 1; 6 September 2021, pilot 2; 8 September 2021, Session 1; 13 September 2021, Session 2; 21 September 2021, Session 3] (Palan & Schitter, 2018). The experiment is programmed in oTree (Chen, Schonger, & Wickens, 2016).[27] Pre-screening filters on Prolific restrict the experiment to students aged 18-34, fluent in English with a 90% approval rating on the platform. The sample is 53% female, with a mean age of 22.6 years. To familiarise themselves with the task, all subjects scroll through written instructions taken from Enke (2020). There is no required reading time set nor a limitation on reading time in the online experiment. They can click on a link to download the instructions as a pdf, which opens in a new tab.[28] Experimental instructions are available in the Appendix.

[26] The small sample size in *Control* (one third of the total sample) reflects the fact that it serves a "straw man" with very little expected noise (Enke, 2020: 9).
[27] Enke's (2020) experiment was programmed in zTree (Fischbacher, 2007) and conducted in the BonnEconLab of the University of Bonn. My decision to go online allowed for data collection during Covid-19 lockdown which closed labs in Germany.
[28] Enke's participants had 10 minutes to read instructions in the lab.

TABLE 3

Treatment overview

Treatment	# of subjects	Ave. earnings (GBP)
Control	96	2.84
Selected	112	2.57
Training	92	3.17
Total	300	

Notes. This table shows variable earnings from decisions only. Total payments include a show-up fee of 5 GBP in all treatments, which is not included here. Treatments are randomised within the same experimental session with equal probability.

After the instructions page, participants complete five multiple-choice control questions. I devised these control questions to ensure participants understood the data generating process.[29] Only 28.33% of the sample made zero mistakes, 35% made 1 mistake, 24.33% made 2 mistakes, 10% made 3 mistakes and 2.33% made 4 mistakes.[30]

Subjects then enter the first round. Each round consists of two computer screens. On the first screen, subjects are informed of the first signal and issue a binary guess. The initial part of the estimation task effectively acts as an incentivised control for reading instructions and is paid out with 10% probability. On the second screen, participants receive messages from the information source and state a point belief.

All decisions are financially incentivised in expectation. In total, subjects make 16 decisions, one of which is randomly selected for payment. This constitutes an incentive-compatible reward mechanism in such a setup (Azrieli, Chambers, & Healy, 2018). The probability that a second (point) belief is randomly selected for payment is 90%, while one of the binary first guesses is chosen with probability 10%. The binary first guess is such that subjects receive 180 points if the guess is correct and nothing otherwise. Points are converted into British Pounds (GBP) at the end of the experiment, where 100 points correspond to 5 GBP. The continuous point beliefs are incentivised using a quadratic scoring rule with maximum variable earnings of 9 GBP, *i.e.* variable

[29] For example, participants were asked, "Assume that you submitted a first guess of larger than 100. Which draws will the information source show you no matter what? (a) None. (b) Those above 100. Those below 100."

[30] There were two feedback messages from prolific.co participants about the instructions being difficult to understand for second-language English speakers even though I pre-screened for English fluency. The original instructions could possibly have been clearer. In future replications, it may be helpful for the instructions to be rewritten to be more accessible to the average reader.

earnings in given task j equalled $\pi^j = max\ \{0; 18 - 0.2 \times (b^j - \theta^j)^2\}$, where b denotes the belief. There is also a show-up fee (fixed payment) equal to 5 GBP.

2.5. Theoretical Considerations

This subsection outlines Enke's simple formal framework underlying the experimental design of the model of selection neglect. This model is used for the empirical analyses in Section 3. The true state of the world is given by $\theta = \sum_{k=1}^{6} s_k /6$. Let N denote the number of signals a subject actually sees on their screen. Given a set of signals, a Bayesian updater would calculate the mean posterior belief b_B as

$$b_B = \frac{\sum_{v=1}^{N} s_v + \sum_{l=N+1}^{6} E[s_l|b_1]}{6}, \tag{2}$$

where s_v denotes a signal that appears on the decision screen, and s_l denotes an unobserved (coarse) signal.

If individuals simply base their beliefs on what they see (that is, the set s_v), this introduces a selection problem in the sense that $E[s_i] = 100$, but $E[s_l|b_1] \neq 100$. The set of signals that subjects do not observe is systematically different from the unconditional expectation. As an analogue to the benchmark of rational inference above, Enke defines the full neglect benchmark as:

$$b_N = \frac{\sum_{v=1}^{N} s_v}{4} \tag{3}$$

It is possible that individuals partially adjust for selection. Let $\chi \in [0,1]$ parameterise the degree of neglect such that $\chi = 1$ implies full neglect. Then, the belief formed b can be expressed as a weighted average of b_B and b_N plus decision noise ϵ:

$$b = (1-\chi)b_B + \chi b_N + \epsilon = b_B + \chi \frac{6-N}{6}(\bar{s}_v - \bar{s}_l) + \epsilon \tag{4}$$

$$\equiv b_B + \chi d + \epsilon, \tag{5}$$

where $\bar{s}_v \equiv 1/N \sum_{v=1}^{N} s_v$ is the average visible signal, $\bar{s}_l \equiv 1/(6-N-1) \sum_{i=N+1}^{6} E(s_l|b_1)$ the average expected "non-visible" (coarse) signal, and ϵ is a mean zero random computational error. That is, the systematic component of a subject's belief b can be expressed as the Bayesian belief plus a distortion term d times the neglect parameter χ. Throughout the paper, this formal framework is used to compute estimates of neglect $\hat{\chi}$ and decision noise $|\hat{\epsilon}|$. The treatment *Training* is expected to affect the neglect parameter as follows: $\chi_{Control} \leq \chi_{Training} < \chi_{Select}$. This leads to beliefs closer to Bayesian or completely eliminates the bias.

3. RESULTS

3.1. Replication and extension

The object of interest in Enke's analysis is a potential treatment difference in the second (point) beliefs that subjects state. In the present experiment, in *Control*, 70.57% of all first binary guesses follow the first signal and enter a high (low) first guess if the first signal is above (below) 100. In *Selected*, 72.77% follow the first signal.[31] In *Training,* 75.41% follow the first signal. The analysis was conducted in Stata/IC 15.

Table 4 presents an overview of the results in each of the eight tasks for the three treatment groups. The first two treatments (*Control* and *Selected*) serve as a replication of Enke (2020). The third treatment, *Training*, tests whether beliefs can be debiased with visual representation training in Bayesian reasoning. For ease of comparison, I provide the benchmarks of full neglect and Bayesian beliefs, respectively. Panel A considers the full sample, while Panel B considers those observations for which the participants follow the first signal in the first part of the estimation task, *i.e.* pass an incentive-compatible attention check. Considering Panel B, the median beliefs in the *Control* condition mostly follow the Bayesian prediction closely, suggesting that the experimental setup is probably not systematically misconstrued by participants. In the absence of selected information, subjects state rational beliefs in five out of eight tasks. In the *Selected* treatment, median beliefs

[31] As a comparison, across *Control* and *Selected* treatments, 93% of Enke's sample follow the first signal.

are mostly distorted away from the Bayesian benchmark towards the full neglect belief. In seven out of eight tasks, the distributions of beliefs significantly differ between *Control* and *Selected* treatments at least at the 5% level, and, in five out of eight tasks, beliefs differ at the 1% level (Wilcoxon rank-sum tests).[32] In four out of eight tasks, the median belief in *Training* is closer to the Bayesian belief than *Selected*, but the difference between the distributions is only significantly different in one out of eight tasks (Task 6) at the 10% level. In four out of eight tasks, the median belief is the same in *Selected* and *Training*.

TABLE 4

Overview of beliefs across tasks

Panel A: Full Sample								
True State	First Signal	Bayesian Belief	Neglect Belief	Median Belief *Control* Treatment	Median Belief *Selected* Treatment	Median Belief *Training* Treatment	p-value (*C v. S*) (Ranksum test)	p-value (*S v. T*) (Ranksum test)
96.67	High	103.33	122.00	93	110	105	0.0002***	0.4851
110.00	High	110.00	130.00	110	118	115	0.1327	0.5851
93.33	Low	96.67	80.00	97	90	90	0.0023***	0.3032
90.00	High	90.00	100.00	110	100	90	0.4234	0.1023
103.33	High	100.00	115.00	100	103	103	0.0039***	0.9878
116.67	Low	110.00	100.00	90	103	107	0.0092***	0.6820
116.67	High	120.00	130.00	120	120	123	0.7596	0.3919
96.67	High	90.00	100.00	100	110	102	0.0171**	0.0553*

Panel B: Followed first signal to estimate prior (binary) belief								
True State	First Signal	Bayesian Belief	Neglect Belief	Median Belief *Control* Treatment	Median Belief *Selected* Treatment	Median Belief *Training* Treatment	p-value (*C v. S*) (Ranksum test)	p-value (*S v. T*) (Ranksum test)
96.67	High	103.33	122.00	100	115	110	0.0002***	0.3202
110.00	High	110.00	130.00	110	120	120	0.0029***	0.3066
93.33	Low	96.67	80.00	97	90	90	0.0007***	0.2779
90.00	High	90.00	100.00	110	100	97	0.0104**	0.1460
103.33	High	100.00	115.00	100	110	105	0.0001***	0.9554
116.67	Low	110.00	100.00	90	100	100	<0.0001***	0.0957*
116.67	High	120.00	130.00	120	130	130	0.0618*	0.7177
96.67	High	90.00	100.00	100	110	107	0.0373**	0.1233

Notes. Overview of the estimation tasks in order of appearance. See Table 2 for details on the signals in each task as well as the computation of the Bayesian and full neglect benchmarks. High (low) first signals are defined as signals above (below) 100. Median beliefs are rounded to the nearest integer. The p-value (*C v. S*) refers to a Wilcoxon rank-sum test between beliefs in *Selected* and *Control*, while the p-value (*S v. T*) refers to the same test between beliefs in *Selected* and *Training*. Panel A includes the full sample (300 subjects). Panel B includes those observations for which the participant followed the first signal in the first part of the task, *i.e.* passed an incentive-compatible attention check (217 subjects).

[32] As a comparison, in all eight tasks, beliefs differ between treatments at least at the 10% level, and five out of eight at the 1% level in Enke (2020: 12).

3.2. Econometric analysis

In the remainder of the paper, treatment comparisons will be conducted by pooling the data across tasks. Enke (2020: 12) justifies this approach with a view to brevity and to remove potential multiple testing concerns. Pooling the data requires transforming the beliefs data into a scale that has the same meaning across tasks. For this purpose, Enke (2020) makes use of the simple belief formation rule, which has the additional advantage that, going forward, all estimated quantities will have direct theoretical counterparts. In particular, he uses equation (5) to compute the estimated neglect implied in the belief of subject i in task j:

$$\hat{\chi}_i^j = E\left[b_i^j\right] = \frac{b_i^j - b_B^j}{d^j} = \frac{6(b_i^i - b_B^j)}{(6 - N^j)(\bar{s}_v^j - \bar{s}_i^j)} \tag{6}$$

Enke (2020: 12) notes that this analytical tool corresponds to a simple linear transformation of the raw beliefs data (subtract the Bayesian belief and divide by the distortion term *d*, which is only a function of the signal realisations). This method thus only converts the data into a consistent interval, so that subjects' beliefs are, firstly, on the same scale across tasks and, secondly, can be directly interpreted as reflecting Bayesian ($\hat{\chi}_i^j$ = 0), full neglect ($\hat{\chi}_i^j$ = 1), or intermediate levels.

While $\hat{\chi}_i^j$, for individual i in task j, should in principle be between zero and one, in the experimental data not all observations lie within this interval. Enke (2020: 13) attributes this in part to typing mistakes and random computational errors. This produces outliers that are partly severe. Across the treatments in Table 3 (N=2334 belief statements), the minimum implied naïveté parameter $\hat{\chi}_i^j$ = -69 and the maximum $\hat{\chi}_i^j$ = 60 To avoid arbitrary exclusion criteria – while at the same time dealing with outliers – Enke (2020: 13) presents three different sets of regression specifications. I follow the same strategy as Enke. First, I present an analysis with median regressions that includes the full sample of beliefs, including large outliers. Second, I present an OLS analysis in which I winsorise the data at $|\hat{\chi}_i^j|$ = 3. That is, I replace each belief that is larger (smaller) than 3 (-3) by the corresponding value. This affects 17.92% of all 2400 observations. Third, I present an OLS analysis on a trimmed sample, where I drop all observations with $|\hat{\chi}_i^j|$ > 3.

Table 5 presents the results. In these analyses, the unit of observation is a subject-task, for a total of usually eight observations per subject, unless the subject does not enter a belief on time, in which case the observation would be missing. Timeout happened in only 0.25% of observations. The standard errors are clustered at the subject level. All regressions include experimental session fixed effects, exploiting random assignment into the three treatments within sessions. In Table 5 Panel A, the replication of *Control* versus *Selected* treatment is shown. In column (1), the median regression controls for session fixed effects. Column (2) adds a vector of controls with fixed effects for each experimental task interacted with the first guess (high / low) of the subject, as well as controls for individual characteristics. In columns (3)-(4), the dependent variable is winsorised at $|\hat{\chi}_i^j| = 3$, and I estimate OLS regressions. In columns (5)-(6), the sample excludes observations with $|\hat{\chi}_i^j| > 3$. Table 5 Panel B, columns (7)-(12), follows the same estimation approach but for *Selected* versus *Training*.

In Table 5 Panel A, the median regression coefficient is quantitatively large and suggests that relative to the control treatment, subjects in *Selected* exhibit a neglect of 0.25-0.50 units of $\hat{\chi}_i^j$. The winsorised OLS model does not show a significant selection neglect effect. Likewise, for the OLS trimmed model. The replication is successful to some extent. However, Enke's range of 0.4-0.6 units of $\hat{\chi}_i^j$ was robust across all three regression specifications, whereas my online sample was not.

In Table 5 Panel B, the median regression coefficient is negative and large, 0.23-0.25 units of $\hat{\chi}_i^j$, and becomes significant after adding controls for session and task-fixed effects as well as controls for participant characteristics. This suggests that the treatment works in the expected direction, shifting the median belief towards the Bayesian benchmark ($\hat{\chi}_i^j = 0$). Again, this significant effect is not apparent in the other two regression specifications.

The selection bias implies lower earning of subjects in the *Selected* treatment compared to those in *Control*. The expected profit over all eight belief formation tasks (*i.e.* the average hypothetical profit from each belief) is 2.01 GBP in *Selected*, 2.55 GBP in *Control*, and 2.65 GBP in *Training*. Actual profits, which include a show of fee of 5 GBP and depend on a random draw, are 7.57 GBP in *Selected*, 7.84 GBP in *Control* and 8.19 GBP in *Training*. Kruskal-Wallis equality-of-populations rank tests rejected the hypothesis that the distributions were equivalent: (1) *Control* versus *Selected* (p = 0.0065), and (2) *Selected* versus *Training* (p = 0.0003).

TABLE 5

Treatments Control *vs.* Selected*, and* Selected *vs.* Training

	Dependent variable: Neglect χ					
	Median regression		OLS winsorized		OLS trimmed	
Panel A:	(1)	(2)	(3)	(4)	(5)	(6)
0 if *Control*, 1 if *Selected*	0.50***	0.25*	0.025	0.0070	0.15	0.11
	(0.17)	(0.14)	(0.15)	(0.14)	(0.12)	(0.11)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE x prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	980	972	1008	1000	835	832
R^2	0.00	0.04	0.00	0.24	0.00	0.22

Panel B:	(7)	(8)	(9)	(10)	(11)	(12)
0 if *Selected*, 1 if *Training*	-0.25	-0.23**	-0.074	0	-0.10	-0.14
	(0.16)	(0.12)	(0.14)	(.)	(0.11)	(0.09)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE x prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	1110	1110	1144	632	979	979
R^2	0.00	0.05	0.00	0.24	0.00	0.28

Notes. Regression estimates with robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect that is implied in a given belief. In the top panel, the sample includes each of the subjects' eight beliefs in *Control* and *Selected* conditions (replication). In the bottom panel, the sample includes each of the subjects' eight beliefs in *Selected* and *Training* conditions (extension). Columns (1)-(2) and (7)-(8) report median regressions, and columns (3)-(6) and (9)-(12) are OLS regressions. In columns (3)-(4) and (9)-(10), the dependent variable is winsorised at $-3 \leq \hat{\chi}_i^j \leq 3$. In columns (5)-(6) and (11)-(12), the sample is trimmed at $-3 \leq \hat{\chi}_i^j \leq 3$. Controls include age, gender and highest education. The sample is restricted to participants who achieved at least 4/5 control questions correctly. The median coefficients remain significant, and OLS trimmed in the top panel becomes significant when mistakes in my control questions are ignored. * p < 0.10, ** p < 0.05, *** p < 0.01.

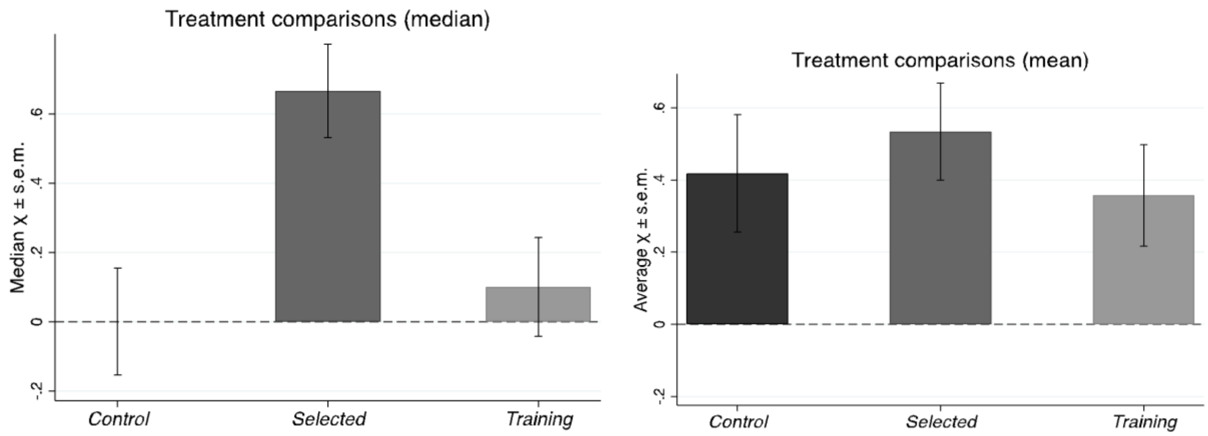
3.3. Shifting mental models: The rationalising effect of visual representation

Relative to *Selected*, subjects' median beliefs in *Training* tended to exhibit coefficients closer to the Bayesian belief. The negative sign indicates the treatment works to make beliefs closer to rational ($\hat{\chi}_i^j = 0$), as opposed to full neglect ($\hat{\chi}_i^j = 1$). This suggests a significant debiasing effect on beliefs of the visual representation training activity. However, there is only weak support for this treatment effect as the results are not robust to the other regression specifications.

Figure 1 gives an overview of neglect $\hat{\chi}_i^j$ across the three treatments. It plots the median (left) and mean (right) of all subject-task-specific $\hat{\chi}_i^j$ across treatments along with standard error bars. Motivation for this analysis is establishing grounds for a practicable treatment for debiasing beliefs in a setting in which participants do not seem to learn over the course of the experiment. Previously, in a feedback treatment arm, Enke (2020:15) finds that feedback on the true state and expected profits from a subject's submitted belief does not lead to less neglect. Indeed, feedback does not significantly reduce neglect beliefs between treatments, nor does neglect decrease across periods as participants accumulate feedback. For applied contexts, a relevant question is whether people can learn to reason the Bayesian way and not fall prey to the heuristic "what you see is all there is". My short training exercise draws on previous work on visual representation training with adaptations suited to convey Bayesian reasoning logic in the context of selection information.[33] It was an open question whether such training would be effective in a different context with incentivised beliefs.

Figure 1 shows that at the median, the *Control* group is roughly at the rational benchmark, the *Selected* group exhibits significant neglect, and the *Training* group shows significantly lower selection neglect compared to *Selected* treatment group. These between-group differences are not significant at the mean, however. The median for *Training* is $\hat{\chi}_i^j$= 0.10, so the treatment reduces neglect by 0.50 units of $\hat{\chi}_i^j$.

[33] Unlike in the original feedback treatment, assignment to *Training* treatment was randomised within the session.



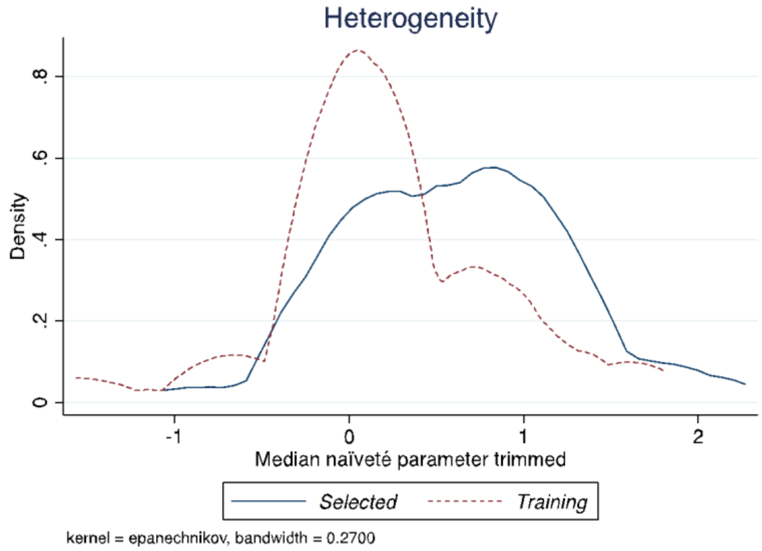
Notes. The left panel shows the median $\hat{\chi}_i^j$ across all subject-task observations. The right panel shows the average $\hat{\chi}_i^j$ across all subject-task observations, whereas, in columns (3)-(4) of Table 5, the data are winsorised at $|\hat{\chi}_i^j| = 3$. Standard error bars are computed based on clustering at the subject level. N = 680 – sample was restricted to participants who answered all my additional control questions correctly.

FIGURE 1

Overview of neglect across treatments

3.4. Heterogeneity Analysis

Type distribution. To characterise participants' belief patterns in more detail, I examine the subject-level distribution of neglect. In order to do this, I plot the median naïveté coefficient of each subject. Figure 2 presents kernel density plots of median naïveté parameter for each subject in treatments *Selected* and *Training*. The weight of the distribution of *Training* is shifted more towards the Bayesian belief ($\hat{\chi}_i = 0$) than full neglect ($\hat{\chi}_i = 1$). This is in contrast to *Selected*, which shows many more subjects forming naïve beliefs. Overall, these patterns suggest that the extreme types of $\hat{\chi}_i^j = 0$ and $\hat{\chi}_i^j = 1$ represent a substantial part of the empirical data compared to intermediate types. A Kruskal Wallis equality-of-populations rank test rejected the hypothesis that the distributions were equivalent at the 10% level (p=0.0677). This approximately bimodal distribution is consistent with Enke (2020: 17) and the literature from cognitive science and economic theory, which stresses the importance of misspecified mental models for belief formation. There are still many cases of partial adjustment of selected information, however. *Training* appears to be quite effective at shifting median beliefs towards the Bayesian benchmark.



Notes. Heterogeneity in the distribution of median neglect parameter $\hat{\chi}_i$ in *Selected* and *Training*. A subject's type is determined by taking the median implied neglect parameter over the eight tasks. The sample is restricted to subjects who answered all of my 5 additional control questions correctly. The sample is trimmed at $|\hat{\chi}_i^j|$=3.

FIGURE 2

Heterogeneity in distribution of types

4. CONCLUSION

This paper tests a practicable policy of visual representation training aimed to increase the proportion of Bayesian updaters in a setting in which a large proportion of individuals typically exhibit selection neglect in belief formation. The present online experiment replicates Enke's (2020) laboratory experiment, which manipulated the degree of selected information in treatments *Control* and *Selected*. In the extension, a novel visual representation exercise I designed is tested in treatment *Training*, in which individuals faced the same selected information problem as *Selected*. The replication of the original treatment groups is largely successful with the online sample of participants. The results aligned with Enke's finding that the majority of people can be classified into Bayesian updaters and full neglect types. As previously observed, there were substantial amounts of extreme outliers. These were addressed in several different regression models. Only the effects in the median regressions were significant, however. The coefficients were not robust to the OLS winsorised and OLS trimmed regression specifications.

The main contribution of this paper is to show that the selection neglect mental model can potentially be shifted towards a Bayesian one, with a simple Bayesian reasoning exercise that uses visual representation and a real-world scenario to explain abstract statistical problems. However, only weak evidence supports this treatment effect. The paper relates to other research that seeks to design practicable strategies for debiasing beliefs relevant to applied settings, e.g. the effect of team rather than individual decision-making (Monteiro, Praxmarer & Sutter, forthcoming) and a model of technological learning in which people "learn through noticing" (Hanna, Mullainathan, & Schwartzstein, 2014). A limitation of the online setting is lower attention than the laboratory and potentially poorer comprehension of the experimental instructions, judging from the high number of mistakes in my additional control questions. Pre-screening online participants for first-language-English speakers instead of English fluent would be advisable, if not a rewriting of the instructions to be at the high school reading level and more accessible. To my knowledge, this paper is the first to test visual representation training in an incentivised beliefs setting, that is, where participants are rewarded financially for accuracy. This is an important methodological contribution since belief accuracy is significantly higher with incentivisation (Gächter & Renner, 2010). The results weakly support literature from psychology (e.g. Hoffrage et al., 2015), which demonstrates that visual representation training can be a practicable means to combat collective statistical illiteracy in non-incentivised settings. A follow-up experiment with the *Selection* and *Training* groups would be potentially interesting to see whether any treatment effect remains after some time has elapsed.

Acknowledgements. Michael Kuelper provided excellent research assistance with programming the experiment in oTree.

Data. The data are available upon request from the author.

5. APPENDICES

A. Training Treatment Exercise (The Dice Game)

Notes: The following contain the text explanation and screenshots from the visual representation training. It proceeds over several screens which subjects click through.

SCREEN 1

Your task will be explained to you shortly. First, let’s warm up with a quick game.

Warm up: The Dice Game
Imagine you are stuck at home with your girlfriend, boyfriend or housemate during Covid-19. They come up with a game to amuse both of you. You decide to switch off Netflix, and agree to play, since the stakes are appealing. If you guess correctly you will not have to do the washing up for one week.

They grab six dice. They throw them, you hear them land on the table but you can’t see the what numbers they landed on.

They tell you, “your task will be to guess the average of the six dice”.

Click on the interactive dice a few times to experiment with the possible outcomes.

“I will give you several clues,” they say.

“Get ready for your first clue!”

CONTINUE .

SCREEN 2

"Here is your first clue." They randomly show you one of the dice.
Here it is, below. It was a 2.

They say, "do you think the average of the six dice is above or below 3.5? (That is, HIGH or LOW). You can base your decision on the die I just showed you.

You say, "LOW".

She says, "OK, great. Now I will show you three more dice. You will also be able to see ALL of the dice that fell on a LOW number."

Back .

CONTINUE .

SCREEN 3

The dice game

They show you the next three dice, including all the remaining dice that fell on LOW. You see a 3, a 1, and a 6.

Finally, she asks you, "guess the average of the dice I threw.

What is your best guess? 3

How certain are you about your belief? Rate how you feel from certain to not certain at all.

- ○ 1. "Certain! - I'm confident not doing the washing up for the next week"
- ○ 2.
- ○ 3.
- ○ 4.
- ◉ 5. "Not certain at all"

Continue

FIGURE A1

Example screenshot of the dice game

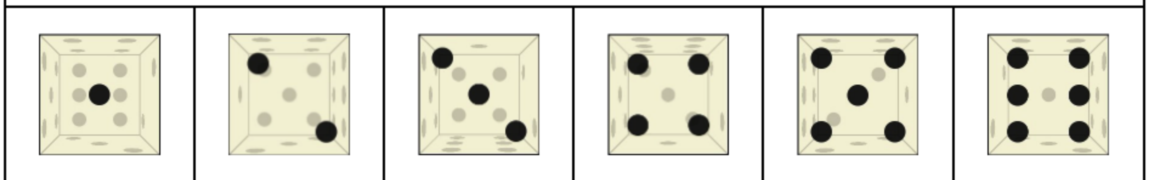
SCREEN 4

The dice game

Let's see how visual representation helps make this problem a bit easier.

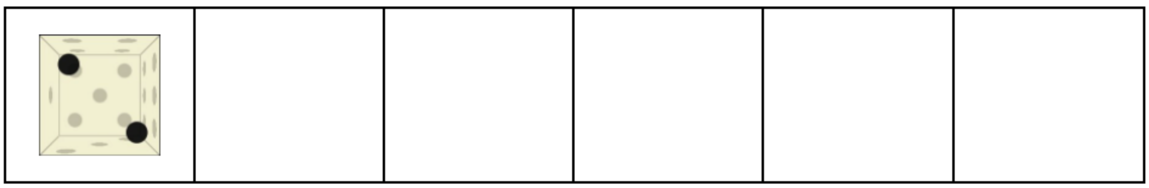
You know from experience that dice have numbers 1 through 6. Half are LOW and half are HIGH. Each number has an equal chance of showing up when your girl-/boyfriend throws the six dice.

ROLL					

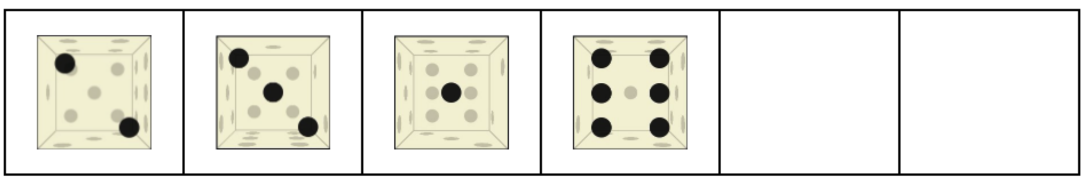


1	2	3	4	5	6
TOTAL					

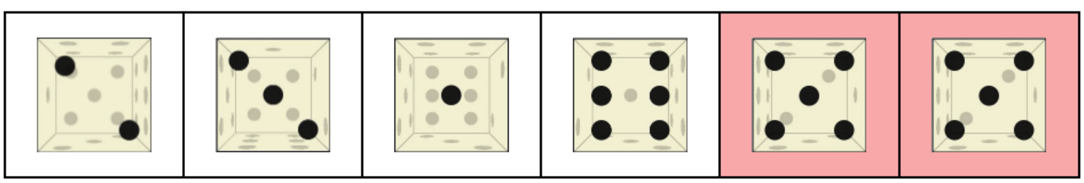
Your first random clue was LOW. The remaining dice could be LOW or HIGH. You don't know what they are yet.



Then they said, you will get three more clues and they would show you ALL the other LOW dice. Let's add these clues that were revealed.



But wait! Since you have seen ALL the LOW dice, you actually have one bonus clue. The remaining dice must be HIGH.



The expected value of the hidden HIGH dice is 5 (i.e. the midpoint of 4, 5 and 6).

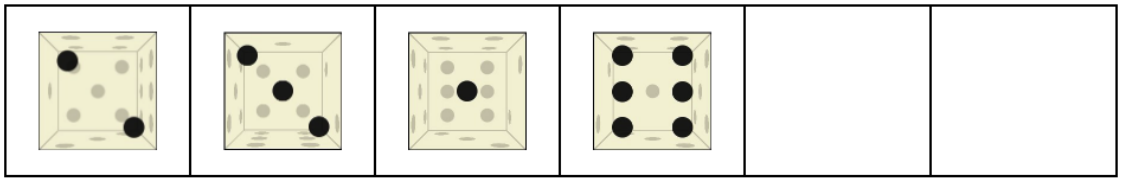
You can now use all the information to solve your girl-/boy-/friend's question:

What is the average of the dice I've thrown?

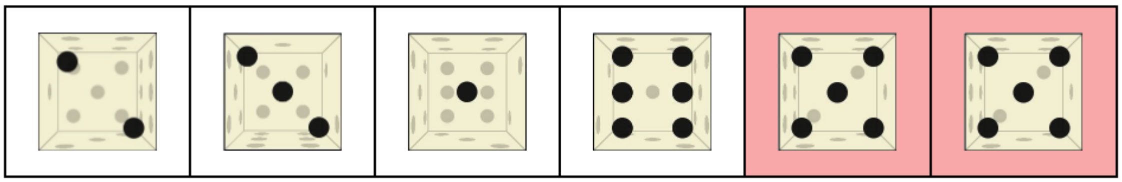
2 + 3 + 1 + 6 + 5 + 5 = 22/6 = 3 ¾ or 3.75

Contrary to your initial guess LOW (after seeing the first die only, which was a 2), the average of all the six dice is HIGH.

Then they said, you will get three more clues and they would show you ALL the other LOW dice. Let's add these clues that were revealed.



But wait! Since you have seen ALL the LOW dice, you actually have one bonus clue. The remaining dice must be HIGH.



The expected value of the hidden HIGH dice is 5 (i.e. the midpoint of 4, 5 and 6).

You can now use all the information to solve your girl-/boy-/friend's question:

What is the average of the dice I've thrown?

$2 + 3 + 1 + 6 + 5 + 5 = 22/6 = 3\frac{3}{4}$ or 3.75

Contrary to your initial guess LOW (after seeing the first die only, which was a 2), the average of all the six dice is HIGH.

FIGURE A2

Example screenshots of the dice game explanation

Notes: After this, participants proceed to the task below.

B. Experimental Instructions for Selected Treatments (including Training)

Your task

In this experiment, your task is to estimate two so-called "variables." In what follows, we will refer to these variables as variable A and B. You will receive information about these variables from an information source. Based on this information, you will need to provide your estimates. In what follows, we will explain how these variables are generated and which type of information you will receive.

How variables A and B get determined

Every variable is determined through random draws from an urn. **Figure 1** depicts these urns. Each urn contains exactly six balls with numbers 50, 70, 90, 110, 130, and 150. The letters are meant to help you in distinguishing between the urns.

The computer randomly determines variables A and B by drawing balls from the respective urn. From each urn, the computer draws six balls, i.e., six balls from urn A and six balls from urn B. Please note that, at each draw, each ball is equally likely to get drawn.

When a ball gets drawn, it gets replaced by another ball with the same number. That is, if the computer draws, say, a 130, then a new ball with number 130 is put into the urn before the next ball gets drawn. Thus, any given number can get drawn multiple times from the same urn.

Urn A	Urn B
A-50	B-50
A-70	B-70
A-90	B-90
A-110	B-110
A-130	B-130
A-150	B-150

Figure 1: The urn, from which the computer draws six balls each. Please note that balls that get drawn get replaced by another ball with the same number, so that every number can get drawn multiple times. Only the numbers in this table can get drawn

Thus, the computer draws six balls from each urn. The average of the six balls then equals the respective variable:

- The average of the six balls from urn A equals variable A.
- The average of the six balls from urn B equals variable B.

In this experiment, you need to estimate the value of variables A and B. As you can see, these variables are fully independent from each other, so that you cannot learn anything from one variable about the other one. Thus, you should always distinguish between these variables in the course of the experiment.

Your information
You receive your information from an information source. This information source does not draw balls from the urn itself. Rather, it observes all 12 balls that got drawn from urns A and B, i.e., all balls that determine the value of variables A and B. The experiment proceeds in multiple steps:

1. The information source observes the balls that got drawn from urns A and B.
2. For each variable, the information source shows you one of these randomly selected balls. Each ball is equally likely to be shown to you. You then have one piece of information about each variable, A and B.
3. Subsequently, you need to provide your first estimate about each variable. In this first step, you only need to estimate whether a variable is greater or smaller than 100. You can base this decision on the first ball that was shown to you by the information source. As will be explained to you in greater detail below, you will earn the highest amount of money on average if:
    - You estimate that the variable is greater than 100 if the first ball had a number greater than 100.
    - You estimate that the variable is smaller than 100 if the first ball had a number smaller than 100.
4. Subsequently, you receive further information from the information source. It shows you some balls for variable A and some balls for variable B. In doing so, the information source depends on your first estimates:

• If you estimated that a variable is greater than 100, the information source definitely shows you all balls with numbers greater than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers smaller than 100, all of which previously got drawn from the urn. The information source continues with this process until you have seen three balls per variable.
• If you estimated that a variable is smaller than 100, the information source definitely shows you all balls with numbers smaller than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers greater than 100, all of which previously got drawn from the urn. The information source continues with this process until you have seen three balls per variable.

5. This means that for each variable you will see at least three additional balls on your decision screen. In addition, as a reminder, you will also see the first ball that you had already seen in the first step.
6. Then, you need to provide an estimate for each variable. These estimates can take on any value between 50 and 150. You have a total of up to six minutes to do so. As will be explained to you in greater detail below, you will maximize your earnings with your second estimate if your estimate is as close as possible to the value of the respective variable.

We will implement this entire procedure four times. These four "rounds" are entirely independent from each other: each time, the variables A and B get drawn anew and you receive new information about these variables. This means that variables A and B are determined in each round separately, so that you cannot learn anything from one round about another one.

Your payment
In addition to your fixed payment, you will be paid based on your estimates.

In each round, you can earn up to 180 points with your first estimate if you correctly estimate whether the respective variable is greater or smaller than 100. You receive 0 points if your estimate is not correct.

In each round, you can also earn up to 180 points with your second estimate. The further away your response from the truth, the less you earn. This is determined by the following equation (in points):

$$\textbf{Earnings} = \textbf{180 - 2 * (Difference between estimate and truth)}^2$$

This means that the difference between your estimate and the truth gets squared and multiplied by 2. This value then gets subtracted from the potential maximum earnings of 180. While this formula may look complicated, the underlying principle is very simple: **the smaller the difference between your estimate and the true value, the higher your earnings.** However, your earnings can never be less than zero, i.e., you cannot incur losses. You can also see that your earnings only depend on the absolute difference. For example, it is hence immaterial for your earnings whether you over- or underestimate the true value by 5.

In total, you will provide 16 estimates in the course of this experiment (two for each of the two variables, in each of four rounds). The computer will randomly determine one of these estimates, and your payment will then depend on this estimate. In every round, one of your first estimates gets randomly selected with probability 10% and one of your second estimates with probability 90%. Thus, you should work on each estimate as well as you can because each estimate may be relevant for your payment.

IMPORTANT: Please note that in this experiment you maximize your earnings on average if you always truthfully report your estimates! Because only one of your decisions gets selected for payment, there is no point for you in, say, strategizing by sometimes providing a high and sometimes a low estimate. You should simply try to make the best decision possible to maximize your earnings.

Example
Suppose that the computer has drawn six balls from each urn and has thereby determined the values of variables A and B. The information source now shows you a first randomly selected ball. As depicted in **Figure 2**, you then need to estimate, for each variable, whether it is greater or smaller than 100.

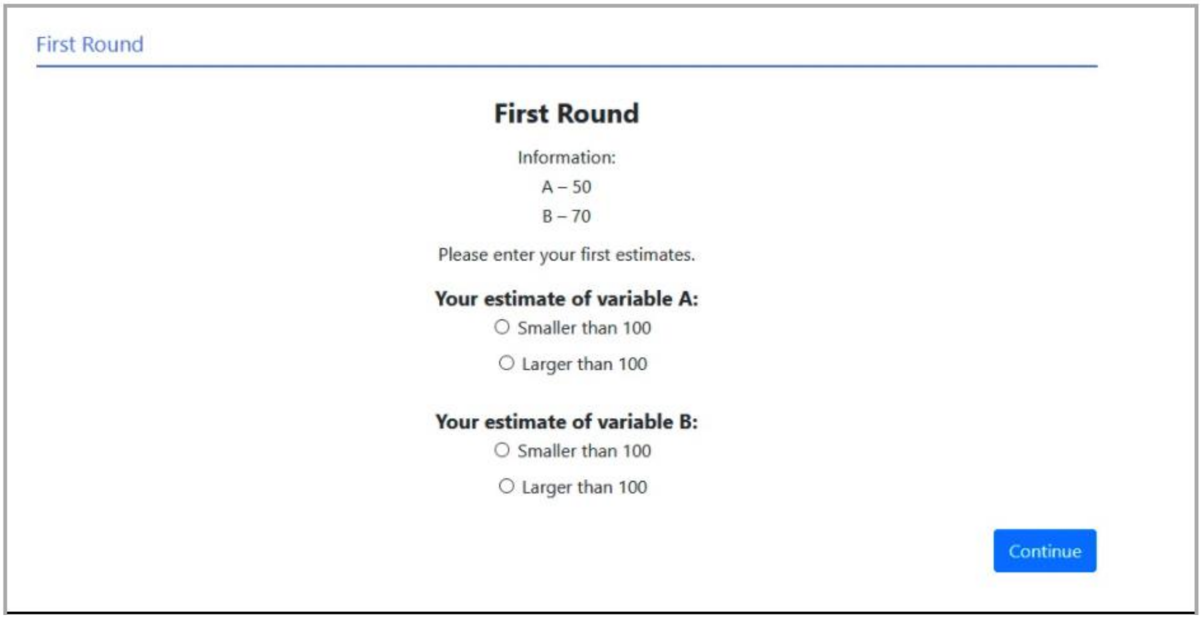


Figure 2. Example screenshot for the first estimates

Subsequently, the information source shows you additional balls. **Figure 3** presents an example. As you can see, you also get reminded of the first ball that you have already seen on the previous screen.
Then, you need to provide an estimate about each variable.

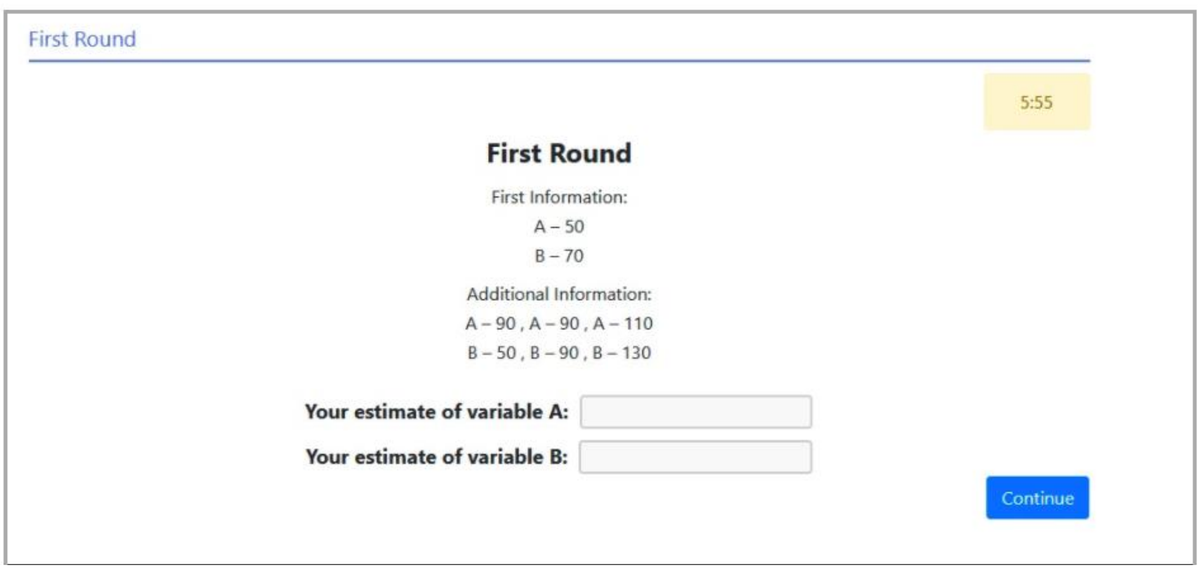


Figure 3. Example screenshot for the second estimates.

You can get a piece of paper and pen and write personal notes for yourself if you like.

Are you ready? If you are finished reading the instructions, then click OK to begin the experiment.

C. *Experimental Instructions for Control Treatment*

Your task

In this experiment, your task is to estimate two so-called "variables". In what follows, we will refer to these variables as variable A and B. You will receive information about these variables from an information source. Based on this information, you will need to provide your estimates. In what follows, we will explain how these variables are generated and which type of information you will receive.

How variables A and B get determined

Every variable is determined through random draws from an urn. **Figure 1** depicts these urns. Each urn contains exactly six balls with numbers 50, 70, 90, 110, 130, and 150. The letters are meant to help you in distinguishing between the urns.

The computer randomly determines variables A and B by drawing balls from the respective urn. From each urn, the computer draws six balls, i.e., six balls from urn A and six balls from urn B. Please note that, at each draw, each ball is equally likely to get drawn.

When a ball gets drawn, it gets replaced by another ball with the same number. That is, if the computer draws, say, a 130, then a new ball with number 130 is put into the urn before the next ball gets drawn. Thus, any given number can get drawn multiple times from the same urn.

Urn A	Urn B
A-50	B-50
A-70	B-70
A-90	B-90
A-110	B-110
A-130	B-130
A-150	B-150

Figure 1: The urn, from which the computer draws six balls each. Please note that balls that get drawn get replaced by another ball with the same number, so that every number can get drawn multiple times. Only the numbers in this table can get drawn

Thus, the computer draws six balls from each urn. The average of the six balls then equals the respective variable:

- The average of the six balls from urn A equals variable A.
- The average of the six balls from urn B equals variable B.

In this experiment, you need to estimate the value of variables A and B. As you can see, these variables are fully independent from each other, so that you cannot learn anything from one variable about the other one. Thus, you should always distinguish between these variables in the course of the experiment.

Your information
You receive your information from an information source. This information source does not draw balls from the urn itself. Rather, it observes all 12 balls that got drawn from urns A and B, i.e., all balls that determine the value of variables A and B. The experiment proceeds in multiple steps:

1. The information source observes the balls that got drawn from urns A and B.
2. For each variable, the information source shows you one of these randomly selected balls. Each ball is equally likely to be shown to you. You then have one piece of information about each variable, A and B.
3. Subsequently, you need to provide your first estimate about each variable. In this first step, you only need to estimate whether a variable is greater or smaller than 100. You can base this decision on the first ball that was shown to you by the information source. As will be explained to you in greater detail below, you will earn the highest amount of money on average if:
   • You estimate that the variable is greater than 100 if the first ball had a number greater than 100.
   • You estimate that the variable is smaller than 100 if the first ball had a number smaller than 100.
4. Subsequently, you receive further information from the information source. More specifically, the information source in some way shows you all balls that determine variables A and B:
   • Case A: If your first estimate was greater than 100:
     –Then, the information source definitely shows you all balls with numbers greater than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers smaller than 100 until you have seen three balls per variable.
     – In addition, the information source shows you a "70" for all remaining balls with numbers smaller than 100, which corresponds exactly to the midpoint of this interval.
   • Case B: If your first estimated was smaller than 100:
     – Then, the information source definitely shows you all balls with numbers smaller than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers greater than 100 until you have seen three balls per variable.
     – In addition, the information source shows you a "130" for all remaining balls with numbers greater than 100, which corresponds exactly to the midpoint of this interval.
   • Thus, in some way, you will see all variables that determine variables A and B. It is not important for you why you observe these balls in different ways.
5. This means that for each variable you will see five additional balls on your decision screen. In addition, as a reminder, you will also see the first ball that you had already seen in the first step.
6. Then, you need to provide an estimate for each variable. These estimates can take on any value between 50 and 150. You have a total of up to six minutes to do so. As will be explained to you in greater detail below, you will maximize your earnings with your second estimate if your estimate is as close as possible to the value of the respective variable.

We will implement this entire procedure four times. These four "rounds" are entirely independent from each other: each time, the variables A and B get drawn anew and you receive new information about these variables. This means that variables A and B are determined in each round separately, so that you cannot learn anything from one round about another one.

Your payment
In addition to your fixed payment, you will be paid based on your estimates.

In each round, you can earn up to 180 points with your first estimate if you correctly estimate whether the respective variable is greater or smaller than 100. You receive 0 points if your estimate is not correct.

In each round, you can also earn up to 180 points with your second estimate. The further away your response from the truth, the less you earn. This is determined by the following equation (in points):

$$\textbf{Earnings} = \mathbf{180 - 2 * (\textbf{ Difference between estimate and truth})^2}$$

This means that the difference between your estimate and the truth gets squared and multiplied by 2. This value then gets subtracted from the potential maximum earnings of 180. While this formula may look complicated, the underlying principle is very simple: **the smaller the difference between your estimate and the true value, the higher your earnings**. However, your earnings can never be less than zero, i.e., you cannot incur losses. You can also see that your earnings only depend on the absolute difference. For example, it is hence immaterial for your earnings whether you over- or underestimate the true value by 5.

In total, you will provide 16 estimates in the course of this experiment (two for each of the two variables, in each of four rounds). The computer will randomly determine one of these estimates, and your payment will then depend on this estimate. In every round, one of your first estimates gets randomly selected with probability 10% and one of your second estimates with probability 90%. Thus, you should work on each estimate as well as you can because each estimate may be relevant for your payment.

IMPORTANT: Please note that in this experiment you maximize your earnings on average if you always truthfully report your estimates! Because only one of your decisions gets selected for payment, there is no point for you in, say, strategizing by sometimes providing a high and sometimes a low estimate. You should simply try to make the best decision possible to maximize your earnings.

Example

Suppose that the computer has drawn six balls from each urn and has thereby determined the values of variables A and B. The information source now shows you a first randomly selected ball. As depicted in **Figure 2**, you then need to estimate, for each variable, whether it is greater or smaller than 100.

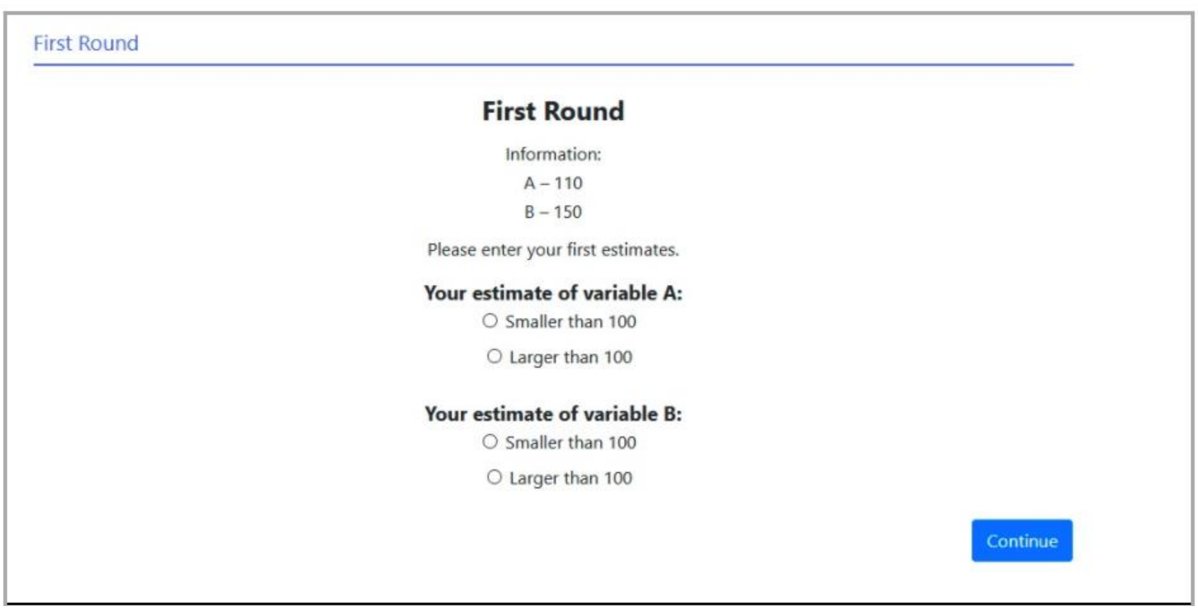


Figure 2. Example screenshot for the first estimate

Subsequently, the information source shows you additional balls. **Figure 3** presents an example. As you can see, you also get reminded of the first ball that you have already seen on the previous screen.

Then, you need to provide an estimate about each variable.

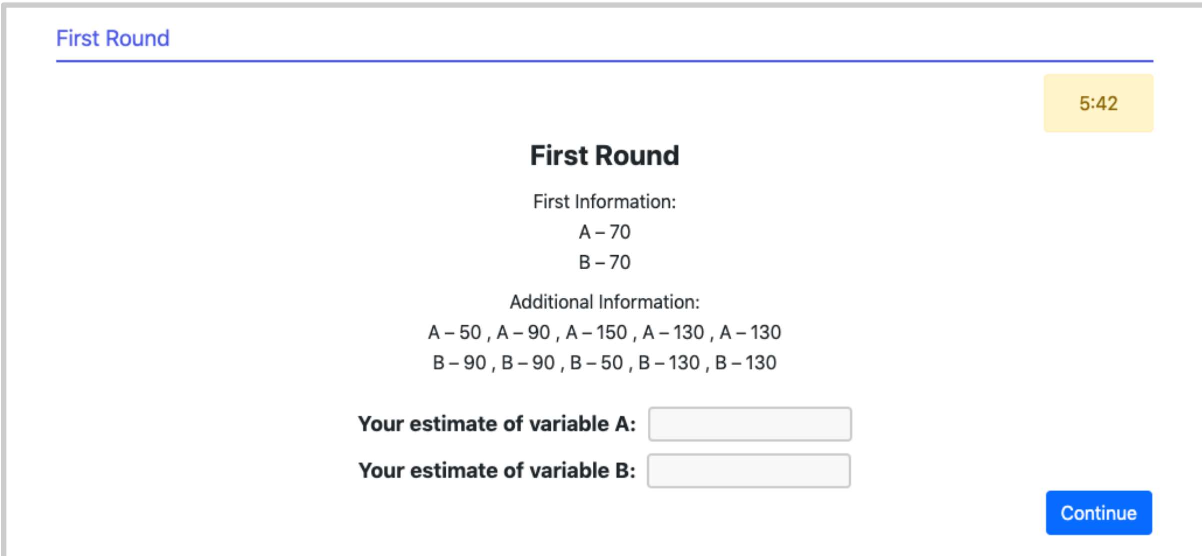


Figure 3. Example screenshot for the second estimate

You can get a piece of paper and pen and write personal notes for yourself if you like.

Are you ready? If you are finished reading the instructions, then click OK to begin the experiment.

6. REFERENCES

Azrieli, Y., Chambers, C. P., & Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, *126*(4), 1472–1503. https://doi.org/10.1086/698136

Bohren, J., & Hauser, D. (2017). Bounded rationality and learning: A framework and a robustness result. *Working Paper Papers.Ssrn.Com*.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2017). Memory, attention, and choice. *The Quarterly Journal of Economics*.

Bushong, B., & Gagnon-Bartsch, T. (2016). Learning with misattribution of reference dependence. *Working Paper Scholar.Harvard.Edu*.

Charness, G., Oprea, R., & Yuksel, S. (2021). How do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment. *Journal of the European Economic Association*, *19*(3), 1656–1691. https://doi.org/10.1093/jeea/jvaa051

Chen, D., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioural and Experimental Finance*, *9*, 88–97.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, *86*(1), 313–332.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics, 135*(3), 1363-1398.

Eyster, E., & Rabin, M. (2010). Naive herding in rich-information settings. *American Economic Journal: Microeconomics*, *2*(4), 221–243. https://doi.org/10.1257/mic.2.4.221

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, *129*(4), 1661–1710. https://doi.org/10.1093/qje/qju024

Gächter, S., & Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, *13*(3), 364–377. https://doi.org/10.1007/S10683-010-9246-4

Gagnon-Bartsch, T., Rabin, M., & Schwartzstein, J. (2018). *Channeled attention and stable errors*. Boston: Harvard Business School.

Gennaioli, N., & Shleifer, A. (2010). What comes to mind. *Quarterly Journal of Economics*, *125*(4), 1399–1433.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, Supplement*, *8*(2), 53–96. https://doi.org/10.1111/J.1539-6053.2008.00033.X

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, *129*(3), 1311–1353.

Heidhues, P., Kőszegi, B., & Strack, P. (2018). Unrealistic expectations and misguided learning. *Econometrica*, *86*(4), 1159–1214. https://doi.org/10.3982/ECTA14084

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*.

Hoffrage, U., Krauss, S., Martignon, L., & Gigerenzer, G. (2015). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Frontiers in Psychology*, *6*(OCT), 1473. https://doi.org/10.3389/FPSYG.2015.01473/FULL

Jehiel, P. (2001). Analogy-based expectation equilibrium. *Journal of Economic Theory*, *123*(2), 81–104.

Kurzenhäuser, S., & Hoffrage, U. (2002). Teaching Bayesian reasoning: an evaluation of a

classroom tutorial for medical students. *Medical Teacher*, *24*(5), 516–521. https://doi.org/10.1080/0142159021000012540
Monteiro, S., Praxmarer, M., & Sutter, S. (forthcoming). Correlation neglect in belief formation with team versus individual decision making. *Working Paper*
Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioural and Experimental Finance*, *17*, 22–27.
Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, *12*(6), 1423–1452.
Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology*, *130*(3), 380–400. https://doi.org/10.1037/0096-3445.130.3.380
Spiegler, R. (2016). Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, *131*(3), 1243–1290. https://doi.org/10.1093/qje/qjw011
Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287–308.

Chapter 3. Improving food preferences through a nutrition education programme: An evaluation comparing survey evidence with behavioural measures

Sofia Monteiro*[1,2], Georgina Pujol-Busquets Guillén[3,4], James Smith[5] & Kate Larmuth[5]

[1]Max Planck Institute for Research on Collective Goods, Bonn, Germany
[2]University of Cologne
[3]Catalunya Open University UOC, Division of Nutrition, Health Sciences Faculty, Barcelona
[4]University of Cape Town division of Physiology, Department of Human Biology, Faculty of Health Sciences, University of Cape Town
[5]University of Cape Town division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Sciences, University of Cape Town

Abstract. There are doubts that what people say is what they actually do. There is reasonable concern that self-reported assessments of dietary intake do not reflect actual intake. Yet, a correspondence between both is imperative to evaluate any interventions on food preferences. This paper makes such a comparison. It provides the first evidence from a nutrition education programme, which is evaluated with both survey measures and actual behavioural measures of food choice. The main result is that there is a large correspondence between survey and behavioural measures.

JEL classification: C81, C83, I12, I14, N37, Q18
Keywords: Methodology, Survey Methods, Food choice behaviour, Women, Education, Nutrition

* Corresponding author: sofia.g.monteiro@gmail.com

1. INTRODUCTION

According to the Global Burden of Disease Study in 2015, much of non-communicable diseases (NCDs) is driven by what people eat and drink, so influencing dietary behaviour for the prevention and management of NCDs is a key challenge for policymakers (Wang et al., 2016). For example, how do we persuade people to eat better and lose weight? (Roberto & Kawachi, 2015). How do we make the healthier choice the easier choice? (Hallsworth, 2017). There is arguably a need for evidence-based, sustainable dietary interventions that consider how people allocate the scarce resources of money and time to improve health (Cawley, 2004). Yet, on top of the challenge to change people's dietary preferences, there is the additional concern of how to measure such changes properly. While this is typically done with self-reported survey measures, there is reason to believe that what people say is not always what they do – as is, for instance, well known with regards to physical activity (Prince et al., 2008), but also debated with respect to dietary intake (Archer et al., 2018; Medic et al., 2016). In this paper, we provide both an intervention to study how to improve food preferences and an evaluation of its effects by measuring them with two different instruments – a survey and a behavioural measure – in order to study how well aligned what people say is to what they actually do.

Epidemiological observational studies, such as the much-cited Nurses' Health Study, have significantly influenced public health policy and practice globally (Liu et al., 1999; Colditz and Hankinson, 2005; Crous-Bou et al., 2014; Colditz et al., 2016). Such studies of dietary risk factors use assessments that require subjective responses, namely the Food Frequency Questionnaire (FFQ), 24-hour food recalls and food diaries, to make conclusions about the optimal diet for populations. It is understood that subjective measures of food preferences may suffer from hypothetical bias and a lack of incentive for accuracy (Lusk and Schroeder, 2004; Alemu and Olsen, 2018), and in Food Frequency Questionnaires, there may be noise due to people's inattention to what they eat, inability to recall fully, and social desirability bias (Archer et al., 2018; Medic et al., 2016). The motivation for the present research is to provide the first evidence on whether a behavioural measure of food choice corresponds to what people say they do in the Food Frequency Questionnaire, and in so doing provide a more credible channel to understand whether a programme may be said to affect dietary behaviour change. Our study contributed to an emerging literature on the behavioural economics of food choice (Ginon et al., 2014; List and Samek, 2015; Crosetto et al., 2016; Müller and Prevost, 2016; VanEpps et al., 2016; Alemu and Olsen, 2018; Charness et al., 2020).

Various diets are touted both in the nutrition literature and in public opinion, but the quality of evidence for them varies (Freire, 2020). We posit that the behavioural measurement of food choice is a complementary tool that could increase methodological credibility of nutrition intervention studies. There is, however, no gold standard for directly assessing the validity of the FFQ (Cade et al., 2002). Moreover, an evidence base founded on self-report survey instruments is arguably problematic (Archer et al., 2018). Therefore, we pose the question, do revealed food preferences corroborate self-report responses in the FFQ? This is an open empirical question since in another health-related behaviour, physical activity, no correlation between self-report responses and behaviour is found (Prince et al., 2008). The present cross-sectional study compares food choice behaviour of 44 women that had taken part in a nutrition education programme to 51 similar women who had not yet taken part but qualified for the programme, using complimentary diet assessment tools.

We collaborated with the Eat Better South Africa! (EBSA) programme in Ocean View and Atlantis, two under-resourced communities in the Western Cape.[34] Past EBSA programmes were six weeks long and involved weekly two-hour educational sessions at a central community hall with a group limited to about 30 women. These sessions aim to teach participants about nutrition, NCDs, shopping on a budget, cooking and how to access healthier foods. Peer support via the instant messaging group is central to the programme, and engagement in the group chat continues after the six-week course ends. In our previous qualitative study, the nutrition education programme is evaluated using a different methodology, namely focus group discussions on women’s perceptions of the programme, lifestyle choices and shopping habits (Pujol-Busquets Guillen et al., 2020).

[34] EBSA is a non-profit organisation, created to address health and socioeconomic issues faced by underprivileged South Africans, particularly women.

In the present research, the behavioural component is a grocery shopping activity at a local supermarket where the women could spend a ZAR 250 (~16 USD) retail voucher. Participants take home whatever food they buy and send photos of their groceries and the receipt via instant message. Eligible participants are drawn from the communities where the programme already operates or plans to operate in the near future. Adult women could participate in the study if they had taken part in the programme or were eligible to take part in future programmes. Thus, two groups of women are included and their responses compared.

Data were collected in 2020 during a Covid-19 lockdown. Despite the adverse conditions, we observe a marked difference in food choice behaviour by the programme group in line with its advice. While only an approximation of usual eating habits, the study shows that the inexpensive FFQ is representative of women's revealed food preferences. The results from this research suggest that when it comes to food: what she says is what she does. Future nutrition research could make use of complementary diet assessment tools that take into consideration revealed preferences in an ecologically valid setting. In particular, how people choose to allocate scarce resources of time and money when faced with lifestyle choices (Cawley, 2004).

2. METHODS

2.1. Study procedures

Participation in the study takes about 90 minutes. Each participant completes: (a) informed consent; (b) a questionnaire on socioeconomic status, medical conditions, food insecurity and shopping habits; (c) a task, in which they purchase food and (non-alcoholic) drink items in a local supermarket with a retail voucher and photograph their groceries and receipt; (d) a Food Frequency Questionnaire; and (e) a feedback questionnaire about their experience. Interview 1 comprises (a) and (b), while Interview 2 later in the week comprises (d) and (e).

2.2. Participant eligibility

Eligible participants are: women; 18-69 years old; capable of providing informed consent; able to understand and speak English or Afrikaans, not using private health care, and without access to health insurance. To be eligible as a previous programme participant, they must have attended at least four out of six weekly sessions from one of the four previous Western Cape programmes. Controls are drawn from similar communities through nomination to the study by a participant or EBSA community coach/ambassador. This is typically how EBSA recruits future programme participants who are relatively naïve to its Low-Carbohydrate High-Fat diet.

2.3. Participant recruitment

Our sampling strategy allows us to identify the effect of participating in the programme. To form a convincing counterfactual to the treatment group, control participants should be similar in observable characteristics. The educators and community coaches are asked to assist with recruiting women who were enrolled in previous programmes, either using the WhatsApp group or by telephone. For the non-programme participants, a person identified in the community is asked to help recruit women and participants are invited to nominate a non-programme woman to take part in the study. Those women interested in taking part are given the appropriate participant information sheet electronically (Appendix A, Control; Appendix B, Treatment). Women are not necessarily expected to read it but are asked if an investigator may contact them to explain the study. Women who communicate their willingness to participate in the study and meet the recruitment criteria are all contacted by a researcher. The infographic explaining the 10 steps to complete the study (Appendix C) is sent electronically to each person's mobile phone. A researcher confirms eligibility, explains the study, answers any questions and invites interested women to take part.

2.4. Data

The participant receives a R250 (~16 USD) voucher via SMS. This amount equates to about a week's typical grocery shopping in our underprivileged communities. The shopping activity is held at a major local supermarket. Participants can choose from a selection of grocery items (at current retail prices), in particular from fresh produce and cupboard groceries. The researcher explains the activity according to Interview 1 Transcript (Appendix D) and conducts the demographic questionnaire (Appendix E). Each participant has a week to purchase her groceries at her convenience. An online shopping account with which the vouchers are purchased allows the investigators to track which vouchers have been spent. Purchasing decisions are made privately in the store, as per regular grocery shopping. The participant has access to her phone to keep track of the total cost of her shopping. When she has finished selecting items, she proceeds to the checkout. At home, she takes a photo of both her groceries and the till slip, and sends them to the researcher. The researcher ticks off on a checklist with the participant's code the types of food she chose to buy and notes the total spent on the receipt.

When the shopping activity is completed, the follow-up interview is booked, usually on the same day or the following day. Participants complete an interviewer-administered FFQ on a phone call with a researcher asking about foods eaten in the past four weeks (Appendix F). This questionnaire was developed and used previously as an online FFQ. The FFQ was developed according to guidelines proposed by Cade et al. (2002). The South African Medical Research Council's (MRC) FFQ (Steyn and Senekal, 2004) was adapted to include food items frequently eaten by people following a Low-Carbohydrate High-Fat (LCHF) diet and foods reportedly eaten by previous programme participants. It also included standard portion sizes and frequency options. It underwent a period of pilot testing in volunteers that habitually followed a LCHF diet and was modified accordingly (Webster et al., 2019). The FFQ data was primarily used to assess the types of foods eaten.

2.5. Data cleaning and variable transformation

To measure average dietary intake, the FFQ asks respondents how frequently they have eaten a standard portion of a particular food in the past 4 weeks. For example, "In the past 4 weeks, how often did you eat a slice of white bread?" to which the participant may choose one of nine fixed responses, e.g. "None", "1 or less per month", "2-3 per month, "1-2 per week", 3-4 per week", "5-6 per week", "1 per day", "2-3 per day" or "3 or more per day". For analysis, categorical variables are converted into continuous numerical variables, in particular, an estimate of the number of standard portions eaten per week. e.g. if a participant responded "1 per day", this was converted to "7 per week".

The food purchases that the sample of women make with a retail voucher are coded as a list of dummy variables. If a food item (e.g. bread) is purchased, 1 is captured, and 0 otherwise. The list of food types corresponds to the EBSA programme's traffic lights list of foods (Appendix G) and the FFQ items for ease of comparison. 5 Control observations are excluded from the analysis because these women were part of a gym group run by an EBSA community coach and already eating according to the programme's dietary guidelines. A RED index score is created that adds up the binary food choice variables, with a maximum score of 8 if all RED foods are purchased with the R250 voucher. An ORANGE index score is created that adds up the binary food choice variables, with a maximum score of 3 if all three ORANGE list foods are purchased with the R250 voucher. A GREEN index score is created that adds up the binary food choice variables, with a maximum score of 8 if all eight GREEN list types of food are purchased with the R250 voucher.

2.6. Data analysis

Individual characteristics are examined to motivate that the Control group can be used as a valid counterfactual to evaluate the impact of the programme. The non-parametric Kruskal-Wallis H test is conducted to test for significant differences in the distributions of individual and household characteristics. The self-report dietary intake of participants is categorised into RED, ORANGE and GREEN traffic-light lists published by EBSA (Appendix G). In order to evaluate the impact of the programme on women presumably attempting to follow the recommended LCHF diet, the types of foods eaten by the Control and treatment groups are compared. If the programme influenced eating habits according to its recommendations, one would expect to see fewer RED food items and more GREEN items consumed by the treatment group compared to the Control group. We control for observable characteristics in the Ordinary Least Squares (OLS) regression analysis with robust standard errors. The types of food purchases of the Treatment and Control groups are compared. In the Linear Probability Model (LPM) regression analysis, we control for observable characteristics. Finally, we consider the relationship between surveyed responses and purchasing behaviour, which allows us to make an evaluation of the validity of the FFQ. If the FFQ correlates well with our behavioural measure, this would support its use as a valid instrument to assess the impact of the programme on food preferences.

3. RESULTS

3.1. Sample characteristics

The Treatment (EBSA programme) and Control groups are balanced in terms of household characteristics: number of household members, number of employed household members, shopping frequency and food insecurity.[35] About half of all participants worried they would not have enough food because of lack of money in the preceding four weeks. In addition, Table 1 shows that the two groups are well balanced and similar on a number of observable individual characteristics. Participants predominantly reported their ethnic heritage as Cape Coloured ($\chi 2(1) = 1.639$, $p = 0.2005$). According to the 2011 Census, the sample is representative of the ethnicity and language of our Western Cape communities (Ocean View and Atlantis). It should be noted, however, that the treatment group is older on average ($\chi 2(1) = 10.760$, $p = 0.0010$). The mean age for the Control group is about 44 years old, while the mean age for the programme group is about 51 years old. In the regression analysis, we control for observable characteristics.

Almost all the women have children (mode: 2; max: 5), and 75% share a household with a partner. The women are usually responsible for household groceries ($\chi 2(1) = 1.639$, $p = 0.2005$). On average, households have four members, with one or two members typically employed. 53% of the Control group are employed compared to 32% in the Treatment group ($\chi 2(1) = 4.251$, $p = 0.0392$). We examine if there is a difference between the two groups in employed, working-age individuals (*i.e.* less than 60 years old). The test shows no significant difference between the two groups at the 5% level, $\chi 2(1) = 2.833$, $p = 0.0924$. In the pooled sample, 43% completed primary school only, 35% completed high school and 21% higher education. The Treatment group has

[35] See Supplementary Table S1 in the Appendix.

more participants with primary school only (less educated), and this difference is significant at the 5% level, $\chi 2(1) = 5.226$, $p = 0.0222$. However, there is no significant difference at the 5% level between the groups in basic education (where primary school and high school are pooled), $\chi 2(1) = 3.378$, $p = 0.0661$. In South Africa, the returns to education are convex, in that the marginal rate of return is extremely high for tertiary levels of education and approaches zero for lower levels of education (Keswell and Poswell, 2004). Thus, completing school versus not is arguably the more relevant characteristic with which to consider differences in life chances between the two groups. There are no significant differences between groups in NCD risk factors at the 5% level, however, the Control group has a lower incidence of diagnosed high blood pressure at the 10% level.

TABLE 1

Sample characteristics of participants

Treatment Group	All (N = 95)	CONTROL (n = 51)		EBSA (n = 44)
Age (years)	47,04	43,76	***	50,84
Cape Coloured	0,90	0,94		0,86
Black, Indian or White	0,09	0,06		0,14
Shares HH with partner	0,75	0,75		0,75
Mother	0,95	0,96		0,93
Buys HH groceries	0,91	0,94		0,86
Employed	0,43	0,53	**	0,32
Employed & working age	0,48	0,43	*	0,38
Covid unemployed	0,16	0,15		0,16
Education				
No formal schooling	0,01	0,00		0,02
Primary school completed	0,43	0,37	**	0,55
High school completed	0,35	0,37		0,32
College/Uni completed	0,19	0,27		0,09
Post graduate degree	0,02	0,02		0,02
Risk factors				
Smoker (daily)	0,09	0,12		0,07
Alcohol (past year)	0,40	0,39		0,41
High BP diagnosed ever	0,42	0,33	*	0,52
BP meds in past 2 weeks	0,23	0,16		0,32
Diabetes diagnosed ever	0,09	0,12		0,07
CVD event or chest pain	0,08	0,04		0,14

Note: Summary means rounded to two decimal places.
Kruskal-Wallis H tests are reported as a balance check between CONTROL and EBSA groups. Significant differences are starred.
* shows p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01.

3.2. Women's self-reported dietary intake in the past four weeks

Figure 1 shows histograms of total weekly consumption of standard portions of foods from the RED-list (e.g. fast food, sugar, bread, pasta, processed seed oils), ORANGE-list (e.g. fruit, such as bananas and apples) and GREEN-list (e.g. eggs, unprocessed meat, unsweetened full cream dairy, green leafy vegetables), respectively.[36] In each histogram, blue bars indicate the distribution of the Control group's responses, gold bars indicate the distribution of the Treatment group's responses, and brown bars indicate where the two group's distributions overlap. The Treatment group's total weekly consumption of RED foods is clearly left-shifted compared to the Control group (*i.e.* fewer standard portions of unhealthy items were consumed, Kolmogorov-Smirnov test: Treatment < Control, $p < 0.001$), while for GREEN foods, the weight of the distributions is reversed (Control < Treatment, $p < 0.001$).[37] In general, for ORANGE list foods, there is not much interesting variation. However, as we did not have particular hypotheses for the consumption of ORANGE list foods, we focus on the GREEN and RED lists in the main paper and discuss ORANGE in the Supplementary Information for completeness.

---

[36] See Appendix A for the EBSA programme's traffic-lights lists which summarise its low-carbohydrate dietary advice.

[37] For total consumption of standard portions of ORANGE list foods there is no difference (Treatment < Control, $p = 0.500$). We had no strong prior expectation for the ORANGE list.

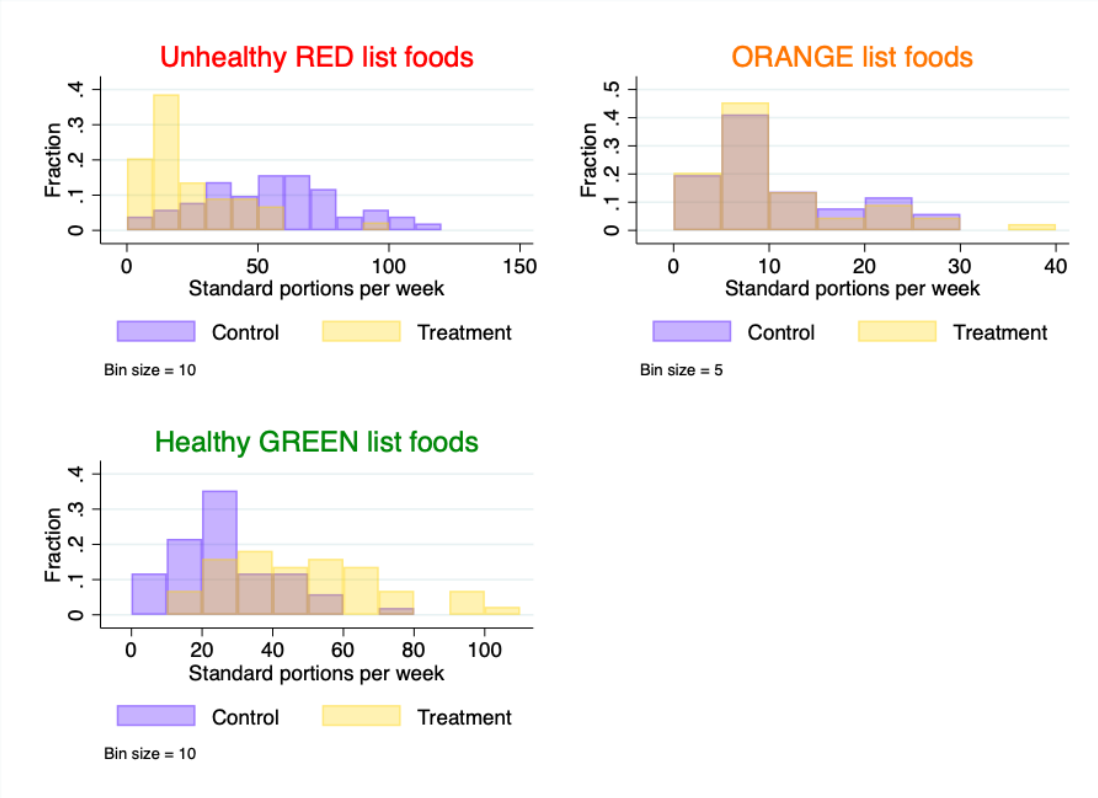
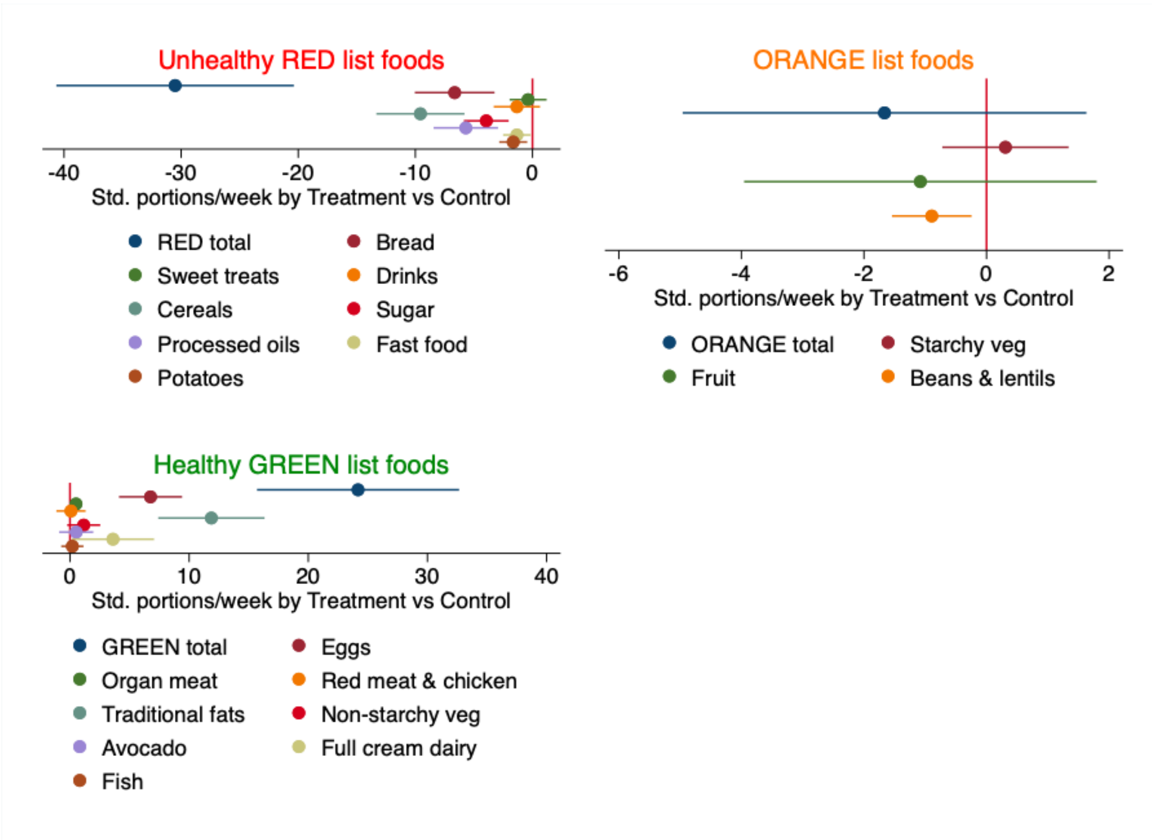


FIGURE 1

Histograms of reported intake from the various food lists

Figure 2 shows coefficients from Ordinary Least Squares (OLS) regressions of the number of standard portions of food consumed per week on a treatment dummy variable = 1 if Treatment, and 0 if Control, controlling for age and highest education. Each regression coefficient illustrated originates from a separate model. We model the treatment effect of participating in the nutrition education programme on total weekly intake of RED, ORANGE, and GREEN list foods, respectively. The treatment effect on particular foods should be interpreted with caution because of multiple hypothesis testing concerns with a sample of 95 women. The OLS models in Figure 2 consider only the self-reported FFQ data. The corresponding regression results can be seen in Table 2.

The treatment group reports significantly fewer total portions per week from the RED list: *i.e.* 56 in Treatment versus 86 unhealthy standard portions per week in the Control, a 35% lower intake by the Treatment group. Conversely, the treatment group reports significantly higher intake from the GREEN list.[38]



Note: OLS regression coefficient plots include controls for age and education.

FIGURE 2

Post-intervention treatment effect on weekly intake from the various food lists

[38] "*Eggs in particular were highly regarded as a versatile food. Participants expressed an appreciation that eating certain fats like butter and animal fat was permitted more freely with the EBSA diet and said that this improved the taste of their food. Adding coconut oil or butter to coffee (Bulletproof coffee) as a meal substitute was reported quite often.*" (Pujol-Busquets Guillen et al., 2020: p. 7).

TABLE 2

OLS models of treatment effect on total weekly intake of unhealthy and healthy foods

	Red Total	Orange Total	Green Total
EBSA group	-30.512***	-1.664	24.177***
	(5.11)	(1.66)	(4.27)
Age (years)	-0.486*	0.066	0.027
	(0.24)	(0.08)	(0.20)
Education level	-4.930	-1.313	5.181*
	(3.08)	(1.00)	(2.58)
Constant	86.087***	10.813*	15.170
	(14.16)	(4.60)	(11.84)
Observations	95	95	95
R-squared	0.353	0.037	0.279

Standard errors reported in parentheses
* p<0.05 ** p<0.01 &*** p<0.001

Result 1. *Women who have participated in the nutrition education programme reported significantly lower (higher) weekly intake of standard portions of unhealthy RED (healthy GREEN) list foods than the Control group.*

3.3. Women's incentivised food choices with a supermarket voucher

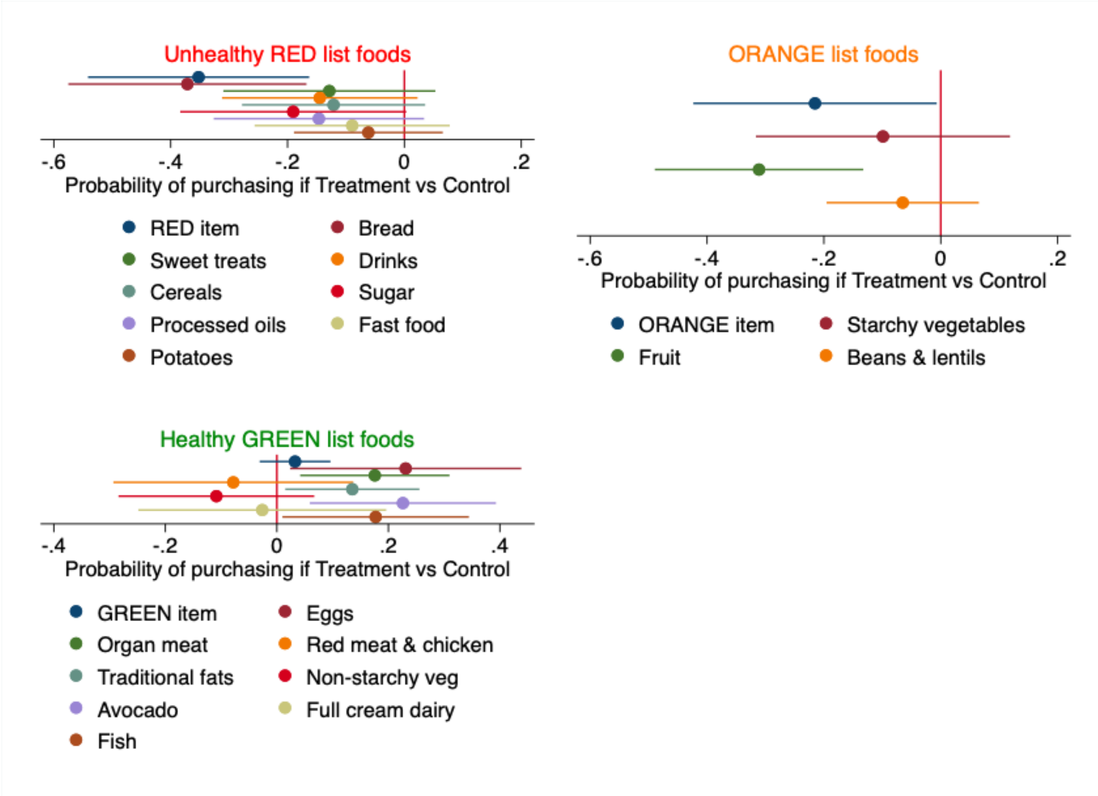
In the Control group, 86% of participants purchase at least one RED-list item with their voucher, compared to 52% of programme participants. On average, the Control group selects 2.39 (sd = 1.42) out of eight different types of RED list foods, while the Treatment group selects 1.02 (sd = 1.23). The modal RED index score in the Control is 3 (out of 8), while the Treatment group's modal RED index score is 0. A Kolmogorov-Smirnov test rejects the hypothesis that the two distributions are equivalent (Treatment < Control; $p < 0.001$).

In the Control group, 96% of participants purchase at least one GREEN-list item with their voucher, compared to 100% of programme participants. On average, the Control group selects 2.45 (sd = 1.06) out of 8 possible GREEN foods, while the treatment group selects 3.43 (sd = 1.32). The modal GREEN index score in the Control group is 3 (out of 8), while the Treatment group's modal GREEN index score is 4. A Kolmogorov-Smirnov test rejects the hypothesis that the two distributions are equivalent (Control < Treatment, $p < 0.001$).

The following coefficient plots are from Linear Probability Model (LPM) regressions of binary dependent variables equal to 1 if the food is purchased, and 0 otherwise, on a treatment dummy = 1 if Treatment and 0 if Control. These models use only the behavioural food choice data from receipts, verified by photos of groceries. All the LPM models include controls for age and highest education. The second main result confirms the first main result: The nutrition education programme influences not only self-reported eating habits but also behaviour.

In Figure 3, the LPM model "RED" shows that programme participants are about 40% less likely to purchase a RED-list item with the retail voucher. The LPM model "GREEN" shows that programme participants are as likely as the Control group to purchase at least one GREEN list food. It is necessary to consider the separate LPM models of each of the GREEN list foods to learn more about the likelihood of healthy food choices. Programme women are significantly more likely to spend their voucher on five out of eight GREEN items compared to the Control group.

Result 2. *Women who had participated in the programme were less likely to purchase a RED (unhealthy) item compared to the Control group. The Treatment group bought more GREEN (healthy) items.*



Note: LPM coefficient plots include controls for age and education.

FIGURE 3

Probability of purchasing foods from the various lists by treatment group

3.4. The relationship between incentivised choices at the supermarket and self-reported dietary intake in the past four weeks

The following coefficients in Figure 4 are from OLS models of total weekly intake of RED, ORANGE and GREEN list foods, respectively. Controls for age and highest education are included in all regressions. In the first model, total weekly intake is regressed on a binary variable, e.g. "RED" = 1 if one or more RED foods are purchased with the voucher, and 0 otherwise. This model allows us to test the hypothesis that buying any RED food with a voucher is positively associated with reporting to have eaten RED food in the past 4 weeks. In the second model, total weekly reported intake is regressed on an index score variable which reflects the number of different types of RED list items purchased with the voucher. This model allows us to test the strength of the association between buying more types of RED foods and reported eating habits in the past month. Buying more types of RED items should also be represented in greater reported consumption of RED food if the behavioural measure validates the survey.

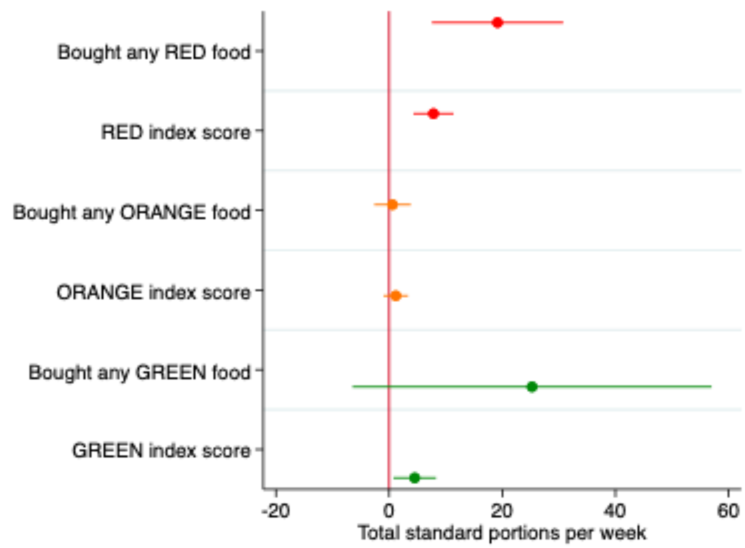
Figure 4 shows that the association between behavioural and survey responses is significant and positive for RED items. Buying any RED-list food is associated with an increase in total weekly intake of about 20 standard portions.[39] There are eight different types of food in the RED list that were used to generate the RED index score. A one-unit increase in the RED behavioural index is significantly associated with reporting eight more standard portions of RED list foods in the survey. Overall, buying RED in the supermarket is associated with greater reported total weekly RED intake in the past four weeks.

The association between behavioural and survey responses is significant and positive for GREEN list foods. There are eight different types of food from the GREEN list, which form the behavioural GREEN index score we generated. A one-unit increase in the GREEN index score (*i.e.* buying one more GREEN product) is significantly associated with reporting five more standard portions of GREEN in the survey.[40]

In general, the relationship between purchasing behaviour and reported food intake tends to be positive and significant. The more unhealthy or healthy products bought with the voucher, respectively, the more the women report to have consumed in the past month. This finding validates the FFQ as a tool to evaluate eating habits. Therefore, we infer there is utility in measuring the impact of the nutrition education programme using the FFQ in this population.

---

[39] An example of a standard portion of a RED list food in the FFQ is “a slice of white bread”.
[40] An example of a standard portion of a GREEN list food in the survey is “an egg”.



Note: OLS regression coefficient plots control for age and education

FIGURE 4

The association between purchasing behaviour and reported dietary intake

Result 3. *Spending the voucher on at least one RED (unhealthy) item is associated with 20 more unhealthy standard portions reported in total weekly consumption in the past four weeks. A one-unit increase in the behavioural RED/GREEN index score is significantly positively associated with surveyed consumption.*

4. CONCLUSION

This paper examines the question of methodology in studies of food choice and, in particular, how to evaluate an intervention designed to improve healthy eating. Much of global NCDs, such as type 2 diabetes, are driven by dietary behaviour (Wang et al., 2016). Behaviour is the mechanism or channel through which to explain changes in blood markers, weight and blood pressure. Thus, being able to show that a nutrition education programme affects revealed food preferences is critical for demonstrating evidence of sustainable behaviour change in lifestyle. In hugely influential observational studies, such as the Nurses' Health Study (e.g. Colditz and Hankinson, 2005), the measurement of dietary intake is self-reported in Food Frequency Questionnaires at intervals (*i.e.* every two years). Such measures are open to bias (Archer et al., 2018), and it is an empirical question whether the FFQ would be validated by revealed food preferences when spending a retail voucher on real food in the supermarket. Previously, Lusk and Schroeder (2004) showed that hypothetical responses predict higher probabilities of purchasing beef steaks than non-hypothetical responses, suggesting that people overestimate their "willingness-to-pay". The present study contributes to an emerging literature on the behavioural economics of food choice (e.g. Ginon et al., 2014; List and Samek, 2015; Crosetto et al., 2016; Müller and Prevost, 2016; VanEpps et al., 2016; Charness et al., 2020).

We address three questions: (a) Does self-reported dietary intake in the past month differ between past programme participants and a similar Control group of women? (b) Does incentivised decision-making with a supermarket voucher differ between women who have been through the programme versus Control? (c) Do incentivised food choices validate self-reported dietary intake in the past four weeks? Participants spend a voucher valued at one week's shopping and complete an interviewer-administered Food Frequency Questionnaire over the phone. Data are analysed using the programme's traffic-lights lists. The treatment group reports 30 fewer RED (unhealthy) standard portions per week and 24 more GREEN (healthy) portions per week. Women from the nutrition education programme are also less likely to purchase an unhealthy item and instead buy more healthy items. Spending the voucher on an unhealthy RED food is associated with reporting an additional 20 unhealthy food portions per week. Our results on reported versus behavioural dietary measures are in contrast to Prince et al. (2008), which shows no correlation between reported and behavioural measures in physical activity. The researcher's independence from the education programme and anonymity of responses was emphasised to participants to address concerns of possible experimenter demand effects. The study design puts the behavioural measure first, so there is no concern about participants trying to match what they buy to their survey responses, although the reverse is possible. Participants may potentially have the desire to be consistent, but there is no reason to believe that this desire would be greater amongst our participants than Prince et al. (2008). We show how the credibility of evaluations can be strengthened

using incentivised measures of revealed food preferences when they corroborate inexpensive complementary diet assessment tools such as the Food Frequency Questionnaire.
This study contributes to the limited literature on behavioural economics of food choice, building on behavioural experiments such as List and Samek (2015), which examined incentives for improving children's snack choices and Charness et al. (2020), which examined the impact of several non-incentivised interventions on children's snack choices. Our study offers interdisciplinary behavioural insights into the nutrition literature by examining the validity of a standard survey instrument for measuring food intake. Building on Cawley's (2004) economic framework for understanding eating behaviour, our study empirically shows that economics offers useful insights into nutrition because it is the study of how people allocate their scarce resources of time and money. We show that the EBSA programme influences food choice behaviour and how this methodology could strengthen the quality of evidence generated by impact evaluations since it is reflected in women's survey responses about their eating habits.

While purchased groceries are likely strongly correlated with actual food intake, it is not a measure of direct intake. Food could be wasted or given to other household members. When asked, *"Are there any grocery items that you bought for someone else that you don't plan to eat at all yourself?"* less than 20% of participants answered in the affirmative. When analysing food choice behaviour, we did an "intention to eat" analysis since we cannot rule out that the women would not eat the treats they bought since all groceries become part of the home food environment. A second limitation of our shopping activity is the value of the supermarket voucher. What women choose to buy with R250 (16 USD) is certainly not a complete picture of their dietary intake. However, we argue that the purchases we observe are nevertheless a reasonable sample of the types of foods our participants would buy given a limited budget. Allowing for unlimited food expenses would also not be a valid measure of their usual behaviour. Our voucher is equivalent to about a week's shopping value, which is less than what participants normally spend on a particular shop.[41] Over 90% of the sample answered affirmatively to *"Is what you bought similar to what you normally buy?"*. The choice of R250 was taken with consideration of the research budget and achieving the

[41] The modal shopping frequency reported by participants was bulk buys with R1000 or more, followed by R500 - R700 and R300 - R500.

sample size needed for statistical analysis. A third limitation is the cross-sectional nature of the data.

Acknowledgements. We thank our research assistant Laris Gomez for her excellent work. Helpful comments were received from members of the EEG MPI group, as well as attendants at ASHEcon 2021. This research was funded by Sofia Monteiro's research budget at The Max Planck Institute for Research on Collective Goods, Bonn. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone.

Data. All anonymised data used in this study are available from the corresponding author upon request.

Code. All data analysis was carried out in Stata SE 15. All figures were produced in Stata. The code used for data analysis and figure production is available from the corresponding author upon request.

Author Contributions. Conceptualisation, S.M., K.L., P-B.G, J.S., analysis and data collection S.M., writing – original draft preparation S.M.; writing – review and editing S.M., K.L., project administration S.M., funding S.M.

Ethics Declarations. The authors declare no competing interests.

5. SUPPLEMENTARY INFORMATION

5.1. Supplementary Table S1: Household Characteristics

TABLE S1

Balance test of household characteristics

Treatment Group	All (N = 95)	CONTROL (n = 51)	EBSA (n = 44)
Number of household members	4,15	4,16	4,14
Number of employed household members	1,43	1,45	1,41
HH grocery shopping frequency			
- Less than once a week	0,44	0,41	0,48
- Once a week	0,44	0,41	0,48
- 2-3 times per week	0,09	0,14	0,05
- Every day	0,02	0,04	0,00
Food Security in the household (past 4 weeks)			
1. Worried you would not have enough food	0,47	0,51	0,43
2. Not able to eat the kinds of food you wanted	0,46	0,43	0,50
3. Not able to eat any luxury foods	0,46	0,53	0,39
4. Had to eat a limited variety of foods	0,48	0,51	0,45
5. Had to eat some foods you did not want to	0,33	0,35	0,30
6. Had to eat smaller meals	0,44	0,45	0,43
7. Had to eat fewer meals	0,43	0,43	0,43
8. Ever no food of any kind in your household	0,09	0,12	0,07
9. Any HH member went to sleep hungry	0,03	0,04	0,02
10. Any HH member went 24 hours without food	0,02	0,04	0,00
Number of affirmative answers /10	2,77	2,88	2,64

Note: Summary means rounded to two decimal places. Kruskal-Wallis H tests are reported as a balance check between CONTROL and EBSA. Significant differences are starred.
* shows p-value < 0.1, ** shows p-value < 0.05, *** shows p-value < 0.01.

5.2. The ORANGE list foods

As we did not have particular hypotheses about consumption of the ORANGE list foods, we provide discussion of the ORANGE list here in the Appendix for completeness. There is no significant difference in total weekly intake of ORANGE list foods between the two groups in the past four weeks. In the Control group, 66.67% of women purchase at least one ORANGE food item with their voucher, compared to 54.55% of programme participants. On average, the Control selects 0.94 (sd = 0.79) out of 3 and the treatment group selects similarly 0.61 (sd = 0.62) out of 3. The modal ORANGE index score is 1 for both groups. A Kolmogorov-Smirnov test fails to reject the hypothesis that the two distributions are equivalent (Treatment < Control; $p = 0.134$).

There is no significant difference between the two groups' ORANGE list purchases. The LPM model "ORANGE" indicates that programme participants are about 20% less likely to purchase an ORANGE list food. The association between behavioural and survey responses is not significant for ORANGE. Buying any ORANGE list food is not significantly associated with an increase in total weekly intake. There are three different foods in the ORANGE list that form the ORANGE index score we create. A one-unit increase in the ORANGE index score is not significantly associated with reporting more standard portions of ORANGE list foods in the survey. Overall, buying ORANGE is not significantly associated with reporting greater total weekly intake of ORANGE in the FFQ.

6. APPENDICES

A. Participant information sheet Control group

Appendix I, HREC REF 295/2019, Version 3, 3 June 2020

STUDY INFORMATION SHEET - Group A Participants

Behavioural and self-report measures of food choice in women who have taken part in a nutrition education program

Dear potential study participant,

You have been contacted because you are a woman who is eligible for a future Eat Better South Africa (EBSA) nutrition education program in your community. We are researchers from the University of Cape Town (UCT) and the Max Planck Institute for Research on Collective Goods (MPI) and we would like you to take part in our study on measuring food choices. Below you will find more information about what is involved. Please feel free to contact us if you have any questions or would like to take part.

Georgina Pujol-Busquets Guillén (UCT)	georginapbg@gmail.com	061 659 2692
Sofia Monteiro (MPI)	monteiro@coll.mpg.de	066 311 8981
James Smith (UCT)	ja.smith@uct.ac.za	083 3057593
Kate Larmuth (UCT)	kateus65@gmail.com	074 217 0042

What is the study about?

This study will use two different methods to measure the food choices of people to better understand nutrition and shopping habits in communities like yours.

Are you eligible to take part?

You are allowed to participate if you are a woman; are 18-69 years old; can understand and speak English or Afrikaans; do not have private healthcare or medical insurance and are able to visit a local supermarket and do a telephone survey interview (we will give you details about the time and place). You must routinely shop for your household in order to participate.

What will be required of you?

If you want to take part, you will be given a R250 shopping voucher to spend on foods and non-alcoholic drinks at Pick n Pay. You can keep the groceries you buy but we ask that you send us a photo of everything you bought, and the till slip. You will also be asked to answer some questionnaires about yourself over the phone. This will take about 1.5 hours in total. You will be given a specific time for the telephone calls that are convenient for you. In the first interview, the researcher or research assistant will ask you some questions about yourself. Sofia will SMS you the retail voucher and you will then go to the shop in your own time during the following week. You have up to 7 days to complete your shopping trip. You will also receive R100 for your transport and data costs via electronic payment. We want you to feel that you can buy things that you want, just like when you normally do your grocery shopping. Shopping will ideally take place on

1

Sofia WhatsApp: +49 152 5542 7686

your own, not with friends or family members. No-one except the researchers will be able to see what you have bought. After you have completed the shopping activity we will ask you some questions about what you usually eat and what you thought about the shopping activity on the phone. We really want to hear your honest answers, even if you think they are negative or not what we want to hear.

How do I take part?

If you would like to take part in the study, you can contact one of us via phone, email, or SMS (see details above). We will phone you back to check that you are eligible and answer any questions you have. If you still want to take part and we decide that you are eligible, we will give you a specific time period in which to complete the shopping activity on your own and do the two telephone interviews with a researcher for about 30 minutes each. Sofia will SMS you the shopping voucher after you are enrolled in the study (after informed consent).

What will be done with the information collected?

The information from this study will be used to improve the way that researchers measure food choices and it will also help EBSA improve their program for women like you in the future. The results (but not your names) may also be published in research and magazine articles and may be presented at public and scientific talks.

Will my information be kept confidential?

You will be given a code that only the study investigators will know. This code will be used instead of your name for all the information we collect from the shop and feedback form. Additionally, EBSA will know who took part in the shopping experience, but we will not tell them who purchased what in the shop or who said what in the feedback form or in the questionnaire about yourself.

What are the benefits for taking part?

You will help to improve the way that researchers measure diet and also help EBSA to improve their program for women like you. We will share our study conclusions with you. You will take home up to R250 in groceries that you purchase. You will also be given R100 to reimburse you for your transport and air time.

What are the risks for taking part?

The risks for taking part in this study are no greater than you would currently experience when shopping for groceries at [shop details]. However the world is currently experiencing an outbreak of Covid-19 which was declared a pandemic in March 2020. The virus responsible of Covid-19 is highly infectious and is transmitted via person to person contact, via respiratory droplets and via surfaces that have been touched by infected persons. Covid-19 can have severe health consequences and can even cause death in certain individuals. Health compromised individuals (eg those with diabetes or cardiovascular disease) are at higher risk for adverse Covid-19 outcomes. South Africa, and particularly the Western Cape are experiencing an increase in new cases and deaths from Covid-19. Therefore by travelling to shops and purchasing groceries you will be exposed to greater risk of Covid-19 than if you stayed at home. We do not want to increase your risk of becoming infected with the virus so will only include you in the study if you would be going shopping anyway. Therefore, you will only be eligible for this study if you have routinely been going shopping in your community during the past 2 months.

Please follow the up-to-date Covid-19 guidelines that the government has provided. At the moment these include washing your hands often, wearing a face mask in public and maintaining social distancing of 1.5m where possible. Please follow the supermarkets recommended precautions such as using hand sanitizer. Please contact the researchers if they have any concerns or negative experiences. Please contact your doctor if you fall ill. Do not attend a supermarket study visit if you have experienced shortness of breath, dry cough, fever or high temperature in the previous 2 weeks. We will only contact you via telephone and Whatsapp to limit the need for person to person contact in the study.

You will make your choices in the shop privately and no one outside the research team will know what you buy. You will be encouraged to give us your honest feedback about your experience buying groceries in the pop-up shop. For some people, these topics may be sensitive and may cause negative emotions. Please keep in mind that you will not be judged for anything that you choose to buy or say in your feedback form. Supermarkets take measures to prevent food going off. Please let us know if you are allergic to any foods.

Ethical considerations

This study has been approved by the faculty of Health Sciences, Human Research Ethics Committee (UCT). You will only start this study once you have confirmed that you understand what the study is about by signing the informed consent form on the day of visit. Participation in the study is entirely voluntary and you have the right to withdraw from the study at any time. The investigator may also withdraw you from the study at any time should he/she believe it is in your best interests to do so, or you are found to be ineligible for the study.

What if Something Goes Wrong?

The University of Cape Town (UCT) undertakes that in the event of you suffering any significant deterioration in health or well-being, or from any unexpected sensitivity or toxicity, that is caused by your participation in the study, it will provide immediate medical care. UCT has appropriate insurance cover to provide prompt payment of compensation for any trial-related injury according to the guidelines outlined by the Association of the British Pharmaceutical Industry, ABPI 1991. Broadly-speaking, the ABPI guidelines recommend that the insured company (UCT), without legal commitment, should compensate you without you having to prove that UCT is at fault. An injury is considered trial-related if, and to the extent that, it is caused by study activities. You must notify the study doctor immediately

Sofia WhatsApp: +49 152 5542 7686

of any side effects and/or injuries during the trial, whether they are research-related or other related complications.

UCT reserves the right not to provide compensation if, and to the extent that, your injury came about because you chose not to follow the instructions that you were given while you were taking part in the study. Your right in law to claim compensation for injury where you prove negligence is not affected. Copies of these guidelines are available on request.

Please feel free to contact the UCT Human Research Ethics Committee if you have any queries: Floor E53-Room 46, Old Main Building, Groote Schuur Hospital, Observatory, 7925 Phone: (021) 406 6338 **Email:** hrec-enquiries@uct.ac.za

B. Participant information sheet EBSA treatment group

Appendix H, HREC REF 295/2019, Version 3, 3 June 2020

STUDY INFORMATION SHEET - Group B Participants

Behavioural and self-report measures of food choice in women who have taken part in a nutrition education program

Dear potential study participant,

You have been contacted because you are a woman who took part in an Eat Better South Africa (EBSA) program in your community. We are researchers from the University of Cape Town (UCT) and the Max Planck Institute for Research on Collective Goods (MPI) and we would like you to take part in our study on measuring food choices. Below you will find more information about what is involved. Please feel free to contact us if you have any questions or would like to take part. You can also contact your coach if you would like to take part.

Georgina Pujol-Busquets Guillén (UCT)	georginapbg@gmail.com	061 659 2692
Sofia Monteiro (MPI)	monteiro@coll.mpg.de	066 311 8981
James Smith (UCT)	ja.smith@uct.ac.za	083 3057593
Kate Larmuth (UCT)	kateus65@gmail.com	074 217 0042

What is the study about?

This study will use two different methods to measure the food choices of people to better understand nutrition and shopping habits in communities like yours.

Are you eligible to take part?

You are allowed to participate if you are a woman; are 18-69 years old; can understand and speak English or Afrikaans; do not have private healthcare or medical insurance and are able to visit a local Pick n Pay supermarket and do a telephone survey interview. You must routinely shop for your household in order to participate.

What will be required of you?

If you want to take part, you will be given a R250 shopping voucher to spend on foods and non-alcoholic drinks at Pick n Pay. You can keep the groceries you buy but we ask that you send us one photo of everything you bought, and one photo of the till slip. You will also be asked to answer some questionnaires about yourself over the phone which will take about 1.5 hours in total. You will be given a specific time for the telephone calls that are convenient for you. A researcher or research assistant will ask you some questions about yourself. Sofia will SMS you the retail voucher and you will then go to the shop in your own time during the following week. You will also receive R100 for your transport and data costs via electronic payment. We want you to feel that you can buy things that you want, just like when you normally do your grocery shopping. Shopping will ideally take place on your own, not with friends or family members. No-one except the

1

Sofia WhatsApp: +49 152 5542 7686

researchers will be able to see what you have bought. After you have completed the shopping activity we will ask you some questions about what you usually eat and what you thought about the shopping activity. We really want to hear your honest answers, even if you think they are negative or not what we want to hear.

How do I take part?

If you would like to take part in the study, you can contact one of us via phone, email, or SMS (see details above). We will phone you back to check that you are eligible and answer any questions you have. If you still want to take part and we decide that you are eligible, we will give you a specific time period in which to complete the shopping activity on your own and do the two telephone interviews with a researcher about 30 minutes each. Sofia will SMS you the shopping voucher after you are enrolled in the study. The infographic we will send you contains everything you need to know in 10 easy steps.

What will be done with the information collected?

The information from this study will be used to improve the way that researchers measure food choices and it will also help EBSA improve their program for women like you in the future. The results (but not your names) may also be published in research and magazine articles and may be presented at public and scientific talks.

Will my information be kept confidential?

You will be given a code that only the study investigators will know. This code will be used instead of your name for all the information we collect from the shop and feedback form. Additionally, EBSA will know who took part in the shopping experience, but we will not tell them who purchased what in the shop or who said what in the feedback form or in the questionnaire about yourself.

What are the benefits for taking part?

You will help to improve the way that researchers measure diet and also help EBSA to improve their program for women like you. We will share our study conclusions with you. You will take home up to R250 in groceries that you purchase. You will also be given R100 to reimburse you for your transport and air time.

What are the risks for taking part?

The risks for taking part in this study are no greater than you would currently experience when shopping for groceries at Pick n Pay. However the world is currently experiencing an outbreak of Covid-19 which was declared a pandemic in March 2020. The virus responsible of Covid-19 is highly infectious and is transmitted via person to person contact, via respiratory droplets and via surfaces that have been touched by infected persons. Covid-19 can have severe health consequences and can even cause death in certain individuals. Health compromised individuals (eg those with diabetes or cardiovascular disease) are at higher risk for adverse Covid-19 outcomes. South Africa, and particularly the Western Cape are experiencing an increase in new cases and deaths from Covid-19. Therefore by travelling to shops and purchasing groceries you will be exposed to greater risk of Covid-19 than if you stayed at home. We do not want to increase your risk of becoming infected with the virus so will only include you in the study if you would be going shopping anyway. Therefore, you will only be eligible for this study if you have routinely been going shopping in your area during the past 2 months.

Please follow the up-to-date Covid-19 guidelines that the government has provided. At the moment these include washing your hands often, wearing a face mask in public and maintaining social distancing of 1.5m where possible. Please follow the supermarkets recommended precautions such as using hand sanitizer. Please contact the researchers if they have any concerns or negative experiences. Please contact your doctor if you fall ill. Do not attend a supermarket study visit if you have experienced shortness of breath, dry cough, fever or high temperature in the previous 2 weeks. We will only contact you via telephone and Whatsapp to limit the need for person to person contact in the study.

You will make your choices in the shop privately and no one outside the research team will know what you buy. You will be encouraged to give us your honest feedback about your experience buying groceries in the pop-up shop. For some people, these topics may be sensitive and may cause negative emotions. Please keep in mind that you will not be judged for anything that you choose to buy or say in your feedback form. Supermarkets take measures to prevent food going off. Please let us know if you are allergic to any foods.

Ethical considerations

This study has been approved by the faculty of Health Sciences, Human Research Ethics Committee (UCT). You will only start this study once you have confirmed that you understand what the study is about by signing the informed consent form on the day of visit. Participation in the study is entirely voluntary and you have the right to withdraw from the study at any time. The investigator may also withdraw you from the study at any time should he/she believe it is in your best interests to do so, or you are found to be ineligible for the study.

What if Something Goes Wrong?

The University of Cape Town (UCT) undertakes that in the event of you suffering any significant deterioration in health or well-being, or from any unexpected sensitivity or toxicity, that is caused by your participation in the study, it will provide immediate medical care. UCT has appropriate insurance cover to provide prompt payment of compensation for any trial-related injury according to the guidelines outlined by the Association of the British Pharmaceutical Industry, ABPI 1991. Broadly-speaking, the ABPI guidelines recommend that the insured company (UCT), without legal commitment, should compensate you without you having to prove that UCT is at fault. An injury is considered trial-related if, and to the extent that, it is caused by study activities. You must notify the study doctor immediately of any side effects and/or injuries during the trial, whether they are research-related or other related complications.

UCT reserves the right not to provide compensation if, and to the extent that, your injury came about because you chose not to follow the instructions that you were given while you were taking part in the study. Your right in law to claim compensation for injury where you prove negligence is not affected. Copies of these guidelines are available on request.

Please feel free to contact the UCT Human Research Ethics Committee if you have any queries: Floor E53-Room 46, Old Main Building, Groote Schuur Hospital, Observatory, 7925 Phone: (021) 406 6338 **Email:** hrec-enquiries@uct.ac.za

C. Infographic for participants: 10 steps to complete the study

EVERYTHING YOU NEED TO KNOW
IN 10 EASY STEPS ☺

BEFORE THE STUDY (5 MINUTES):

STEP 1: SCREENING QUESTIONS

PHONE INTERVIEW 1 (30 MINUTES):

STEP 2: INFORMED CONSENT
STEP 3: DEMOGRAPHIC QUESTIONNAIRE

SHOPPING ACTIVITY (20 MINUTES):

STEP 4: RECEIVE SMS VOUCHER
STEP 5: CHOOSE GROCERIES AT PNP
STEP 6: PAY WITH VOUCHER BARCODE
STEP 7: PHOTOGRAPH YOUR GROCERIES

PHONE INTERVIEW. 2 (30 MINUTES):

STEP 8: FEEDBACK FORM
STEP 9: FOOD QUESTIONNAIRE

AFTER THE STUDY:

STEP 10: SHARE RESEARCH FINDINGS.

STEP 1: ELIGIBILITY & SCREENING

WOMEN WHO ARE ELIGIBLE ARE

18-69 YEARS OLD

ENGLISH OR AFRIKAANS SPEAKING

COMFORTABLE TO SHOP
AT PICK N PAY

HAVE ACCESS TO A SMART PHONE
WITH WHATSAPP AND CAMERA

ONLY 60 WOMEN FROM YOUR COMMUNITY
WILL BE ABLE TO PARTICIPATE.

YOU WILL RECEIVE A STUDY INFO SHEET
VIA WHATSAPP OR EMAIL.

IF YOU ARE INTERESTED,
WE WILL ASK YOU A FEW QUESTIONS
TO CONFIRM YOU ARE ELIGIBLE.

YOU WILL RECEIVE A LINK IN WHATSAPP
TO A QUICK ONLINE SURVEY.

A RESEARCHER WILL HELP YOU IF YOU NEED IT.

IF WE CONFIRM YOU ARE ELIGIBLE THEN WE
SCHEDULE YOUR FIRST STUDY PHONE CALL.

STEP 2: INFORMED CONSENT & CONFIDENTIALITY

A RESEARCHER WILL CALL YOU AT YOUR
SCHEDULED DATE AND TIME.
SHE WILL EXPLAIN THE STUDY TO YOU.

YOU WILL RECEIVE A LINK VIA WHATSAPP TO AN
ONLINE CONSENT FORM.
CLICK ON THE LINK TO OPEN THE FORM.

A RESEARCHER WILL HELP YOU IF YOU NEED IT.

WE WILL USE A PARTICIPANT CODE
INSTEAD OF YOUR NAME.

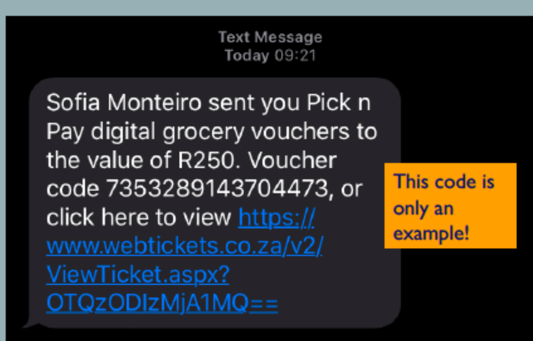
NO ONE OUTSIDE THE RESEARCH TEAM WILL
KNOW WHAT YOU PERSONALLY BUY
IN THE SHOPPING ACTIVITY
OR SAY IN THE QUESTIONNAIRES.

YOUR PRIVACY IS OUR PRIORITY.

STEP 3: DEMOGRAPHIC QUESTIONNAIRE

ON THE FIRST STUDY PHONE CALL
A RESEARCHER WILL ASK YOU SOME QUESTIONS
ABOUT YOURSELF AND YOUR HOUSEHOLD.

STEP 4: RECEIVE YOUR SMS VOUCHER



STEP 5: CHOOSE GROCERIES AT PICK 'N PAY SUPERMARKET

YOU HAVE ONE WEEK TO SPEND YOUR R250
VOUCHER AT YOUR LOCAL PNP

BUY FOOD AND
(NON-ALCOHOLIC) DRINKS ONLY

SIMPLY BUY WHAT YOU NORMALLY CHOOSE AT
THE SUPERMARKET ☺

STEP 6: PAY WITH YOUR VOUCHER

This barcode is only an example!

CLICK ON THE SMS LINK (SEE STEP 4)
TO GET YOUR BARCODE.

SHOW BARCODE TO THE PERSON AT CHECKOUT.

STEP 7: PHOTOGRAPH YOUR GROCERIES AND YOUR RECEIPT

WHEN YOU GET HOME,
TAKE A PHOTO OF YOUR GROCERIES

TAKE A PHOTO OF YOUR RECEIPT.

SEND SOFIA THE PHOTOS TO CONFIRM YOU HAVE DONE YOUR STUDY VISIT TO PNP.

STEP 8: FEEDBACK FORM

WHEN WE RECEIVE YOUR PHOTOS WE WILL SCHEDULE YOUR SECOND PHONE CALL.

A RESEARCHER WILL CONFIRM WHAT YOU BOUGHT.

SHE WILL ASK YOU SOME QUESTIONS ABOUT YOUR SHOPPING EXPERIENCE.

STEP 9: FOOD FREQUENCY QUESTIONNAIRE

ON YOUR SECOND PHONE CALL
THE RESEARCHER WILL ASK YOU SOME QUESTIONS ABOUT THE DIFFERENT FOODS YOU NORMALLY EAT.

STEP 10: SHARE RESEARCH FINDINGS

AFTER THE STUDY FINISHES WE WILL SHARE OUR RESEARCH FINDINGS WITH YOU.

D. Interview 1 Transcript for Researcher/Research Assistant

TRANSCRIPT FOR INTERVIEWERS

DEAR RESEARCH ASSISTANT, YOUR INSTRUCTIONS TO GUIDE YOU ARE WRITTEN IN **PURPLE.**

DO NOT READ PURPLE OUT LOUD. SAY EVERYTHING WRITTEN IN BLACK - YOU CAN READ OR PARAPHRASE IN YOUR OWN WORDS AS YOU GET MORE COMFORTABLE. TRY TO KEEP TO THE TIME ALLOCATION.

Make sure you know your participant's code!! (CHECKLIST FOR INTERVIEWERS)

Call the participant at {date, time} that you booked with them. Remember to invite Sofia in the calendar event you created and turn the reminder alert on.

INTRO AND SMALL TALK (2 minutes)

Be warm but professional :)

Hello, it's {your name}, Sofia's research assistant from the diet and shopping habits study.

How are you doing? How is the weather? …. Are you comfortable in English? Sal ons Afrikaans praat?

Ok, shall we get started with the interview?

INFORMED CONSENT (10 minutes)

To get informed consent you must (1) explain the study, (2) send the participant the informed consent survey link via whatsapp, (3) end the call for a minute while they complete it, then (4) call them back straight to continue the interview.

(1) EXPLAIN THE STUDY

Ek help die navorsers van die Universiteit van Kaapstad en Max Planck Institute for Research on Collective Goods en ons wil graag hê dat u moet deelneem aan ons studie oor voedselkeuses. Hier is meer inligting oor wat die studie behels.

(English): I help the researchers from UCT and MPI and we are happy that you want to participate in our study about food choices. Here is more information about important aspects of the study.

1

Sofia Monteiro: WhatsApp: +49 152 5542 7686

(please read these question headings too)

Waaroor gaan die studie?

Die studie sal van twee verskillende metodes gebruik maak om mense se voedselkeuses te meet om voedings- en inkopiegewoontes in gemeenskappe soos u eie beter te verstaan.

> (English): The study will use two different methods to better understand food choices and shopping habits in communities like yours.

Wat sal van u verwag word?

Indien u wil deelneem, sal u 'n koopbewys van R250 kry om by Pick n Pay te spandeer op voedsel en nie-alkoholiese drankies. U kan die kruideniersware wat u koop hou, maar ons vra dat u vir Sofia 'n foto stuur van alles wat u gekoop het, sowel as die strokie. U sal ook gevra word om 'n paar vraelyste oor uself per telefoon te beantwoord wat ongeveer 1.5 ure in totaal sal neem. 'n Spesifieke tyd wat vir u gerieflik is sal vir die telefoonoproep gegee word. Vandag sal ek aan u 'n paar vrae oor uself vra. Sofia sal die Pick n Pay koopbewys per SMS stuur en u sal dan in die volgende week in u eie tyd Pick n Pay besoek. U sal ook R100 via elektroniese betaling ontvang vir u vervoer- en datakoste. Ons wil hê u moet met vrymoedigheid koop wat u wil, net soos wanneer u normaalweg inkopies doen. U sal ideaal alleen wees wanneer u inkopies doen, nie met vriende of familielede nie. Niemand behalwe die navorsers sal sien wat u koop nie. Nadat u die inkopie-aktiwiteit voltooi het, sal ek aan u 'n paar vrae vra oor wat u gewoonlik eet en wat u van ons inkopie-aktiwiteit dink. Ons wil graag hê dat u antwoorde eerlik moet wees, selfs as u dink dat hulle negatief is of nie is wat ons wil hoor nie.

> (English): If you want to take part, you will be given a R250 shopping voucher to spend on foods and non-alcoholic drinks at Pick n Pay. You can keep all the groceries you buy. We ask that you send Sofia one photo of everything you bought, and one photo of the till slip. You will also be asked some questionnaires about yourself over the phone which will take about 1.5 hours in total between our two interviews. One now and one later this week. You will be given a specific time for the telephone calls that are convenient for you. A research assistant will ask you some questions about yourself. Sofia will SMS you the R250 Pick n Pay voucher and you will then go to Pick n Pay in your own time during the week. You will also receive R100 for your transport and data costs via electronic payment. We want you to feel that you can buy things that you want, just like when you normally do your grocery shopping. No-one except the researchers will be able to see what you have bought. After you have completed the shopping activity we will ask you some questions about what you usually eat and what you thought about the shopping activity. We really want to hear your

Sofia Monteiro: WhatsApp: +49 152 5542 7686

honest answers, even if you think they are negative or not what we want to hear.

Wat sal met die inligting wat ons versamel gebeur?

Die inligting uit hierdie studie sal gebruik word om die manier waarop navorsers voedselkeuses meet, te verbeter. Dit sal ook Eat Better South Africa (EBSA) help om hul program in die toekoms te verbeter. Die resultate (maar nie u naam nie) sal moontlik ook in navorsings- en tydskrifartikels gepubliseer word en kan tydens openbare en wetenskaplike besprekings aangebied word.

(English): The information from this study will be used to improve the way that researchers measure food choices and it will also help EBSA improve their program for women like you in the future. The results (but not your name) may also be published in research and magazine articles and may be presented at public and scientific talks.

Sal my inligting vertroulik gehou word?

U sal 'n kode ontvang en slegs u, die studienavorsers en ek sal weet wat hierdie kode is. Hierdie kode sal in plaas van u naam gebruik word vir al die inligting wat ons van die winkel en terugvoervorm versamel. Daarbenewens sal EBSA weet wie aan die inkopies-ervaring deelgeneem het, maar ons sal nie vir hulle laat weet wie wat in die winkel gekoop het of wat u in die terugvoervorm of vraelys oor uself gesê het nie.

(English): You will be given a code that only the study investigators will know. This code will be used instead of your name for all the information we collect from the shop and feedback form. Additionally, EBSA will know who took part in the shopping experience, but we will not tell them who purchased what in the shop or who said what in the feedback form or in the questionnaire about yourself.

Wat is die voordele vir deelname?

U sal help om die manier waarop navorsers dieet meet, te verbeter en u sal ook EBSA help om hul program vir vroue te verbeter. Ons sal die gevolgtrekkings van die studie met u deel. U mag kruideniersware ter waarde van R250 wat u gekoop het, huis toe neem. U sal ook R100 kontant ontvang om u te vergoed vir u vervoer en lugtyd .

(English): You will help to improve the way that researchers measure diet and also help EBSA to improve their program for other women. We will share our study conclusions with you. You will take home R250 in groceries. You will also be given R100 to reimburse you for your transport and air time.

Wat is die risiko's vir deelname?

Sofia Monteiro: WhatsApp: +49 152 5542 7686

Die risiko's vir deelname is nie groter as wat u tans sal ervaar wanneer u koop by Pick n Pay nie.

Die wêreld ervaar egter tans 'n uitbraak van Covid-19. Ons wil nie u risiko vir infeksie verhoog nie, en sal u slegs by die studie insluit indien u in elk geval inkopies doen. U sal dus slegs vir deelname aan die studie kwalifiseer indien u oor die afgelope 2 maande gereeld in jou gemeenskap gaan inkopies doen het.

Volg asb. die bygewerkte Covid-19-riglyne wat die regering opgestel het. Kontak asb. die navorsers as hulle enige kommer of negatiewe ondervindinge ervaar. Kontak asb. u dokter as u siek word. Moet asb. nie aan 'n supermarkstudiebesoek deelneem indien u in die afgelope 2 weke kortasem, droë hoes of hoë koors ervaar het nie. Ons sal u slegs per telefoon en Whatsapp kontak om persoon tot persoon kontak tydens die studie sover moontlik te beperk.

(English): The risks for taking part in this study are no greater than you would currently experience when shopping at Pick n Pay.

However the world is currently experiencing an outbreak of Covid-19. We do not want to increase your risk of becoming infected with the virus so will only include you in the study if you would be going shopping anyway.

Please follow the up-to-date Covid-19 guidelines that the government has provided. Please contact the researchers if they have any concerns or negative experiences. Please contact your doctor if you fall ill. Do not attend a supermarket study visit if you have experienced shortness of breath, dry cough, fever or high temperature in the previous 2 weeks. We will only contact you via telephone and WhatsApp to limit the need for person to person contact in the study.

---

(2) COPY THIS LINK (ONLINE INFORMED CONSENT FORM) AND SEND IT TO THE PARTICIPANT IN WhatsApp now.

Goed, ek het u die toestemmingsvorm in WhatsApp gestuur. Oor 'n minuut sal ek die oproep vir 'n paar minute beëindig en u laat voltooi. Dan bel ek u dadelik terug om met ons onderhoud voort te gaan.

Om by die studie ingeskryf te word, moet u hierdie toestemmingsvorm voltooi deur die volgende te doen:

As u klik op 'Ja', sal u aandui dat u bewus is van hierdie belangrike aspekte van die studie).

Sofia Monteiro: WhatsApp: +49 152 5542 7686

As u 'Ja' antwoord op al die stellings en steeds graag wil deelneem, klik dan op 'Ek stem in, begin die studie' en tik dan u naam in.

Klik uiteindelik op OK. Dit beteken dat u ons toestemming gegee het om u vir die studie in te skryf.

As u 'Nee' antwoord op een van die stellings, maar steeds wil deelneem, kontak Sofia vir meer inligting voordat u die vorm invul.

As u van plan verander het en nie wil deelneem nie, klik aan die einde van hierdie vorm op "Ek stem nie in nie, ek wil nie deelneem nie".

U kan te eniger tyd ek om hulp vra om die opname oop te maak.

Vertel my asseblief wanneer u dit voltooi het.

Is dit reg so? [Wait for a question from participant]

Ok, totsiens vir nou!

(English): I have just sent you the consent form in our WhatsApp chat.

In a minute, I will end the call for a few minutes and let you complete the consent form. Then I will call you back straight away to continue our interview.

Clicking '**Yes**' to the 6 statements will indicate you are aware and accept these important aspects about the study that I just explained to you)
If you answer '**Yes**' to all statements and are still happy to take part, click 'I consent, begin the study' and then type your name.
Finally, click ' OK '. Doing this means that you have given us permission to enrol you in the study.

If you answer '**No**' to any of the statements but would still like to take part, please contact Sofia for more information before completing this form. We will reschedule your interview for a later time. If you have changed your mind and do not want to participate, click "I do not consent, I do not wish to participate" at the end of this form.

You can ask me for help at any time. Do you have any questions now?

Ok, please tell me when you are done and I will call you back.

Bye for now!

(3) END THE CALL WHILE PARTICIPANT COMPLETES INFORMED CONSENT FORM ONLINE (1 minute)

[wait to receive text from the participant - "Ek het die opname voltooi". Problem? Try to help them open it "did you click on the blue link?".]

(4) CALL PARTICIPANT BACK TO CONTINUE THE INTERVIEW.

(Sofia will check that they filled out the form properly and follow up if they did not. Just continue.)

BASELINE SURVEY (15 MINUTES)

You will fill out this form for the participant while you are on the phone now.

Do you know the PARTICIPANT CODE? Check what this is in the CHECKLIST FOR INTERVIEWERS.

Click on the blue link here to open --> BASELINE SURVEY.

Great! Well done! You completed Interview 1.

Now return to --> CHECKLIST FOR INTERVIEWERS and fill in how long the call took, and write any comments you want to make note of.

Sofia Monteiro: WhatsApp: +49 152 5542 7686

E. Baseline Questionnaire

Survey Flow

Block: Introduction (7 Questions)
Standard: Demographics (13 Questions)
Standard: Tobacco Use (1 Question)
Standard: Alcohol Consumption (2 Questions)
Standard: History of Raised Blood Pressure (4 Questions)
Standard: History of Raised Blood Sugar or Diabetes (5 Questions)
Standard: History of Cardiovascular Diseases (3 Questions)
Standard: Household Food Insecurity Access Scale (20 Questions)
Standard: Shopping/Cooking Behaviour (9 Questions)

Page Break

Start of Block: Introduction

[Instructions for Interviewers are in PURPLE. Do not read purple instructions to the participant.]

Goed, laat ons begin!
Ek sal u vrae stel oor uself en u huishouding.
Die opname duur ongeveer **15 minute**.

U antwoorde is sterk vertroulik. Slegs die studie-ondersoekers sal u kan identifiseer. As ek enige vrae vra wat u nie wil beantwoord nie, laat weet my dan; dan gaan ek na die volgende vraag. U kan ook die onderhoud te eniger tyd stop. Ons hoop egter dat u aan die opname sal deelneem omdat u antwoorde vir ons belangrik is.

Most of the questions are in English. If you want me to clarify, I can do that in Afrikaans.

____

(English): I will ask you some questions about yourself and your household.

The survey will take about **15 minutes.**

Your answers are strictly confidential. Only the study investigators will be able to identify you. If I ask any questions you do not want to answer, please let me know; then I will move on to the next question. You can also stop the interview at any time. However, we hope that you will participate in the survey because your answers are important to us.

---

participant_code **Enter Participant Code** (see CHECKLIST for Interviewers if you forget what it is):

____

interviewer Interviewer, please select yourself:____________________

consent_given Has participant completed the consent form? Check yes if you have done this step.

○ Yes

○ No

mobile_number What is your cellphone number? (You can add this since you have their number)

____

first_language What is your first or home language?

▢ English

▢ Afrikaans

▢ Other ____

other_languages What other languages can you understand and speak?

▢ None

▢ English

▢ Afrikaans

▢ Other ____

End of Block: Introduction

Start of Block: Demographics

age_years How old are you? (years)

____

height_cm What is your height? (centimetres)

____

weight_kg What is your weight? (kg)

____

cultural_heritage In your own words, please can you describe your racial, ethnic or cultural background. We are interested in this because it can have an impact on food traditions and choices. [Probe: If they are confused, add: So for example, some people may say: Cape Coloured, White European, Black South African, etc]

____

partner_shares_hh Do you share a household with a partner?

○ Yes

○ No

○ Prefer not to answer

parent Do you have children?

○ Yes

○ No

Skip To: highest_education If Do you have children? = No

number_children How many children do you have?

number_children_hh How many children and grandchildren currently live with you in your household?

____

highest_education What is the highest level of education you have completed?

- No formal schooling
- Less than primary school
- Primary school completed
- High school completed (Matric)
- College/University completed
- Post graduate degree
- Prefer not to answer

employed Are you currently employed?

- Yes
- No
- Prefer not to answer

Display This Question:

If Are you currently employed? = No

Covid_unemployed Do you currently have less employment or no job because of Covid-19 or lockdown?

- Yes
- No

number_hh_members How many people in total, including yourself, live in your household? [Enter number]

____

employed_hh_members How many people, including yourself, in your household are employed? [Enter number]

____

End of Block: Demographics

Start of Block: Tobacco Use

smoker Do you currently smoke any tobacco products every day? [If needed prompt, cigarettes, cigars or pipes?]

- Yes
- No
- Prefer not to answer

End of Block: Tobacco Use

Start of Block: Alcohol Consumption

alcohol_past_year Have you consumed any **alcohol** within the past 12 months?

- Yes
- No

Skip To: End of Block If Have you consumed any alcohol within the past 12 months? = No

alcohol_freq During the past 12 months, **how frequently** have you had at least one alcoholic drink?

- Daily
- 5-6 days per week
- 3-4 times a week
- 1-2 days per week
- 1-3 days per month
- Less than once a month
- Never

End of Block: Alcohol Consumption

Start of Block: History of Raised Blood Pressure

high_BP Have you ever been told **you have high blood pressure** by a doctor or health care worker?

○ Yes

○ No

Skip To: End of Block If Have you ever been told you have high blood pressure by a doctor or health care worker? = No

BP_diagnosis_12mo Have you been told you have high blood pressure in the past 12 months?

○ Yes

○ No

Display This Question:

If Have you been told you have high blood pressure in the past 12 months? = Yes

BP_past_6mo Have you been told in the past 6 months?

○ Yes

○ No

BP_meds In the past two weeks, have you taken any medication for high blood pressure prescribed by a physician?

○ Yes

○ No

End of Block: History of Raised Blood Pressure

Start of Block: History of Raised Blood Sugar or Diabetes

diabetes_diagnosed Have you been told **you have diabetes (that is high blood sugar)** by a doctor or health care worker?

Yes

No

Skip To: End of Block If Have you been told you have diabetes (that is high blood sugar) by a doctor or health care worker? = NoDisplay This Question:

If Have you been told you have diabetes (that is high blood sugar) by a doctor or health care worker? = Yes

diabetes_past_12mo Have you been told in the past 12 months?

Yes

No

Display This Question:

If Have you been told in the past 12 months? = Yes

diabetes_past_6mo Have you been told in the past 6 months?

Yes

No

Display This Question:

If Have you been told you have diabetes (that is high blood sugar) by a doctor or health care worker? = Yes

diabetes_meds In the past two weeks, have you taken any medication for diabetes?

Yes

No

Display This Question:

If Have you been told you have diabetes (that is high blood sugar) by a doctor or health care worker? = Yes

insulin Are you currently using insulin for diabetes?

Yes

No

End of Block: History of Raised Blood Sugar or Diabetes

Start of Block: History of Cardiovascular Diseases

CVD_event Have you had a heart attack or chest pain from heart disease or a stroke?

○ Yes

○ No

asprin Are you currently taking Disprin regularly to prevent or treat heart disease? [aspirin]

○ Yes

○ No

statins Are you currently taking statins, which is medicine for high cholesterol?

○ Yes

○ No

End of Block: History of Cardiovascular Diseases

Start of Block: Household Food Insecurity Access Scale

Now I'd like to ask you ten YES or NO questions about food security (Voedselsekuriteit) in your household.

If you answer YES, I will ask you how frequently. Rarely, Somtimes or Often? (Selde, soms of gereeld?)

[Remember this is a sensitive topic for some families. Keep this in mind as you ask the following questions. Afrikaans is in brackets if needed]

V268 In the past four weeks, did you worry that your household would not have enough food? *(Was u die afgelope vier weke bekommerd dat u huishouding nie genoeg voedsel sou hê nie?)*

○ Yes

○ No

Skip To: V270 If In the past four weeks, did you worry that your household would not have enough food? (Was u di... = No

V269 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

- ○ Rarely (once or twice in the past four weeks)
- ○ Sometimes (3 to 10 times in the past four weeks)
- ○ Often (more than 10 times in the past four weeks)

V270 In the past four weeks, were you or any household member not able to eat the kinds of foods you preferred because there was not enough money?

(Kon u of enige huishoudelike lidmaatjie die afgelope vier weke nie die soorte kosse eet wat u verkies nie omdat daar nie genoeg geld was nie?)

- ○ Yes
- ○ No

Skip To: no_luxury_foods If In the past four weeks, were you or any household member not able to eat the kinds of foods you p... = No

V271 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct e.g. "Rarely, so once or twice...? Is that right?"].

- ○ Rarely (once or twice in the past four weeks)
- ○ Sometimes (three to ten times in the past four weeks)
- ○ Often (more than ten times in the past four weeks)

no_luxury_foods In the past 4 weeks were you or any household member not able to eat any luxury foods you wanted to eat because there was not enough money?

(Kon u of 'n huishoudelike lidmaatjie die afgelope vier weke nie luukse kos eet wat u wou eet nie omdat daar nie genoeg geld was nie?)

- ○ Yes
- ○ No

V272 In the past four weeks, did you or any household member have to eat a limited variety of foods due to a lack of money?

(Het u of enige huishoudelike lidmaatjie die afgelope vier weke 'n beperkte verskeidenheid kosse moes eet as gevolg van 'n tekort aan geld?)

◯ Yes

◯ No

Skip To: V274 If In the past four weeks, did you or any household member have to eat a limited variety of foods du... = No

V273 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

◯ Rarely (once or twice in the past four weeks)

◯ Sometimes (three to ten times in the past four weeks)

◯ Often (more than ten times in the past four weeks)

V274 In the past four weeks, did you or any household member have to eat some foods that you really did not want to eat because there was not enough money?

(Het u of enige huishoudelike lid die afgelope vier weke 'n paar kosse geëet wat u regtig nie wou eet nie omdat daar nie genoeg geld was nie?)

◯ Yes

◯ No

Skip To: V276 If In the past four weeks, did you or any household member have to eat some foods that you really di... = No

V275 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

◯ Rarely (once or twice in the past four weeks)

◯ Sometimes (three to ten times in the past four weeks)

◯ Often (more than ten times in the past four weeks)

V276 In the past four weeks, did you or any household member have to eat a smaller meal than you felt you needed because there was not enough money for food?

(Moet u of enige huishoudelike lidmaatjie die afgelope vier weke 'n kleiner maaltyd eet as wat u gevoel het dat u nodig het omdat daar nie genoeg geld vir kos was nie?)

○ Yes

○ No

Skip To: V278 If In the past four weeks, did you or any household member have to eat a smaller meal than you felt... = No

V277 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

○ Rarely (once or twice in the past four weeks)

○ Sometimes (three to ten times in the past four weeks)

○ Often (more than ten times in the past four weeks)

V278 In the past four weeks, did you or any other household member have to eat fewer meals in a day because there was not enough money for food?

(Het u of enige ander huishoudelike lidmaat in die afgelope vier weke minder maaltye op 'n dag moes eet omdat daar nie genoeg geld vir kos was nie?)

○ Yes

○ No

Skip To: V280 If In the past four weeks, did you or any other household member have to eat fewer meals in a day be... = No

V279 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

○ Rarely (once or twice in the past four weeks)

○ Sometimes (three to ten times in the past four weeks)

○ Often (more than ten times in the past four weeks)

V280 In the past four weeks, was there ever no food to eat of any kind in your household because of lack of money to buy food?

(Was daar in die afgelope vier weke ooit voedsel in enige huishouding om te eet as gevolg van 'n gebrek aan geld om kos te koop?)

- Yes
- No

Skip To: V282 If In the past four weeks, was there ever no food to eat of any kind in your household because of la... = No

V281 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

- Rarely (once or twice in the past four weeks)
- Sometimes (three to ten times in the past four weeks)
- Often (more than ten times in the past four weeks)

V282 In the past four weeks, did you or any household member go to sleep at night hungry because there was not enough money to buy food?

(Het u of enige huishoudelike lidmaat die afgelope vier weke snags honger geslaap omdat daar nie genoeg geld was om kos te koop nie?)

- Yes
- No

Skip To: V284 If In the past four weeks, did you or any household member go to sleep at night hungry because there... = No

V283 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

- Rarely (once or twice in the past four weeks)
- Sometimes (three to ten times in the past four weeks)
- Often (more than ten times in the past four weeks)

V284 In the past four weeks, did you or any household member go a whole day and night without eating anything because there was not enough money for food?

(Het u of 'n huishoudelike lid die afgelope vier weke 'n hele dag en nag gegaan sonder om iets te eet omdat daar nie genoeg geld vir kos was nie?)

- ○ Yes
- ○ No

Skip To: End of Block If In the past four weeks, did you or any household member go a whole day and night without eating a... = No

Display This Question:

If In the past four weeks, did you or any household member go a whole day and night without eating a... = Yes

V285 How often did this happen? Rarely, Sometimes or Often?

[Let participant respond and then clarify what that option means to confirm it's correct].

- ○ Rarely (once or twice in the past four weeks)
- ○ Sometimes (three to ten times in the past four weeks)
- ○ Often (more than ten times in the past four weeks)

End of Block: Household Food Insecurity Access Scale

Start of Block: Shopping/Cooking Behaviour

We are almost done.
Finally, I'd just like to ask you a few questions about grocery shopping and cooking.

shopping_freq How often do you go shopping for groceries?

- ○ Every day
- ○ 2-3 times per week
- ○ Once a week
- ○ Less than once a week

hh_shopper Who normally does the grocery shopping for your household?

- [ ] Myself
- [ ] My partner
- [ ] Children
- [ ] Other (specify) ____

number_of_mouths When you buy groceries, how many household members excluding yourself eat the food you buy? ____

purchase_factors What is important for you in deciding what food to buy? List the 3 most important things from the following choices:

- [ ] Price
- [ ] Taste
- [ ] Distance to travel to shop
- [ ] Healthy food
- [ ] Brand of food/shop
- [ ] Environmental impact
- [ ] Food my other household members like
    - [ ] Other (specify) ____

favourite_shops What shops do you/your household buy food from most often? List the 3 most common shops from the following:

- ☐ Pick n Pay
- ☐ USave
- ☐ Checkers
- ☐ Food Lovers Market
- ☐ Shoprite
- ☐ Fruit and Veg
- ☐ Woolworths
- ☐ Spar
- ☐ Street Vendors
- ☐ Local supermarket
- ☐ Local farm outlets
- ☐ Other (Specify) ____
- ☐ Don't know

hh_cook Which statement most accurately describes your cooking situation at home:

- ○ I don't normally cook for myself or my household
- ○ I normally cook just for myself
- ○ I normally cook for myself and other members of my household
- ○ I normally don't cook for myself but I cook for other members of my household

Display This Question:

If Which statement most accurately describes your cooking situation at home:(pause for a second betw... = I don't normally cook for myself or my household

other_cook If you don't cook for yourself, who normally cooks for you?

- ○ Partner
- ○ Child
- ○ Other ____

Display This Question:

If Which statement most accurately describes your cooking situation at home:(pause for a second betw... = I normally cook for myself and other members of my household

Or Which statement most accurately describes your cooking situation at home:(pause for a second betw... = I normally don't cook for myself but I cook for other members of my household

hh_members_fed If you cook for other members of your household, how many people do you normally cook for, excluding yourself?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6 or more

End of Block: Shopping/Cooking Behaviour

F. Food Frequency Questionnaire

Survey Flow

Block: Introduction (2 Questions)
Standard: Breads (7 Questions)
Standard: Cereals and grains (7 Questions)
Standard: Fruit and Veg (9 Questions)
Standard: Animal protein (15 Questions)
Standard: Dairy (9 Questions)
Standard: Added fats (11 Questions)
Standard: Legumes, nuts and seeds (8 Questions)
Standard: Drinks (9 Questions)
Standard: Junk food (5 Questions)
Standard: Dessert, biscuits, sweets (10 Questions)

Page Break

Start of Block: Introduction

This is a Food Frequency Questionnaire (FFQ) asking about foods you have eaten in the past month.
It is important that you only answer about foods that you actually ate and just in the past month.
I will read out a list of foods and the portion units for each food. You will be asked to say on average how many of these portions you usually eat at a time and how times you have eaten them. Try to think about all the times that you have eaten the particular food, including in take-aways, as snacks, or as part of other dishes.
For example, I may say: 'A piece of chicken'. If you normally eat only one piece whenever you eat chicken and you eat chicken twice a week, you would answer: 'One piece, twice a week'. If you normally eat 3 pieces but only on one day of the week, you would answer: 'Three pieces, once a week'.
Another more complicated example may be 'A cup of non-starchy vegetables' and I would give examples such as 'broccoli, spinach, cabbage, peppers, salad greens etc'. Try to think of all the times you have eaten any type of non-starchy vegetables like these in the past month. So you would add up all the times you ate any of them in salads, soups, as snacks on their own, in curries etc. You would then think how many cups you usually ate each time. So some answers may be: 'half a cup, once a day' or '2 cups, 3 times a week' or 'none' if no starchy vegetables were eaten in the past month.
Try to be as honest as possible. Please try not to let your answers be affected by what you think you should have eaten. We want to know what you actually ate and will keep your answers strictly confidential. Everyone else besides me will only see a code next to your answers so they won't know they are yours.
Take as much time as you like to think about your answers. Once we get going, it should be more clear what to do but please feel free to ask me questions at any time.

Shall we start?
There are 18 questions in this survey.

id Enter the code that was given to you by the researchers.

____

End of Block: Introduction

Start of Block: Breads

instructions Thinking back over just the past month, on average, how often did you eat the following foods? Remember to state the number of portions you usually ate at a time and how often you ate them.

breads[white] A slice of white bread or roll.

- ○ None
- ○ 1 or less per month
- ○ 2 to 3 per month
- ○ 1 - 2 per week
- ○ 3 - 4 per week
- ○ 5 - 6 per week
- ○ 1 per day
- ○ 2 - 3 per day
- ○ More than 3 per day

Notes: Unless otherwise listed all the dropdown choices are the same for each of the following questions. We omit them here for brevity.

breads[brown] A slice of brown, wholewheat or rye bread or roll.

breads[roti] A roti.

breads[lowcarb] A slice of low carb bread (e.g. Heba or other)

breads[crackers] A savoury cracker or biscuit.

breads[pastry] A savoury pastry (e.g. croissant).

End of Block: Breads

Start of Block: Cereals and grains

cereals[cereal] A bowl of regular breakfast cereal or granola (e.g. All Bran, musli, Cornflakes etc.)

cereals[oats] A bowl of cooked porridge or oats.

cereals[lowcarb] A bowl of cooked (low carb) Heba pap.

cereals[pap] A bowl of regular cooked pap.

cereals[pasta] A cup of cooked pasta or noodles.

cereals[rice] A cup of cooked rice.

cereals[millet] A cup of other cooked grains (e.g. millet, cous cous, sorghum).

End of Block: Cereals and grains

Start of Block: Fruit and Veg

instructions For the following fruit and vegetables, try to think about all the times you have eaten them either alone, as salads or as parts of other dishes.

fruit&veg[potatoes] A medium white potato (excluding hot chips and crips).

fruit&veg[starchy] A cup of starchy vegetables (e.g. sweet potato, butternut, pumpkin, carrots, corn, peas etc.)

fruit&veg[nonstarchy A cup of non-starchy vegetables (e.g. broccoli, spinach, cabbage, peppers, salad greens etc.).

fruit&veg[olives] A small handful of olives.

fruit&veg[avocado] An avocado.

fruit&veg[berries] A small handful of berries.

fruit&veg[banana] A piece of any other type fruit (e.g. an apple, naartjie, orange, banana, grapes, mango etc.)

fruit&veg[driedfruit A small handful of dried fruit.

End of Block: Fruit and Veg

Start of Block: Animal protein

protein[red_meat] A palm sized piece of fresh red med e.g. beef, lamb or pork (including sausage but excluding dried and cured meats below)

protein[poultry] A piece of fresh poultry e.g. chicken, turkey etc.

protein[biltong] A small handful of biltong or droewors.

protein[processed] A slice of any type of cured/cold meats e.g. bacon, salami, sandwich ham, chicken roll, polony, viennas etc.

protein[offal] A cup of organ meats e.g. liver or kidneys.

protein[freshfish] A palm sized piece of fresh or frozen fish.

protein[canned_fish] A can of tinned fish e.g. pilchards, sardines, tuna etc.

protein[shellfish] A palm sized portion of shellfish (e.g. mussels, oysters).

protein[egg] An egg.

protein[shakes] A heaped table spoon of protein powder.

protein[bonebroth] A cup of bone broth or homemade stock.

Display This Question:

If A palm sized piece of fresh red med e.g. beef, lamb or pork (including sausage but excluding drie... != None

meat_fat When you ate red meat (beef, lamb, pork, game etc.) in the past month, how often did you also eat the fat with the meat?

- Never ate the fat
- Some of the time
- About half the time
- Most of the time
- Always ate the fat

Display This Question:

If A piece of fresh poultry e.g. chicken, turkey etc. != None

chicken_skin When you ate poultry (e.g. chicken), how often did you also eat the skin?

- Never ate the skin
- Some of the time
- About half the time
- Most of the time
- Always ate the skin

Display This Question:

If A piece of fresh poultry e.g. chicken, turkey etc. != None

chicken_breaded When you ate poultry (e.g. chicken) in the past month, how often was it breaded (crumbed), or battered (dipped in flour)?

- Never breaded or battered
- Some of the time
- About half the time
- Most of the time
- Always breaded or battered

Display This Question:

If A palm sized piece of fresh or frozen fish. != None

fish_breaded When you ate fish, how often was it breaded (crumbed), or battered (dipped in flour)?

- Never breaded or battered
- Some of the time
- About half the time
- Most of the time
- Always breaded or battered

End of Block: Animal protein

Start of Block: Dairy

instructions In the past month, how often did you have:

dairy[wholemilk] A cup of milk (add up all milk in hot drinks, cereal, on own etc.).

diary[cream] A tablespoon of cream or sour cream.

dairy[yoghurt] A cup of yoghurt.

dairy[hard_cheese] A matchbox sized portion of hard cheese e.g. gouda, cheddar, edam, feta etc.

dairy[soft_cheese] A matchbox sized portion of soft cheese e.g. cottage cheese, cream cheese etc.

milk_full_cream When you drank milk in the past month, how often was it full cream?

- Never full cream
- Some of the time
- About half the time
- Most of the time
- Always full cream

yoghurt_double_cream When you ate yoghurt in the past month, how often was it full or double cream?

- Never full cream
- Some of the time
- About half the time
- Most of the time
- Always full cream

yoghurt_flavoured When you ate yoghurt in the past month, how often was it flavoured?

- Never flavoured
- Some of the time
- About half the time
- Most of the time
- Always flavoured

End of Block: Dairy

Start of Block: Added fats

instructions For the following added fats, try to think of how much you ate altogether including when used for cooking meals, on their own, or when added to prepared food.

fats[butter] A table spoon of real butter or ghee.

fats[margarine] A table spoon of margarine.

fats[lard] A table spoon of lard or animal fat (excluding fat as part of meat).

fats[olive_oil] A table spoon of olive oil.

fats[coconut_oil] A table spoon of coconut oil.

fats[canola] A table spoon of any other vegetable oil like sunflower, canola, soybean, corn, or palm oil.

fats[coconut_milk] A cup of coconut milk or coconut cream.

fats[salad_dressing] A table spoon of salad dressing (only if not included in the oils mentioned above)

fats[mayonnaise] A table spoon of mayonnaise

fats[condiments] A table spoon of condiment sauces added to food e.g. tomato ketchup, chutney, BBQ, chilli, mustard etc.

End of Block: Added fats

Start of Block: Legumes, nuts and seeds

legumes[baked_beans] A cup of baked beans.

legumes[beans_other] A cup of soy beans.

legumes[kidney_beans A cup of other beans e.g. kidney, butter etc

legumes[lentils] A cup of lentils.

legumes[peanuts] A small handful of peanuts or cashew nuts

legumes[nuts_other] A small handful of other nuts e.g. macadamia, pecans, walnuts, almonds, brazil etc.

legumes[nut_butter] A table spoon of nut butter e.g. peanut butter legumes[seeds] A small handful of seeds

End of Block: Legumes, nuts and seeds

Start of Block: Drinks

drinks[coke] A can of regular soft drink e.g. coke, sprite, jive (340 ml)

drinks[diet_coke] A can of diet soft drink (340 ml)

drinks[juice] A glass of fruit juice or squash (300 ml)

drinks[energy] A can of energy or sports drink (340 ml)

drinks[milo] A cup of hot chocolate, milo or other milk drink (250 ml)

drinks[beer] A can of beer (340 ml)

drinks[wine] A glass of wine (175 ml)

drinks[cider] A can of alcoholic cooler or cider (340 ml)

drinks[spirits] A shot of spirits e.g. whiskey, vodka, gin etc. (25ml)

End of Block: Drinks

Start of Block: Junk food

snacks[pizza] A medium pizza

snacks[samosa] A small samosa

snacks[hot_chips] A cup of hot chips or french fries

snacks[pie] A regular sized pie

snacks[crisps] A small packet of crisps or popcorn

End of Block: Junk food

Start of Block: Dessert, biscuits, sweets

sugar[dessert] A bowl of any kind of pudding dessert e.g. ice cream, chocolate mouse etc. but not cakes below.

sugar[cake] A medium sized piece of any kind of cake or sweet pastry e.g. cake, muffin, brownie, waffle, pancake, milk tart, doughnuts, koeksisters, chelsea bun etc.

sugar[biscuit] A sweet biscuit

sugar[rusk] A rusk

sugar[sweets] A packet of sweets (50g)

sugar[milk_choc] A small bar of milk chocolate (50 g)

sugar[dark_choc] A small bar of dark chocolate (50 g)

sugar[tsp_sugar] A teaspoon of sugar - on its own or added to foods and drinks e.g. in tea, coffee, cooked meals, cereal etc

sugar[honey_jam] A teaspoon of honey, syrup, jam, or nectar eg coconut nectar - on own or added to foods and drinks

sugar[non_caloric] A teaspoon of non-caloric sweeteners added to foods or drinks e.g. aspartemine, stevia, sucralose, xylitol etc. (excluding in food products such as diet sodas)

End of Block: Dessert, biscuits, sweets

G. Traffic lights lists of the Eat Better South Africa programme

G

THE GREEN LIST

THE GREEN FOOD LIST IS THE ONLY LIST THAT YOU CAN EAT FROM ON A DAILY BASIS. THESE ARE THE FOODS THAT ARE NUTRITIOUS, LOW IN CARBS PER PORTION AND EXTREMELY HEALTHY. PRACTICING PORTION CONTROL IS STILL IMPORTANT WHEN EATING FROM THIS LIST

THE BANTING POCKET GUIDE QUICK REFERENCE LISTS

For a more extensive list of foods, including the macro nutrient breakdown, please refer to our book The Banting Pocket Guide.

ANIMAL PROTEIN
All eggs
Beef / veal
Mutton / Lamb
Pork
Venison/game
Ostrich
All Poultry :
Chicken
Duck
Turkey
Offal:
Brain
Brawn
Tripe
Trotters
Liver, heart, kidneys
Tongue
Chicken feet/heads/gizzards
Naturally cured meats and sausage
Bacon
Chorizo
Pancetta
Salami
Sausage

SEAFOOD
Fish – fresh and canned in brine
Calamari
Crab
Oysters
Prawns

DAIRY*
Amasi
Buttermilk
Coconut milk
Cow's milk – full cream
Cheese hard and soft
Cottage cheese
Cream cheese
Cream – fresh/sour
Yoghurt full cream/Greek
*Using dairy products may stall weight loss in some people.

FATS AND OILS
Beef tallow
Butter
Duck fat
Ghee
Lard
Almond oil
Avocado oil
Coconut oil
Olive oil
Macadamia nut oil

FLAVOURING & CONDIMENTS
All natural herbs and spices are acceptable if they do not contain sugars and chemical additives. Includes
Aniseed, Basil, Capers, Caraway seed, Cardamom, Chillies, Cinnamon, Coriander, Curry powder, Dill, Fennel, Garlic, Ginger, Horseradish, Marjoram, Masala, Organum, Paprika, Parsley, Pepper, Peppermint, Rosemary, Sage, Thyme, Turmeric. Vinegar, including Apple cider.

BEVERAGES
Coffee (100% pure coffee)
Tea- including green tea and Rooibos
Water, soda water, sparkling mineral water.

NUTS & SEEDS
Almond, Brazil nuts, Coconut, Macadamia nut, Pecans, Pine nuts, Pistachio nuts, Walnuts.
Chia seed, Flax seed, Linseed, Pumpkin seed, Sesame seed, Sunflower seed
HEBA, Psyllium husk

SWEETENERS
Xylitol granules
Erythritol granules
Stevia powder
NOTE: We do not recommend artificial sweeteners of any kind. It is our opinion that if you want to stay lean and healthy for the rest of your life you need to avoid all foods that taste sweet. The desire to eat sweet foods is the addiction that drives poor food choices leading to obesity and ill health.

VEGETABLES
Amaranth/marog
Artichokes - globe
Asparagus
Aubergine
Broccoli
Brussels sprouts
Cabbage
Calabash / gourd
Cauliflower
Celery
Chives
Collards
Cucumber
Endive
Gherkins (dill, sugar free)
Green beans
Kale
Kohlrabi
Leek - boiled
Lettuce
Mixed frozen vegetables (cauliflower, carrot, green beans)
Mushrooms
Okra
Onion
Pepper- green, red, yellow
Pumpkin
Radish
Sauerkraut
Seaweed
Sousou/ chayote
Spinach
Spring onion
Squash - gem, hubbard,
Squash – baby marrow
Sugarsnap peas
Tomato
Turnip
Waterblommetjies
Wild rocket

FRUITS
Avocado
Olives

THE ORANGE LIST

THE ORANGE FOOD LIST IS FOR PEOPLE WHO HAVE REACHED THEIR GOAL WEIGHT AND WANT TO INCLUDE SOME VEGETABLES AND BERRIES ON THIS LIST, OR FOR THOSE WHO ARE NOT SENSITIVE TO CARBOHYDRATES AND CAN TOLERATE THESE VEGETABLES AND FRUITS. THIS LIST IS ALSO FINE FOR AN OCCASIONAL SWEET TREAT, **BUT ONLY ONCE YOU HAVE REACHED YOUR GOAL WEIGHT.** WE HAVE INSERTED THE CARB COUNT HERE SO YOU CAN BE AWARE OF THE HIGHER CARB VALUES:

THE BANTING POCKET GUIDE QUICK REFERENCE LISTS

For a more extensive list of foods, including the macro nutrient breakdown, please refer to our book The Banting Pocket Guide.

VEGETABLES per 100g

Artichoke 14.3g
Beetroot 7.96g
Carrot boiled 5.3g
Carrot raw 6.4g
Leek – raw 12.4g
Parsnip 13.01g
Squash – Butternut 10.2g
Sweet potato – orange 17.4g
Sweet potato – white 15.1g
Tomato – sundried (per 25g) 10.9g

PROTEINS

Abalone (per 125g) 14.6g
Mussel (per 100g) 7.4g
Perlemoen (per 125g) 14.6g
Snails (per 75g) 11.6g

FRUIT per 50g

Apple 6.5g
Apricot 6.5g
Banana 9.4g
Blackberries 4.3g
Blueberries 6.1g
Cranberries 3.8g
Figs 6.8g
Gooseberries 6.0g
Granadilla 6.5g
Grape 7.4g
Guava 7.7g
Kiwifruit 6.5g
Kumquat 4.7g
Lemon 7.0g
Lime 7.7g
Litchi 8.6g
Melon green flesh 4.5g
Melon orange flesh 4.1g
Naartjie 5.0g
Nectarine 5.2g
Orange 4.6g
Papaya 4.6g
Pawpaw 4.3g
Peach 4.3g
Pear 7.2g
Pineapple 6.1g
Plum 5.5g
Raspberries 2.6g
Strawberries 3.0g
Watermelon 3.0g
Youngberries 2.15g

SWEETENERS

Honey (per 5g) 4g

NUTS per 30g

Betel nut 16.1g
Chestnut 13.3g
Cashew nut 8.9g

THE BANTING POCKET GUIDE

EAT BETTER SOUTH AFRICA!

THE NOAKES FOUNDATION

10 BASIC RULES OF BANTING

1. Banting is about eating when hungry and stopping when satisfied.

2. Eat clean, fresh, real food. Real food goes off and has a very short shelf life. Do not eat processed or pre-packaged foods.

3. Make sure that you include fats, proteins and healthy carbs in all your meals, whether you are eating three meals a day or only two. Meals must be nutrient dense and well balanced.

4. Do not eat more than three meals a day; there is no rule dictating which time of the day you should eat or that you have to eat all three meals.

5. Do not have sweeteners in your coffee or tea; go cold turkey if you want to see results.

6. Drink water throughout the day, but only when you are thirsty.

7. Make sure you are getting enough vitamins and minerals. If you experience energy loss in the beginning, you may supplement.

8. Do not drink any fizzy drinks, fruit juices or 'slimming' drinks, not even if they claim to be sugar free. They all contain artificial sweeteners and additives that can have a negative effect on your health and weight.

9. Do not snack between meals unless you are really hungry. Snacking between meals can lead to weight gain.

10. What works for you may not necessarily work for others. We are all unique.

R

THE RED LIST

THE RED FOOD LIST ITEMS MUST BE AVOIDED AT ALL COSTS. WE DON'T EVEN RECOMMEND THESE FOODS AS A ONCE-IN-A-WHILE TREAT, AS THEY ARE HIGHLY PROCESSED AND CONTAIN UNHEALTHY ADDITIVES AND CHEMICALS.

RED ITEM FOODS WILL ALMOST ALWAYS CONTAIN INGREDIENTS THAT ARE HARD TO PRONOUNCE

THE BANTING POCKET GUIDE QUICK REFERENCE LISTS

For a more extensive list of foods, including the macro nutrient breakdown, please refer to our book The Banting Pocket Guide.

ALL PRODUCTS CONTAINING ANY OF THESE INGREDIENTS
Atta (chapatti flour)
Breaded or battered foods
Cake flour, Chickpea flour
Corn flour, Durum (wheat)
Malt, Matzo meal, Modified wheat starch
Oatmeal, Oat bran, Whole oats
Potato starch, Rice flour
Semolina, Sorghum, Soy flour
Dried beans, Couscous
Lentils, Pasta, Polenta
Rice, Samp
Split peas, Stampkoring
Wheat germ, Wheat starch

BEVERAGES
Canned coffee – generally containing other ingredients like dextrose, etc
Tea with added artificial ingredients
Fizzy drinks including diet or lite drinks
Cordials, Fruit drinks, Fruit juice
Shakes of any kind
Energy drinks

ALCOHOL
Beer
Ciders
Dessert wine
Liqueurs & Shooters

DAIRY
All low fat/ fat free products
Cheese spreads, Processed cheese
Canned cream, Dessert cream
Coffee creamer
Condensed milk
Custard
Flavoured yoghurt
Ice cream
Powdered milk, Rice milk, Soy milk

FATS AND OILS
All commercial fat spreads/ margarine
Flavoured butters
Canola oil, Corn oil
Cottonseed oil, Grapeseed oil,
Soybean oil,Sunflower oil

SAUCES AND DRESSINGS
All commercial sauces and dressings
Barbeque sauce, Cook in sauce, Marinades, Mustard sauce, Peri-peri sauce, Pasta sauce, Salad creams and dressings
Tomato sauce
Sweet sauces

FAST FOOD AND TAKEAWAYS
Burgers, Hot dogs, Spare ribs,
Crumbed chicken or fish
Fries, Wraps, Pizza, Hotdogs

MEAT AND FISH
All meat that has been cured with sugar and/or marinated meats with added ingredients
Corned meat
Cold processed meats, e.g. sandwich ham/ham/chicken/beef, etc generally found at the deli
Crumbed/battered meat, e.g. crumbed chicken, hamburger patties, chicken nuggets, meat pies, readymade meals, meat free products (soy), fish bakes, crumbed fish fingers
Pilchards in tomato sauce
Tuna in vegetable oil

FRUIT AND VEGETABLES
Dried fruit – all varieties
Legumes
Corn
Potatoes

SWEETENERS
Agave
Aspartame
Blackstrap molasses
Cane sugar, Beet sugar
Castor sugar
Coconut sugar, Date sugar
Carob syrup, Corn syrup, Maple syrup
Dextrose
Fructose
Glucose
Maltitol
Saccharin
Sorbitol
Sucralose
Table sugar
Tapioca sugar
Treacle

Eat Better South Africa!

THE BANTING POCKET GUIDE

THE NOAKES FOUNDATION

7. REFERENCES

Alemu, M. H., & Olsen, S. B. (2018). Can a Repeated Opt-Out Reminder mitigate hypothetical bias in discrete choice experiments? An application to consumer valuation of novel food products. *European Review of Agricultural Economics, 45*(5), 749-782.

Archer, E., Lavie, C. J., & Hill, J. O. (2018). The Failure to Measure Dietary Intake Engendered a Fictional Discourse on Diet-Disease Relations. *Frontiers in Nutrition*, *5*. https://doi.org/10.3389/fnut.2018.00105

Cade, J., Thompson, R., Burley, V., & Warm, D. (2002). Development, validation and utilisation of food-frequency questionnaires–a review. *Public Health Nutrition*, *5*(4), 567–587.

Cawley, J. (2004). An economic framework for understanding physical activity and eating behaviors. *American Journal of Preventive Medicine*, *27*(3), 117-125.

Charness, G., Cobo-Reyes, R., Eyster, E., Sánchez, Á., & Sutter, M. (2020). Improving healthy eating in children: Experimental evidence. *MPI Collective Goods Discussion Paper*, (2020/1).

Colditz, G. A., & Hankinson, S. E. (2005). The Nurses' Health Study: lifestyle and health among women. *Nature Reviews Cancer, 5*(5), 388-396.

Colditz, G. A., Philpott, S. E., & Hankinson, S. E. (2016). The impact of the Nurses' Health Study on population health: prevention, translation, and control. *American Journal of Public Health, 106*(9), 1540-1545.

Crosetto, P., Muller, L., & Ruffieux, B. (2016). Helping consumers with a front-of-pack label: Numbers or colors?: Experimental comparison between Guideline Daily Amount and Traffic Light in a diet-building exercise. *Journal of Economic Psychology, 55*, 30-50.

Crous-Bou, M., Fung, T. T., Prescott, J., Julin, B., Du, M., Sun, Q., ... & De Vivo, I. (2014). Mediterranean diet and telomere length in Nurses' Health Study: population based cohort study. *BMJ*, *349*.

Freire, R. (2020). Scientific evidence of diets for weight loss: Different macronutrient composition, intermittent fasting, and popular diets. *Nutrition*, *69*, 110549.

Hallsworth, M. (2017). Rethinking public health using behavioural science. *Nature Human Behaviour*, *1*(9), 612-612.

Keswell, M., & Poswell, L. (2004). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. *South African Journal of Economics, 72*(4), 834-860.

List, J. A., & Samek, A. S. (2015). The behavioralist as nutritionist: leveraging behavioral economics to improve child food choice and consumption. *Journal of Health Economics*, *39*, 135-146.

Liu, S., Stampfer, M. J., Hu, F. B., Giovannucci, E., Rimm, E., Manson, J. E., … Willett, W. C. (1999). Whole-grain consumption and risk of coronary heart disease: results from the Nurses' Health Study. *The American Journal of Clinical Nutrition*, *70*(3), 412–419. https://doi.org/10.1093/ajcn/70.3.412

Lusk, J. L., & Schroeder, T. C. (2004). Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. *American Journal of Agricultural Economics*, *86*, 467–482. https://doi.org/10.2307/30139569

Medic, N., Ziauddeen, H., Forwood, S. E., Davies, K. M., Ahern, A. L., Jebb, S. A., … Fletcher, P. C. (2016). The Presence of Real Food Usurps Hypothetical Health Value Judgment in

Overweight People. *ENeuro*, *3*(2). https://doi.org/10.1523/ENEURO.0025-16.2016
Müller, L., and Prevost, M. (2016). What cognitive sciences have to say about the impacts of nutritional labelling formats. *Journal of Economic Psychology, 55*, 17-29.
Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *The International Journal of Behavioral Nutrition and Physical Activity*, *5*, 56. https://doi.org/10.1186/1479-5868-5-56
Pujol-Busquets, G., Smith, J., Larmuth, K., Fàbregues, S., & Bach-Faig, A. (2020). Exploring the Perceptions of Women from Under-Resourced South African Communities about Participating in a Low-Carbohydrate High-Fat Nutrition and Health Education Program: A Qualitative Focus Group Study. *Nutrients*, *12*(4), 894.
Roberto, C., & Kawachi, I. (2015). *Behavioral Economics and Public Health*. Retrieved from https://books.google.co.za/books?hl=en&lr=&id=WXQ_CgAAQBAJ&oi=fnd&pg=PP1&dq=roberto+and+kawachi+2015&ots=fZh0dkRYM4&sig=4FKHNhTOMwFnFMz1JX1Xa5EoTz0
VanEpps, E. M., Downs, J. S., & Loewenstein, G. (2016). Calorie Label Formats: Using Numeric and Traffic Light Calorie Labels to Reduce Lunch Calories. *Journal of Public Policy & Marketing*, *35*(1), 26–36. https://doi.org/10.1509/jppm.14.112
Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., ... & Bell, M. L. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet, 388*(10053), 1459-1544.
Webster, C. C., Murphy, T. E., Larmuth, K. M., Noakes, T. D., & Smith, J. A. (2019). Diet, Diabetes Status, and Personal Experiences of Individuals with Type 2 diabetes Who Self-Selected and Followed a Low Carbohydrate High Fat diet. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 12*, 2567–2582. http://doi.org/10.2147/DMSO.S227090

Chapter 4. The effect of technology-assisted behavioural interventions in type 2 diabetes[42]

Sofia Monteiro*[1,2] Kate Larmuth[3] Daniel Wiesen[4] Jacolene Kroff[3] Matthias Sutter[1,2,5]

[1]Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-straße 10, 53113 Bonn, Germany

[2]Faculty of Economic and Social Sciences, University of Cologne, Universitätsstraße 24, 50931 Cologne, Germany

[3]Division of Exercise Science and Sports Medicine (ESSM), Department of Human Biology, Faculty of Health Sciences, University of Cape Town, South Africa

[4]Department of Business Administration and Healthcare Management, Faculty of Economic and Social Sciences, University of Cologne, Universitätsstraße 91, 50931 Cologne, Germany

[5]University of Innsbruck, Austria

Abstract. The burden of lifestyle-driven type 2 diabetes (diabetes) is a growing global problem. There is a gap in the literature on experiments evaluating the use of Continuous Glucose Monitoring (CGM) to counsel patients to follow a ketogenic diet with the aim to reverse diabetes. We provide the first evidence. Here, 95 diabetic patients (not using insulin) are randomly assigned to one of three stepwise treatments. CONTROL is a self-monitoring programme that includes ketogenic dietary advice (<25g carbohydrate/day) from a primary care physician, fingerprick testing before and after meals and reviewing a CGM report at doctor's visits only. INFO adds real-time feedback CGM as part of the same programme. COACH adds one-to-one online health coaching sessions on top of INFO. All groups show significant improvement in mean glucose levels as measured by the CGM. The rich report of the CGM closes the information gap between patient and doctor, and offers an accountability mechanism. However, we do not find evidence of

[42] This research was funded by the Max Planck Institute for Research on Collective Goods, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1–390838866. The University of Cape Town Health Sciences Faculty approved the research protocol HREC REF: 310/2020. We thank the study physicians Dr Neville Wellington and Dr Carol Bosch, Sister Nthabeleng Khatiti, and health coach Lauren Angelil. Excellent research assistance was provided by Laris Gomez.

the value added by real-time information on its own. All groups achieve the time in target goal of 70%. After a sustainability period of two months without CGM, INFO has significantly worse outcomes than CONTROL, though an improvement compared to baseline. COACH achieves the greatest reduction in Haemoglobin A1c (HbA1c, average blood glucose over the past 2 months used to diagnose diabetes) of about 19%, while CONTROL reduces by 14% and INFO 11%. OLS regression results show that COACH achieves clinically significantly better HbA1c. Only COACH significantly improves other blood markers of metabolic health, HDL (cholesterol) and triglycerides. In the blood tests, INFO performs significantly worse than CONTROL after the sustainability period in which the CGM effect is allowed to wash out. The success of the coaching group was not due to scanning the real-time CGM more often. COACH scans significantly less often, but the additional weekly support from someone besides the doctor appears to provide important supervision of real-time CGM in this population in South Africa. All treatments see a significant reduction in HbA1c compared to baseline. In the pooled sample, 43% of patients reversed diabetes (HbA1c<6.5%), notable compared to the best practice standard of care reversal rate of <1%.

JEL classification: C19, D91, I12

Keywords: Field Experiments, Health Behaviour, Diabetes, Coaching, Continuous Glucose Monitoring

* Corresponding author: sofia.g.monteiro@gmail.com

1. INTRODUCTION

1.1. The diabetes burden

Diabetes is a growing global public health burden (Cho et al., 2018; IDF, 2019). Globally, an estimated 463 million adults live with diabetes, with 59 million in Europe alone (IDF, 2019). Its prevalence is rising rapidly, with millions of new cases diagnosed each year (Cho et al., 2018). Diabetes increases the risk of kidney disease, vascular disease, blindness and amputations, and thus poses a considerable burden on patients, families, health insurance providers and the wider economy (IDF, 2017). Diagnosed diabetes cost the United States $327 billion in 2017 (up 26% since 2012).[43] It is a serious threat to human health that respects neither socioeconomic status nor national boundaries (IDF, 2019).

In the next 25 years, Sub-Saharan Africa (SSA) will have the greatest increase of diabetes in the world at 143% (Mba & Mbanya, 2020). Only 5.5% of the GDP in SSA was spent on health in 2016 compared to the global estimate of 10% (Mba & Mbanya, 2020). Although SSA has witnessed a recent increase in diabetes incidence, the current diabetes-related expenditure of US$9.5 billion represents only 1% of the overall global expenditure on this disease (Mba & Mbanya, 2020). These include the direct costs associated with medical care and the indirect costs resulting from unemployment or lower productivity because of disease-related disability or premature death. Indirect cost per patient with diabetes in SSA is higher than direct expenditures, contributing over 60% of the total cost (Mba & Mbanya, 2020). In South Africa, diabetes was the leading underlying cause of death among women in 2016 (StatsSA, 2018: Mortality and Causes of Death Report). The Global Burden of Disease Study (Wang et al., 2016) estimates that in 2015, the overall cost of diabetes in sub-Saharan Africa was $19.45 billion or 1.2% of cumulative gross domestic product (GDP). Preventing and reversing metabolic disease has considerable economic value for stakeholders across countries (Mba & Mbanya, 2020). Type 2 Diabetes Mellitus (T2DM) is a disorder of metabolism due to insulin resistance. Lifestyle choices play an important role in the management of all diabetes conditions (IDF, 2019; Wang et al., 2016; Moholdt et al., 2021). Societies have to consider designing health-promoting environments that enable people to make the healthy choice the easier choice (IDF, 2019).

The development of T2DM is regularly preceded by a collection of modifiable risk factors known as Metabolic Syndrome (Alberti et al., 2005).[44][45] There is evidence from multiple randomised control trials (RCTs) (Bazzano et al., 2014; J. S. Volek et al., 2009; J. Volek et al., 2004), meta-analyses (Hashimoto et al., 2016; Meng et al., 2017) and narrative reviews (Hallberg et al., 2019;

---

[43] The American Diabetes Association (2018). Economic Costs of Diabetes in 2017.

[44] These include expanding waist circumference, elevated blood glucose, blood triglycerides and blood pressure, and lowered high-density lipoprotein (HDL) levels (Alberti et al., 2005).

[45] With regard to diet, overconsumption of ultra-processed foods that are rich in added sugar, carbohydrates and ultra-processed vegetable oils are believed to play a predominant role (Simopoulos, 2016; Stanhope, 2016).

Noakes & Windt, 2017; Paoli, Rubini, Volek, & Grimaldi, 2013; J. Volek, Fernandez, & Feinman, 2008) that reducing consumption of sugar and carbohydrates and increasing consumption of healthy fats can improve Metabolic Syndrome and T2DM. Compared to control diets, lower-carbohydrate diets consistently improve metabolic health (e.g. reducing fasting glucose, HbA1c, body weight, body fat, blood triglycerides, and fasting insulin) (Hallberg et al., 2019a; Noakes & Windt, 2017). This motivates the use of the Low-Carb Healthy-Fats diet and blood tests that we use in this paper.

T2DM has long been identified as an incurable, progressive chronic disease.[46] About 50% of patients with T2DM need to start injecting insulin within ten years of diagnosis (Hallberg et al., 2019b). However, the 2016 WHO global report on diabetes acknowledged that reversal (HbA1c under the diabetes threshold of 6.5% for an extended period of time) can be achieved through weight loss and calorie restriction. Best practice standard of care has a remission rate of only 0.23% (Kaiser Permanente, 2014). This is one benchmark that we compare our rate of reversal to in this study. A 2019 review of the evidence suggests that T2DM reversal is possible through alternative approaches of bariatric surgery, low-calorie diets or carbohydrate restriction (Hallberg et al., 2019b). Guidelines for standard of care should arguably reflect the best available evidence.

Continuous glucose monitors (CGM) and mobile applications (apps) can provide the patient and the supervising health practitioner with real-time feedback on the patient's glucose levels over the preceding eight hours with detailed insights into time spent in target glucose range, signalling hyper- and hypoglycaemia – measures that have predictive value for diabetes complications. The value of such technology in behaviour change lies in its ability to teach individuals about their own hyperglycaemic responses to certain foods and drinks, of which they would otherwise be unaware. However, there is limited evidence on using CGM to advise changes in diet. This is discussed further below. Individual-level data from CGM provide an opportunity to understand how behavioural interventions might produce healthier food choices and, in some cases, reverse diabetes.

[46] The best expected outcome with standard of care has been to ameliorate symptoms (high blood glucose and its associated extreme thirst and appetite, frequent urination, fatigue, blurred vision) and slow the inevitable progression with increasingly costly prescriptions.

In this paper, we conduct a behavioural experiment to evaluate the effect of real-time feedback from CGM alone, or in combination with online health coaching, on blood glucose levels and diet adherence. All patients receive the same dietary advice and set a **goal to bring their HbA1c below 6.5% and their time in target range (3.9 - 8 mmol/L) to 70% or more** by restricting total carbohydrate intake to less than 25g per day and cutting out processed foods containing vegetable oil.

This study highlights the use of CGM in a largely overlooked demographic – patients with T2DM – which far exceeds type 1 diabetics, for which CGM are typically prescribed. This research brings attention to the potential of CGM reports as a targeted behaviour-change tool for patients with T2DM. Changing dietary behaviour, ingrained by habit and societal norms, is a challenge that may require complementary accountability mechanisms. This research is highly relevant in the context of the global chronic disease burden, aggravating comorbidities to Covid-19 and preventing diabetes-related complications in particular. Below we examine the related literature and then provide a preview of the results.

1.2. Related literature

Recently, researchers observed 57 participants who were given a CGM to wear (Hall et al., 2018). Most were ostensibly healthy, some had prediabetes, and five had type 2 diabetes. They found that glucose fluctuations or spikes in apparently healthy people were sometimes as high as levels in people with diabetes and occurred after eating specific foods, *i.e.* refined or starchy carbohydrates.[47] Some participants exhibited more extreme glucose spikes than others. The researchers categorised them as low, moderate and severe glycaemic responses. These categories were positively correlated with clinically relevant measures of metabolic health (e.g. HbA1c, fasting glucose, fasting insulin and BMI). The researchers found that glucose fluctuations occurred much more widely than presumed and that traditional measuring methods (e.g. single-measure fingerprick[48]) were unable to predict the observed variability with the same accuracy. The study suggests there are many people experiencing glucose spikes regularly without knowing it. Real-

[47] "A standardised meal of cornflakes and milk caused glucose elevation in the range of prediabetes in 80% of individuals in our study".
[48] A standard fingerprick or fingerstick test of blood capillary glucose level.

time information provision could facilitate learning, healthier choices and habit formation by diabetics compared to fingerprick testing. Our study design allows us to empirically test this hypothesis, addressing a gap in the literature. The effect of real-time feedback in the behavioural economics literature has been explored in papers such as Fang et al. (2020), but it is not clear that their findings on energy conservation behaviour extend to health decision-making. Food habits, unlike energy consumption, are mediated by cultural identities, traditions and social interactions. Fang et al. (2020) argue that when multiple barriers are present, such as imperfect information and limited attention (and in our setting of food choice, self-control and impatience too), a single intervention that does not address all barriers simultaneously might fail to fulfil its potential. This motivates our third treatment that combines real-time feedback with health coaching.

An observational study of worldwide glucose testing patterns with the Abbott FreeStyle Libre CGM using data from their consumers showed greater testing frequency is associated with greater glucose control across countries (Dunn et al., 2018). Flash CGM users scanned on average 16.3 times per day [median (IQR): 14 (10-20)]. This is striking in contrast to rates of fingerprick tests by patients with diabetes doing self-monitoring of blood glucose (2.1 tests per day in the UK). Worldwide, estimated HbA1c gradually reduced from 8.0% to 6.7% as the scanning rate increased from lowest (4.4 scans/day) to highest (48.1 scans/day). Time in hypoglycaemia (< 3.9 mmol/L) decreased by 15% and time in hyperglycaemia (> 8 mmol/L) decreased by 44%, from 10.4 to 5.7 hours/day. Time in range increased by 40% from 12 to 16.8 hours/day. Real-world patterns clearly show that the flash CGM allows for frequent glucose checks and that higher scan rates are correlated with better glucose control indicators. These cross-country data provide useful reference points for us to assess the benefit of our treatments, but they do not include South Africa. There may be challenges to technology adoption in low income and low health-literacy settings.

Patients who only do a fingerprick test before meals do not observe their post-meal glucose excursions. They miss glucose response patterns that would allow them to learn more about their disease and carbohydrate intolerance. Real-time glucose feedback can potentially give individuals a unique understanding of how specific foods and drinks impact their glucose responses and enable them to make informed dietary choices accordingly (Ehrhardt & Al Zaghal, 2019). However, it is necessary to know what the information means in order to know how to base decisions on it.

Behavioural science suggests that public health could focus more on interventions that require little to no effort on the part of individuals (Hallsworth, 2017). Immediate feedback and self-monitoring can potentially teach CGM users through direct experience of linking cause and effect (Ehrhardt & Al Zaghal, 2019). This feature arguably makes it a relatively low-cost tool for making healthier choices, with negligible effort and pain for the diabetic patient. It is an open question whether providing real-time information on glucose patterns leads to better food choices compared to a structured fingerprick testing regime before and after meals with only intermittent feedback on CGM patterns discussed with a doctor.

A few existing studies explore diet quality and CGM data, as well as the potential for CGM in type 1 and type 2 diabetes management (Allen et al., 2009; Block, 2008; Carlson et al., 2017; Ehrhardt & Al Zaghal, 2019; Nansel et al., 2016). Beck et al. (2017) demonstrate that individuals with T2DM increase their Time in Range by 10.3% (from 55.6% to 61.3%) after 24 weeks of CGM use. Lower Time in Range is associated with worse diabetes complications. Beck et al. (2017) show that the hazard rate for retinopathy progression increases by 64% for each 10% reduction in Time in Range. The relationship between Time in Range (3.9 - 8 mmol/L) of 70% and 50% strongly correspond with an HbA1c of approximately 7% and 8%, respectively. An increase in Time in Range of 10% (2.4 h per day) corresponds to a decrease in HbA1c of approximately 0.5%. A change of 0.5% in HbA1c is considered clinically significant (Campbell et al., 2019; Lenters-Westra et al., 2014). Similar associations are made in an analysis of 18 randomised controlled trials by Vigersky and McMahon (2019). In two previous studies, clinicians use CGM data to counsel patients with type 2 diabetes to exercise more, finding a significant increase in physical activity compared to a non-CGM control group (Park & Le, 2018). A meta-analysis of 18 studies suggests there is strong evidence for the efficacy of smartphone apps for lifestyle modification in type 2 diabetes and inconclusive efficacy for other subtypes (Wu et al., 2019). Even small incremental improvements yield clinically significant glycaemic benefits. There is a gap in the literature on how CGM feedback (intermittent or real-time) can assist in nutrition education programmes for type 2 diabetics.

Efforts to change individual behaviour are frequently unsuccessful. This is despite the large amount of evidence-based literature that informs health and wellbeing guidelines (Kelly & Barker, 2016). Information does not necessarily lead to behaviour change. Our environment, social support, routine, emotions, strengths, and internal reward system all feed into our behaviour. Instilling long-term behaviour change arguably requires an understanding of motivation and the social and economic pressures that impact them (Kelly & Barker, 2016). Health coaching has emerged as a promising intervention in the management of non-communicable diseases. A functional medicine health coach supports clients in discovering their own strategies and motivations for change. They help clients to identify and overcome obstacles and to implement protocols that have either been prescribed by a clinician or that the client has chosen to implement on their own. Health coaches use behaviour change techniques such as positive psychology, motivational interviewing, and habit formation and reversal to implement lasting change. Health coaching interventions for patients with diabetes, cardiovascular disease, cancer and chronic obstructive pulmonary disease show positive effects on health outcomes (Thomas et al., 2011; Vale et al., 2003; Wolever et al., 2013; Wang et al., 2018). We incorporate health coaching in an outpatient primary care treatment that includes real-time feedback from a CGM, with the insight from behavioural economics that we expect multiple barriers to healthier food choices besides information, which can be measured using behavioural economics' and psychological scales of time, risk preferences, and grit.

A number of studies explore the effects of a diabetes health coaching intervention compared to usual practice within the standard model of diabetes care (Cho et al., 2011, Cinar et al., 2014, Frosch et al., 2011, Orsama et al., 2013, Ruggiero et al., 2010, Varney et al., 2014, Whittemore et al., 2004, Wolever et al., 2010). The goal of treatment is to achieve glycaemic targets while minimising adverse events to reduce the risk of short-term and long-term negative consequences. The pooled effect of health coaching overall is a reduction of HbA1c levels by 0.32 percentage points over a health coaching period of less than six months. This is one of the benchmarks that we compare our results to. We assume these studies use a standard dietary approach (and not a nutritional ketosis approach) since they did not explicitly report dietary recommendations.

A few studies explore health coaching within a continuous care intervention (CCI) programme as a tool to complement nutritional ketosis for the management of T2DM (Athinarayanan et al., 2019; Hallberg et al., 2018). The CCI includes bio-marker tracking, access to a mobile app, education, and telecommunication with a multidisciplinary team including a health coach and medical practitioner for advice and medical monitoring over a period of one year (Hallberg et al., 2018) or two years (Athinarayanan et al., 2019). Participants are advised to achieve and sustain nutritional ketosis (Beta-Hydroxybutyrate (Blood) level of 0.3-0.5 mmol/L) (Athinarayanan et al., 2019; Hallberg et al., 2018). It is shown that the CCI and nutritional ketosis is associated with significantly lower HbA1c, medication use and weight within 70 days (Hallberg et al., 2018), and this improves or is maintained over a one-year period. Over a two-year period, nutritional ketosis and CCI is associated with improvements in blood glucose, fasting insulin, HbA1c, weight, blood pressure, triglycerides, liver function and inflammation, and reduced dependence on medication over two years[49] (Athinarayanan et al., 2019). The authors do not state the typical frequency or duration of individual coaching sessions. The behavioural techniques utilised by the health coach and medical provider included education of "natural consequences", shaping knowledge, goal-setting and self-monitoring, feedback, monitoring and reinforcement (Hallberg et al., 2018). In both studies, the CCIs focus on education and support regarding dietary intervention, behaviour modification techniques for maintenance of lifestyle change and education modules covering core concepts on achieving nutritional ketosis (Athinarayanan et al., 2019).

In another study, physicians and health coaches are trained in the basic principles of achieving and maintaining nutritional ketosis and the CCI is limited to 90-minute weekly education seminars. A health coach is available for one-to-one texting advice and problem-solving, and clients have access to support via an online community of peers. In addition, they are supervised by a physician (McKenzie et al., 2017). However, no one-on-one virtual health coaching session is provided. At baseline, average HbA1c is 7.6%, and only 19.8% of participants have an HbA1c level of less than 6.5% (*i.e.* most are in the diabetic range). After ten weeks, average HbA1c is lowered by 1.0 percentage point, 56.1% achieved an HbA1c level of 6.5%, 65.1% of individuals achieved an HbA1c level of less than 6.5% (*i.e.* diabetes reversal), and 56.8% of individuals had one or more diabetes medications reduced or eliminated.

Several studies have shown the clinical benefits of early achievement of near-normal glycaemic control in people with diabetes (see Clinical Targets for Continuous Glucose Monitoring Data

[49] For CCI participants, the mean dose of prescribed insulin decreased by 81% from baseline to 2 years.

Interpretation by Battelino et al., 2019). We incorporated the SMART (Specific, Measurable, Achievable, Relevant, Time-bound) goal approach for our intervention, which is directly applicable to setting targets for Time in Ranges and originated with Lawlor and Hornyak (2012). This approach includes five key behavioural change components relevant to goal setting: 1. The goal is specific and defines exactly what is to be achieved, 2. The goal is measurable, and there is tangible evidence when it has been achieved, 3. The goal is attainable but stretches the participant so that they feel challenged, 4. The goal is relevant to the participant's wellbeing and health, and 5. The goal should be attainable over a short period of time. The CGM data strengthens the measurability of this SMART goal. As a biofeedback tool, it potentially allows the participant to identify specific foods that elevate their glucose level beyond the recommended target range of 3.9 - 8 mmol/L.

1.3. Contribution of this study

Behavioural economics is emerging as a key field in the modifying of self-destructive behaviour, such as lifestyle choices contributing to diabetes (John et al., 2011). This research contributes to a growing literature showing that insights from behavioural economics and psychology can help people engage in behaviours that are consistent with their long-term wellbeing. Providing outpatients with education on nutritional ketosis and CGM technology for a limited period of 1 month (alone or combined with personalised online health coaching) is yet to be explored as a relatively cheap, scalable behavioural model to improve glucose time in target range (3.9 - 8 mmol/L). Additional complexity is introduced by technology adoption in a developing country setting where education and health literacy may be limited. The experimental design allows us to rigorously evaluate the impact of real-time feedback from CGM with and without the continuous care component of health coaching. Our protocol of technology-assisted behavioural interventions for T2DM reversal could be built into a mobile app in order to bring it to scale. We contribute to the still controversial literature on nutritional ketosis for the management and reversal of T2DM by providing the first evidence from a developing country on a nutritional ketosis dietary programme with (1) structured fingerprick self-monitoring and intermittent CGM feedback versus (2) real-time CGM feedback alone or (3) real-time CGM feedback as part of a comprehensive CCI intervention with health coaching. Unpacking some components from the black box of a behavioural intervention allows us to draw conclusions about the minimal effective treatment necessary for reversing T2DM.

A second contribution of the study is to address the gap in the literature on experiments that evaluate the utility of CGM as a tool for doctors to advise a ketogenic diet with the aim to reverse diabetes. The extent to which intermittent and real-time glucose feedback shapes the food choices of patients with T2DM is of critical interest. We track dietary intake at baseline and at the 3-month follow-up using the food frequency questionnaire. We also compared the composition of dietary intake between CONTROL, INFO and COACH. Thirdly, this study contributes to the active behavioural economics literature on habit formation. In particular, it considers an approach to solving problems of nutrition through interventions informed by behavioural insights (e.g. List and Samek, 2015; Charness et al., 2020). What measures of support are required in combination with ketogenic nutrition education to help patients to implement physician-prescribed dietary guidelines at 3 months, time in target range at 70% or more and HbA1c below 6.5%?

The results show that all three treatment groups achieve a significant reduction in HbA1c on average at 3-months (endline). All three groups achieve the goal of 70% time in target range at 1-month (mid-line). Contrary to our hypothesis, the real-time CGM group (INFO) does not have better glucose control than CONTROL while wearing the CGM for one month, as shown in CGM patterns. There is no significant difference between groups in average glucose levels after one month of wearing the CGM with or without real-time feedback. This suggests that the rich report of the CGM closes the information gap between patient and doctor, and offers an accountability mechanism. Moreover, when the effect of the CGM is allowed to wash out, we do not find evidence of further value added by real-time CGM feedback on endline HbA1c. The success of the coaching group is not due to scanning the real-time CGM more often than INFO. COACH scans the device significantly less often than INFO but is more able to use the supervised information to implement the diet.

Endline HbA1c is taken after a two-month sustainability period without CGM. The COACH group achieves the biggest reduction in HbA1c of 19%, and regression results indicate that COACH performs significantly better than CONTROL (14% reduction) and INFO (11% reduction). All treatment groups report a significant decrease in unhealthy food intake. At endline, there is no significant difference between treatment groups in reported weekly consumption of unhealthy RED and healthy GREEN list foods. We categorise foods according to the Noakes' Foundation's traffic lights lists (see Appendix). In the regressions, lower RED and higher GREEN consumption at the endline is significantly associated with better glucose control. The rate of reversal is 43% at 3-months in the pooled sample, including CONTROL, INFO, and COACH. This is comparable to the 55% reversal rate at 3 and 6 months from a non-randomised continuous care intervention study including coaching and nutritional ketosis in the United States (Athinarayanan et al., 2019). Our study differs because we have a sustainability period in which the effects could wash out. In addition, we randomise patients into different treatment groups (stratifying on low/medium/high HbA1c level and gender), and we conduct the study in a developing country setting. The blood test results speak to the efficacy of our more conservative interventions. The rest of the paper proceeds as follows: Section 2 details the methods, Section 3 presents the results, and Section 4 concludes.

2. METHODS

2.1. Experimental Design

This study uses two different CGM sensors: one that shares real-time feedback with the wearer through a mobile app (FreeStyle Libre) and one that requires the glucose data to be downloaded by a healthcare practitioner in order to be viewed by the patient (Libre Pro). We use objective health assessments and self-report questionnaires. The study team uses the LibreLinkUp app to remotely monitor participants' scanning activity. The FatSecret food diary app allows the participants to invite the study team to monitor their dietary intake remotely. Patients also have the option to use a paper diary to track their food choices. The sample size is 95 patients (excluding dropouts). To complete the study, participants wear three sensors consecutively over a period of six weeks (i.e. three times two weeks) and attend five study visits to the primary care practice as outlined below (Figure 1). All other contact happens remotely. Participants are compensated R200

(~13 USD) at each of five study visits and receive CGMs, practitioner consultations and blood tests as part of their participation in the study. They purchase their own fingerprick test strips.

Eligible participants are men or women aged 30-65 years old, with a T2DM diagnosis from a doctor, HbA1c 6.5 - 11%, non-insulin-dependent, with no previous experience using Abbott's FreeStyle Libre, a current patient of Kenilworth Medicross Medical Centre or a referral from the Cape Town catchment area, able to read and speak English, use a smartphone, able to attend the five required study visits, wear the three CGM sensors and keep a food diary.[50] A waiting room notice was circulated (Appendix A).

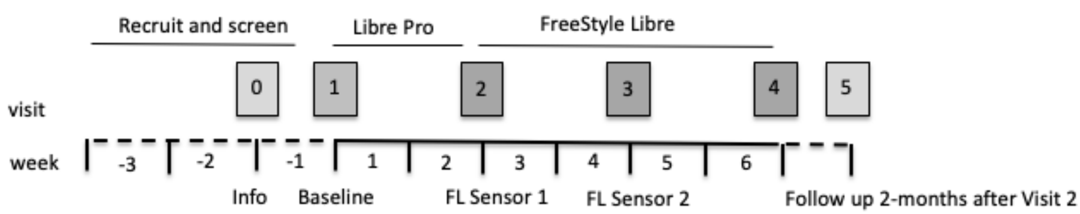
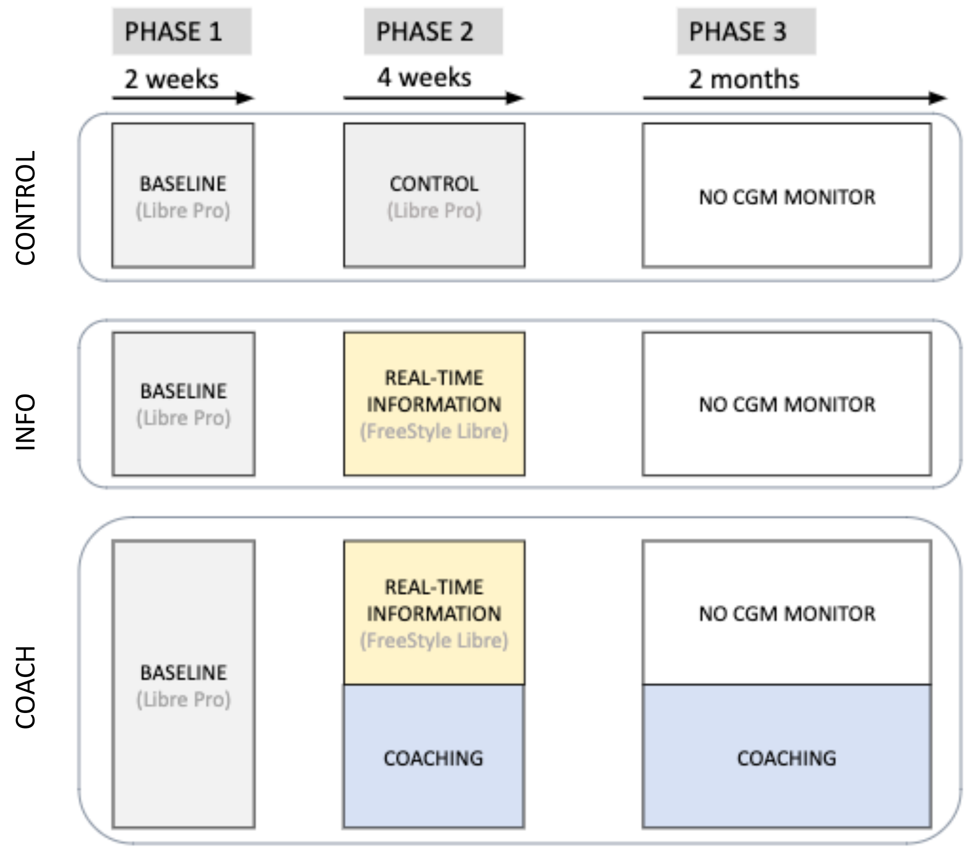


FIGURE 1

Study timeline.

There are three distinct phases, detailed below. Briefly, Phase 1 includes the informed consent and baseline procedures, as well as assignment to one of the three treatment groups by a person independent from the research team, stratifying on gender and HbA1c level (low, medium, high). Phase 1 is the baseline period, then Phase 2 concerns Visit 2, 3, and 4 and is the major intervention phase, while Phase 3 is the follow-up and sustainability assessment (Figure 2). Refer to the Appendix for the informed consent procedure, baseline questionnaire, self-report dietary intake, health-coaching agreement and medical records release form, nutritional education materials.

[50] Gestational diabetes is beyond the scope of the study. Exclusion criteria include anaemia, unmanaged hypothyroidism, disordered eating (e.g. anorexia and bulimia), renal disease (e.g. poor kidney function or kidney stones) and lactation.



Notes. This figure shows the course of the experiment and stepwise treatment groups. All participants complete a baseline period of 2 weeks, during which time their bloodwork is taken, and they are assigned to one of three treatment groups. From top to bottom, CONTROL, INFO, and COACH treatment groups are shown. The difference between CONTROL and INFO is real-time monitoring. The difference between INFO and COACH is remote one-to-one health coaching sessions. All patients see the study doctor intermittently for nutrition education and medical advice.

FIGURE 2

Experimental design: CONTROL, INFO, and COACH treatment groups

2.1.1. Phase 1 (2-weeks) Baseline

At Visit 1, a researcher guides the participant through the informed consent procedure (Appendix B). After Visit 1, the patient completes the baseline questionnaire (Appendix C), Food Frequency Questionnaire (Appendix D) and Covid-19 Questionnaire (Appendix E) online or via phone interview. At Visit 1, the patient is seen by the nurse who applies a Libre Pro sensor to be worn until Visit 2 (14 days if no device failure). Participants are instructed to continue their usual habits during the baseline. The researcher instructs them on how to use the FatSecret food diary app or paper diary for a minimum of four days and as much as they like. Once baseline blood tests are received, patients are randomly assigned on a rolling basis to one of the three treatment groups by an independent person not part of the study team. The algorithm stratifies on low/medium/high HbA1c level and gender. The randomisation code is available upon request. It is a single-blinded study. The physicians have no influence over treatment assignment but know the treatment group of each of his or her patients.

2.1.2. Phase 2 (4-weeks) Intervention: Visits 2-4.

At Visit 2, the nurse removes the Libre Pro. The physician reviews the results with the patient and delivers dietary advice for the patient to follow for the next three months (see Appendix F – Doctor's Low Carb Recommendations and Appendix G – Traffic lights lists of foods). The physician explains the goal to bring HbA1c to 6.4% or lower by following prescribed dietary advice and a structured self-monitoring regime. CONTROL: The second Libre Pro sensor is applied. All patients in booster treatments INFO and COACH use the FreeStyle Libre system for 14 days. Patients in COACH sign a coaching agreement (Appendix H) and receive an initial 50-minute online session, followed by weekly 30-minute online appointments with a certified functional health coach for 4-weeks and then only every 2-weeks for a total of 8 sessions over 12 weeks. In order to keep frequency of contact comparable, the CONTROL and INFO groups receive contact from a member of the study team at similar intervals but no health coaching. At Visit 3, CONTROL: the second Libre Pro sensor is removed, and a new one applied. INFO and COACH: the FreeStyle Libre is removed, and a new one is applied. The patient has a second 30-minute consultation with a study physician who reviews data and progress and helps fine-tune food choices. At Visit 4, CONTROL: the third Libre Pro is removed; INFO and COACH: the second FreeStyle Libre sensor is removed. The physician reviews progress and adjusts medications as needed. The next follow up (Visit 5) is booked for 2 months later to test HbA1c. During this time, the patient is instructed to continue with progress towards the goal of diabetes reversal by following prescribed advice. This allows for the effects of the CGM to wash out since HbA1c reflects average glucose (and dietary behaviour) of the past two months.

2.1.3. Phase 3 (2 months) Sustainability

Patients in CONTROL and INFO are unsupervised and no longer wear the CGM. They can continue to use the food diary and fingerprick testing at their discretion. Patients in COACH continued with the coaching intervention in which they now receive a 30-minute online appointment with their health coach every two weeks instead of weekly. At Visit 5, the patient completes the final study visit to get their HbA1c tested. They also see the physician for a final consultation. The patient completes a FFQ about the last month and an endline questionnaire (Appendix I) as preparation. This sustainability period allows us to evaluate eating habits after the RT CGM is removed. Further details about the study protocol, including survey instruments, are available in the Appendix.

2.1.4. Health coaching: Overview of framework

Prospect stage. This stage commences once a participant falls into the COACHING group during the two-week baseline period. Before the first session, the coach makes contact with the client to connect, explaining the coaching procedure and designing their coaching relationship. The coaching programme is over a 3-month period which includes weekly sessions for 4 weeks, followed by biweekly sessions for an additional 2 months for a total of eight sessions. The first session is 50 minutes, and each following session is 30 minutes. Consent for coaching is verbally given, and the coach follows up the session with electronic copies of the coaching agreement.

(Week 1): Intervention commencement: Session 1: Wellness vision. In this session, the coach explores a wellness vision for the client in which they have reversed their type 2 diabetes through lifestyle change. This wellness vision is the basis of the three-month arc of the coaching relationship. Through this vision, long-term goals are designed that move the client towards that vision.

(Weeks 2-11): Ongoing coaching intervention. Each coaching session that proceeds after session 1 follows the Adapt Model of Health Coaching (Appendix J). Each session is structured around a framework of engaging and opening the conversation, agenda-setting/focus, evoking, pausing before planning, planning, and ends with closing the session.

(Week 12): Closing the contract. In this coaching session, the coach and coachee review the progress that they have made with respect to their initial vision and plan, capture insights and learnings, celebrate success and create a maintenance plan.

2.1.5. Baseline questionnaire (40 minutes, online)

Our baseline questionnaire (Appendix C) includes sections from validated questionnaires. Prevalence and severity of diabetes is strongly associated with socioeconomic and cultural factors, such as income, education, gender, ethnicity and culture. The WHO STEPwise approach to chronic disease risk factor surveillance (STEPS) questionnaire: A few common and preventable risk factors are strongly associated with most non-communicable diseases (NCDs), including type 2 diabetes. This questionnaire has been validated by WHO and includes information on education, marital status, employment, housing density, income, disease history, medication, as well as leading behavioural risk factors for NCDs e.g. tobacco use, harmful alcohol consumption, and physical inactivity.

Standard behavioural economics measures of risk and time preferences from the validated Global Preferences Survey module are used (Falk, Becker, Dohmen, Huffman, & Sunde, 2016; Falk et al., 2018). Subjective wellbeing is measured using a single standard item and scored on a scale of 1 to 10 where 1 means “very dissatisfied” and 10 means “very satisfied”: “How do you feel about your life as a whole right now?”. The wording of this standard measure of subjective wellbeing is sourced from the South African National Income Dynamics Study (NIDS). This measure may be affected by social distancing and economic hardship due to lockdown restrictions. In a separate questionnaire we address the impact of COVID-19 on diabetics.

We use the Diabetes Distress Scale (DDS) for adults with type 2 diabetes. The DDS is a 17-item self-report instrument. Each item is rated on a 6-point scale from (1) “not a problem” to (6) “a very significant problem”. The scale yields an overall distress score based on the average responses on the 1-6 scale for all 17 items. It can also be used to get a measure of emotional burden (average of 5 items), physician distress (average of 4 items), regimen distress (average of 5 items) and interpersonal distress (average of 3 items). Average score < 2.0 reflects little or no distress. Average score 2.0 – 2.9 reflects moderate distress and > 3.0 reflects high distress. A total or subscale score > 2.0 is considered clinically significant. Lastly, we assess patients' confidence to make a lifestyle change, the importance to them of making a lifestyle change, and their stage of readiness.

2.1.6. COVID-19 and type 2 diabetes questionnaire (10 minutes, online)

We survey contact with COVID-19, the impact of the pandemic on access to medical care (the impact of the COVID-19 pandemic on productivity and lifestyle, and related challenges and fears. See Appendix E for the full questionnaire.

2.1.7. Endline questionnaire (15 minutes, online)

In the endline questionnaire (Appendix I) with all participants we measure self-report changes in habits, subjective wellbeing, Diabetes Distress Scale (DDS) and invite participants to give feedback on their experience.

2.2. Research questions

Each of our research questions identifies a potential barrier to T2DM reversal and tests whether the proposed technology-assisted behavioural intervention differentially improves the outcomes of glucose control and dietary intake compared to the control group, which is a limited feedback intervention with nutrition education. All groups receive the same dietary advice.

CONTROL GROUP (Intermittent CGM feedback for one month): Compared to usual habits (baseline), is low-carbohydrate education from a physician with three intermittent checkups sufficient for patients to achieve: (a) Time in target glucose range at 70% or more? (b) HbA1c at 6.4% or lower at 3-months? (c) Compliance with physician's nutrition education on Low-Carb High-Fat (Ketogenic, 25g of carbs per day) at 3-month endline (*i.e.* avoid RED list foods and eat freely from GREEN list foods). We use the intermittent CGM feedback intervention group to isolate the minimal medical intervention of once-off low-carb nutrition education from a physician at a 40-minute appointment with two follow-up checkups of 30 minutes to review progress and adjust medications if needed. Patients are instructed on a self-monitoring programme to fingerprick test just before and then 1.5 hours after meals and write down the results, as well as the meal associated with the glucose spike. They wear a CGM that measures glucose continuously but provides no feedback to the patient until the doctor reviews the downloaded data with them at each of the consultations.

INFO TREATMENT (Real-time CGM feedback for one month): Compared to the control group, does the addition of real-time self-monitoring for four weeks promote better: (a) Time in target glucose range at 70% or more? (b) HbA1c at 6.4% or lower at 3-months? (c) Compliance with physician's Low-Carb High-Fat (Ketogenic, 25g of carbs per day) dietary advice at 3 months. Nutrition education and fingerprick glucose testing may not be sufficient to improve glucose control if the patient has insufficient personalised data to fine-tune their diet. We addressed this information barrier by providing the patient real-time access to their glucose patterns for four weeks via a mobile app. This treatment tests the role of real-time self-monitoring as a limited short-term intervention.

COACH TREATMENT (Real-time CGM feedback for one month and remote continuous care): Compared to the real-time feedback treatment, does the addition of health coaching promote better: (a) Time in target glucose range at 70% or more? (b) HbA1c at 6.4% or lower at 3 months? (c) Compliance with physician's Low-Carb High-Fat (Ketogenic, 25g of carbs per day) dietary advice at 3 months. Nutrition education and real-time self-monitoring may not be sufficient to elicit behaviour change and reverse diabetes. Present bias (impatience or an inability to delay gratification) is a cognitive barrier to achieving glucose control and building sustainable healthy habits. In addition to real-time self-monitoring, dedicating time to identify the goal that is important to the patient, visualising what it would mean to achieve it, and then identifying obstacles that stand in the way and how to overcome them (if/then plans), may be necessary (in addition to self-monitoring) to achieve behaviour change. This is described further below. Health coaching connects short-term goals and rewards to long-term wellbeing through interactive, personalised coaching. We test a comprehensive continuous care model that incorporates physician-prescribed nutritional ketosis advice, real-time CGM glucose feedback and online health coaching.

2.3. Hypotheses

The following pre-registered hypotheses state expected treatment effects on behaviour for each of the primary outcome measures of glucose control and secondary outcome measures of dietary intake (AEA RCT Registry: AEARCTR-0005871).

HbA1c. Expectations for HbA1c blood test are as follows: COACH < INFO < CONTROL < BASELINE. All participants begin with HbA1c 6.5-11%. Within subjects, we expect HbA1c to improve when carbohydrate intake is restricted. Our aim is for patients to achieve HbA1c 6.4% or below (*i.e.* diabetes reversal) at a 3-month follow-up. An HbA1c of up to 5.5% is normal, while 5.6-6.4% is pre-diabetic.

Dietary intake. Participants in the COACH treatment are expected to report greater carbohydrate restriction than INFO and CONTROL treatment groups, respectively, even though they received the same nutrition education. Within subjects, reported dietary intake is expected to favour low-carbohydrate foods after the intervention period relative to BASELINE. Patients' nutrition education is delivered through traffic lights lists that they take home with them. These are available in the Appendix and the same as those used in Chapter 3. We expect to see changes in diet reflected in a reported reduction in RED list foods and an increase in GREEN list foods.

Time in Target. Expectations for Time in Target range are as follows: COACH > INFO > CONTROL > BASELINE. Time in Target is expected to be higher with access to FreeStyle Libre feedback compared to CGM with intermittent feedback every two weeks (Libre Pro). Based on worldwide patterns, we expect the INFO group to achieve at least 60% time in target, or about 14.4 hours/day (Dunn et al., 2018).

Engagement. Participants in the COACH treatment are expected to scan more than the INFO treatment. Storage of glucose values by the sensor is automatic as long as the patient scans the FreeStyle Libre at least once every 8 hours. Patients are advised, "the more scans, the better". If gaps in CGM data are apparent, this advice is reiterated by the nurse.

3. RESULTS

3.1. Sample characteristics

Baseline Questionnaire. Table 1, Panel A shows that the sample is balanced on demographic characteristics measured in the baseline questionnaire, including ethnicity, age, gender, marital status, number of household members, who is responsible for grocery shopping and cooking in the household, employment status, and income category. Panel B shows the sample is balanced on health risks, such as smoking, high blood pressure and cardiovascular events. This indicates the stratified randomisation worked. The majority of our participants describe themselves as Cape Coloured, which reflects the majority population group of Cape Town[51]. The mean age is 50 years old and just under half of the sample are female. Annual household income is reported using eight categories. There is no significant difference in income category across the treatment groups (K-Wallis, p=0.7120), and the average for the whole sample is 2.92, which is below the median income category range [0,7] and implies that, on average, annual household income of patients is R202,500 – R412,000 (~13,200 – 27,000 USD) or less.

Diabetes has only caused one participant to lose their job/retire early. About two-thirds of the sample have been diagnosed with high blood pressure. The three groups are similar in terms of medications and fingerprick testing frequency. Fingerprick testing using a blood glucose meter at home is the most common measurement of glucose. Subjective wellbeing is balanced across groups. In the Global Preferences Survey question, "In general, how willing are you to take risks?", the coaching group is slightly more risk-averse at the 10% significance level. In the Short-Grit Scale, the coaching group is slightly less gritty, significant at the 1% level. The coaching group is also somewhat less confident about their ability to make a change than the other groups, significant at the 5% level. In the regression analysis below, we control for individual characteristics. About half of participants (52.6%) consulted with Dr Neville Wellington and the other half with Dr Carol Bosch, with no significant differences across groups. Assignment of a patient to a doctor is taken with consideration of the availability of the doctors' and patients' schedules. All participants saw

[51] The 2011 Census showed that majority population groups in Cape Town are Coloured (42%) and Black African (39%).

the same diabetes nurse educator. Four patients failed to come in for follow-up blood tests after the intervention, however, there was no significant difference in dropout across groups.

TABLE 1

Sample characteristics (Panel A)

Treatment Group	All (N=95)	CONTROL (n1=32)	INFO (n2=32)	COACH (n3=31)	K-Wallis test p-value
Ethnicity (freq.)					
Coloured	61	22	22	17	0.4189
White	20	4	7	9	0.2751
Indian	6	2	1	3	0.5681
Black	3	0	2	1	
Asian	1	0	0	1	
Other	2	2	0	0	
Prefer not to answer	2	2	0	0	
Age (years)	50	50	53	47	0.0966*
Female (%)	44.2	43.8	43.8	45.2	0.9917
Married (%)	71.6	71.9	75.0	67.7	0.8165
HH members (freq.)	3.7	3.7	3.8	3.6	0.9139
Buys HH groceries (%)	61.1	56.3	68.8	58.1	0.5457
HH cook (%)	48.4	40.6	53.1	51.6	0.5554
Employed (%)	80.0	81.3	81.3	77.4	0.9096
Annual HH income (%)					
0. Up to R20,500	7.87	3.45	16.13	3.45	
1. More than R20,500 but less than R89,000	16.85	10.34	9.68	31.03	
2. More than R89,000 but less than R202,500	10.11	17.24	3.23	10.34	
3. More than R202,500 but less than R412,000	28.09	31.03	35.48	17.24	
4. More than R412,000 but less than R707,000	17.98	17.24	16.13	20.69	
5. More than R707,000 but less than R1,512,000	16.85	17.24	19.35	13.79	
6. More than R1,512,000 but less than R2,414,000	1.12	0.00	0.00	3.45	
7. More than R2,414,000	1.12	3.45	0.00	0.00	

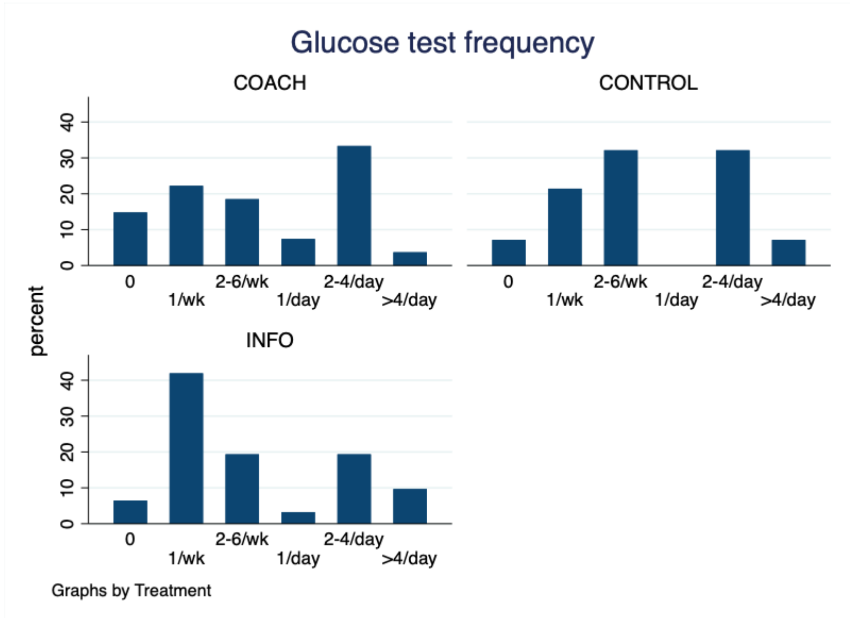
TABLE 1

Sample characteristics (Panel B)

Treatment Group	All (N=95)	CONTROL (n1=32)	INFO (n2=32)	COACH (n3=31)	K-Wallis test p-value
Smoke daily (%)	21.1	31.3	12.5	19.4	0.1802
Alcohol ever (%)	66.3	65.6	65.6	67.7	0.9795
High BP ever (%)	67.4	75.0	59.4	67.7	0.4146
High BP told past yr (%)	82.8	79.2	84.2	85.7	0.8317
BP meds past 2 wks (%)	79.7	79.2	73.7	85.7	0.6427
Heart attack ever (%)	6.3	6.3	6.3	6.5	0.9993
Aspirin (%)	17.9	25.0	18.8	9.8	0.2846
Statins (%)	21.1	21.9	15.6	25.8	0.6092
Years since diabetes diagnosis (yrs)	5.5	6.0	5.0	5.7	0.6975
Less than 1 year since diagnosis (%)	25.3	25.0	37.5	12.9	0.0823*
Testing using fingerprick (%)	72.6	71.9	75.0	71.0	0.9318
Willingness to take risks (1-10)	6.1	6.5	6.4	5.5	0.0861*
Patience (1-10)	7.9	7.9	8.1	7.8	0.8221
Good at math (1-5)	3.0	2.9	2.9	3.1	0.6100
Procrastinate (1-10)	5.2	5.9	4.7	5.1	0.1203
Subjective Wellbeing (1-5)	3.2	3.4	3.1	3.0	0.3480
S-Grit Score (1-5)	3.4	3.5	3.6	3.1	0.0058***
Confidence to change (1-10)	8.0	8.0	8.5	7.4	0.0417**
Importance of change (1-10)	9.4	9.4	9.3	9.5	0.6773
Study doctor Dr W (%)	52.6	62.5	50.0	50.5	0.3658
Dropped out (no follow up bloodwork)	4	1	1	2	0.7532

Notes. Kruskal-Wallis H tests are reported as a balance check between the three treatment groups. Significant differences are starred. * shows p-value < 0.1, ** shows p-value < 0.05, *** shows p-value <0.01. There is no significant difference in mean income category across the treatment groups (p=0.7120). In Panel B, for the eight survey scales, 1 = not willing to take risks, not patient, not good at math, do not procrastinate, low on subjective wellbeing, low on grit, poor confidence to change and low importance of making a change, respectively; while 10 = a high score on the scale.

Figure 3 shows blood glucose testing frequency reported in the baseline questionnaire. Patients typically tested a couple times per week up to a couple times per day. A majority tested less than once per day. A Kruskal-Wallis test for equality of populations finds no significant differences across treatment groups in baseline glucose testing frequency. Baseline testing behaviour is important to describe since all intervention groups receive structured self-monitoring programmes that encourage them to test before and 1.5 hours after a meal to learn what spikes sugar levels. Based on this observed low testing frequency, we would expect a positive impact of increased testing during the intervention on glucose control.



Notes. This figure shows bar charts of percent of frequency of categories arranged by treatment group. Glucose testing frequency at baseline was reported by patients in the baseline questionnaire. K-Wallis test by treatment finds no significant difference in testing frequency across the three treatment groups at the 5% level (p=0.6212).

FIGURE 3

Glucose testing frequency by treatment group

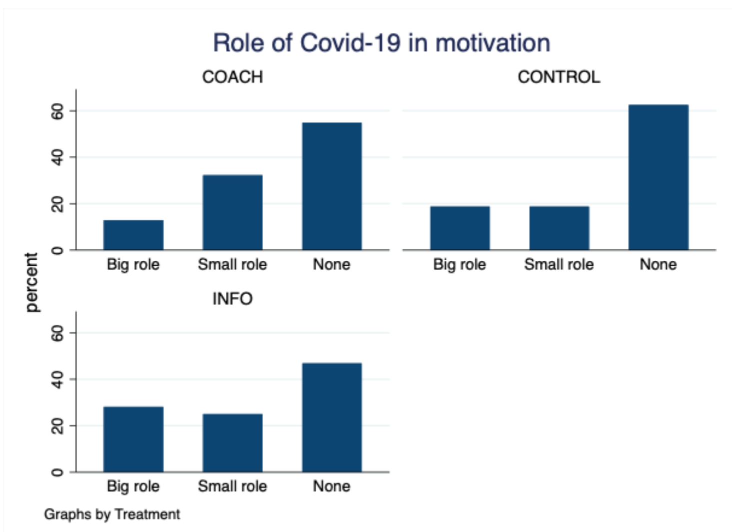
Covid-19 Questionnaire. Table 2 reports on the context of the study during a Covid-19 lockdown. This is important to describe, given that diabetics are notably an at-risk population. At baseline, just under a quarter of the sample have already been diagnosed with Covid and recovered. Fear of Covid has prevented 21% of the sample from seeking non-diabetic medical care, e.g. visiting the dentist. 18% of the sample are being treated for anxiety or depression. Most patients do not live alone, 58% are able to work from home, and 30% have a job classified as an essential service during the hard lockdown, enabling them to continue leaving home for work. The INFO group has received more remote consultations than the other two groups. Figure 4 shows that for about half of the sample, Covid-19 fear plays a role in the personal motivation of the sample of patients to get healthier and so reduce their risk of poor outcomes from the disease. Figure 5 shows that about a third of the sample have been making more unhealthy food choices since the Covid-19 lockdown. Figure 6 shows that physical activity levels have also declined for about half of the sample since the Covid-19 lockdown.

TABLE 2

Influence of Covid-19 lockdown and anxiety on managing diabetes

Treatment Group	All (N=95)	CONTROL (n1=32)	INFO (n2=32)	COACH (n3=31)	K-Wallis test p-value
I had Covid-19 (%)	22.1	25.0	25.0	16.1	0.6236
Someone I know has been diagnosed (%)	83.2	87.5	87.5	74.2	0.2708
Someone I know has died of Covid (%)	63.0	59.3	66.7	63.3	0.8390
Had contact positive person past 2 wks (%)	3.1	9.4	0	0	0.0489**
Had symptoms in past 2 weeks (%)	4.2	3.1	3.1	6.5	0.7532
Covid fear ever prevented seeking diabetes care (%)	9.4	6.3	6.3	16.1	0.3086
Covid fear ever prevented seeking other med. care (%)	21.1	18.8	18.8	25.8	0.7338
Currently being treated for anxiety or depression (%)	17.9	18.8	18.8	16.1	0.9528
Had any remote consultations (%)	36.8	28.1	56.3	25.8	0.0206**
Covid fear prevented picking up meds (%)	8.4	3.1	6.3	16.1	0.1566
I live alone (%)	7.4	6.3	9.4	6.5	0.9715
I can work from home (%)	57.9	50.0	56.3	67.7	0.4700
Job classified as essential service (%)	29.5	25.0	37.5	25.8	0.4760
Eaten at restaurant past two weeks (%)	59.6	65.6	54.8	58.1	0.6717

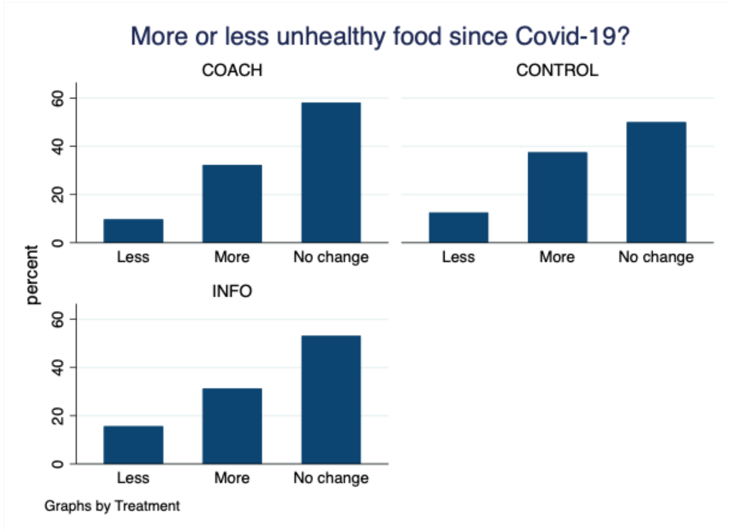
Notes. Kruskal-Wallis H tests are reported as a balance check between the three treatment groups. Significant differences are starred. * shows p-value < 0.1, ** shows p-value < 0.05, *** shows p-value <0.01.



Notes. This figure shows bar charts of percent of frequency of categories, by treatment group. Source of data is online questionnaire taken at baseline period.

FIGURE 4

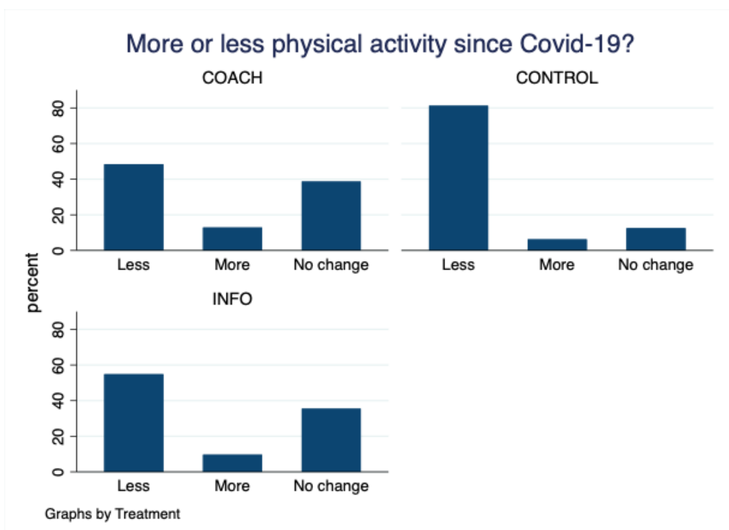
The role of Covid-19 in personal motivation to improve glucose control



Notes. This figure shows bar charts of percent of frequency of categories, by treatment group. Source of data is online questionnaire taken at baseline period.

FIGURE 5

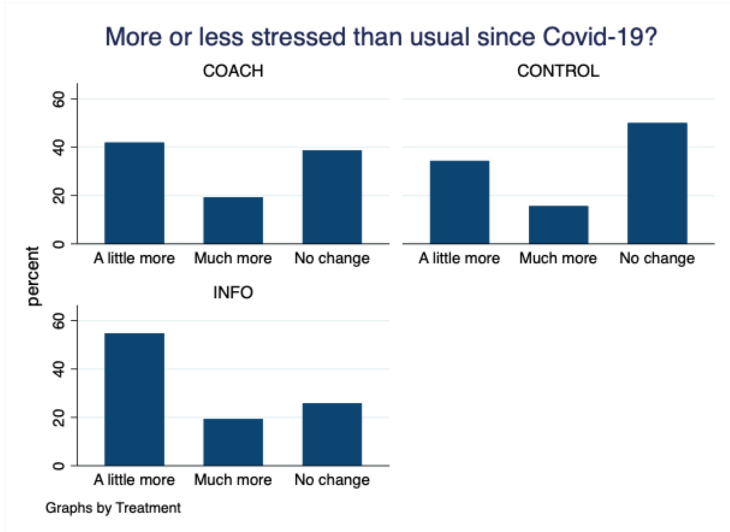
Changes in unhealthy food consumption since Covid-19 lockdown



Notes. This figure shows bar charts of percent of frequency of categories, by treatment group. Source of data is online questionnaire taken at baseline period.

FIGURE 6

Change in physical activity since Covid-19 lockdown



Notes. This figure shows bar charts of percent of frequency of categories, by treatment group. Source of data is online questionnaire taken at baseline period.

FIGURE 7

Change in stress levels compared to usual since Covid-19

3.2. Reported dietary intake pre/post intervention

Figure 8 shows food intake from the unhealthy RED, ORANGE, and healthy GREEN list pre- and post-intervention at the time of the blood tests. The gold bars show food intake before the intervention, and the purple bars show the distribution of food intake after the intervention during the sustainability period (endline). There is no significant difference between the treatment groups at baseline in consumption of RED list foods (K-Wallis, p=0.9937). All three treatment groups see a significant decrease in RED list food intake, consistent with the dietary advice they received (K-Wallis, p<0.001). This is aligned with our hypothesis that foods high in carbohydrates and processed vegetable oils would decrease. At baseline, there is no difference in intake of portions from the GREEN list (K-Wallis, p=0.3564). The CONTROL and INFO groups increased their reported GREEN food intake, significant at the 10% level (p=0.0685, p=0.0857), and there is no significant change in reported GREEN intake for COACH (p=0.3618).

Notes. This figure shows the results of the food frequency questionnaire at baseline (gold bars) and endline (purple bars) and where the two distributions overlap (darker purple bars). The histograms show total weekly intake of RED, ORANGE, and GREEN list foods (columns) by treatment group CONTROL, INFO, and COACH (rows). RED = unhealthy, GREEN = healthy.

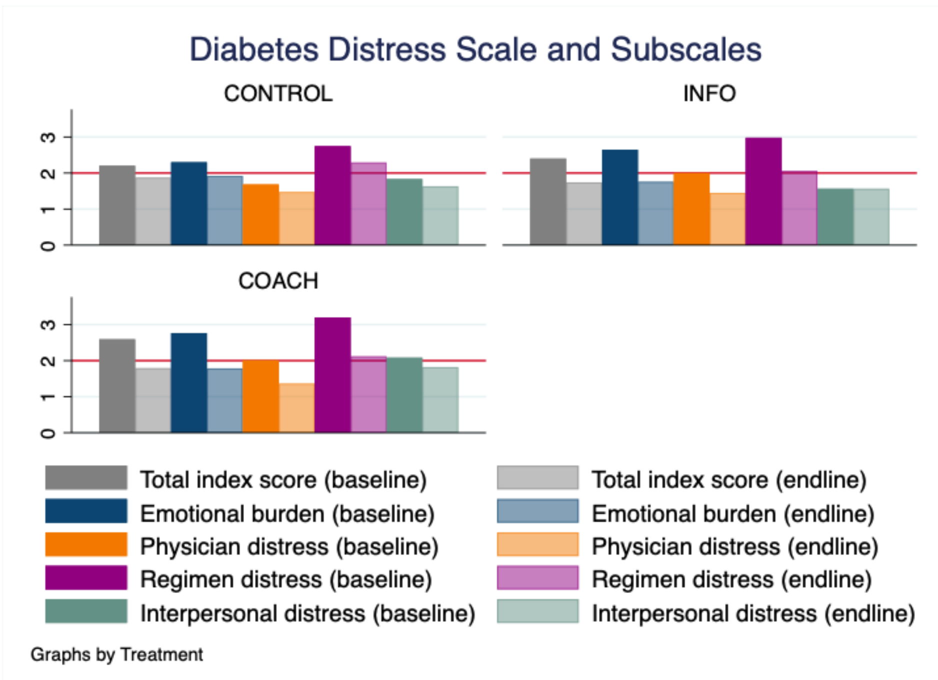
FIGURE 8

Histogram showing portions per week from the various traffic lights food lists at the baseline and endline by treatment group

Result 1. *All treatment groups report a significant decrease in unhealthy RED food intake. At endline, there is no significant difference between treatment groups in reported weekly consumption of unhealthy RED and healthy GREEN list foods.*

3.3. Diabetes Distress Scale pre/post intervention

Figure 9 shows bar charts of the mean of the total scale and subscales for the Diabetes Distress Scale (DDS) across treatment groups, at baseline and endline, respectively. At baseline, the distress in the total DDS scale is greater than 2.0 and so considered clinically significant. The groups are very similar in terms of diabetes distress scores in the total scale (K-Wallis, p=0.5041). The subscale of most concern for all three groups at baseline was regimen distress, *i.e.* the feeling that they are failing by not managing their diabetes well, e.g. meal plan and exercise. For CONTROL and INFO, regimen distress is moderate (2.0-2.9) and for COACH, high (>3.0), but the difference between groups is not statistically significant (K-Wallis, p=0.3471). At endline for all three treatment groups, we observe that the total scale and three out of four subscales fall below 2.0, meaning little to no distress. All groups report moderate regimen distress, and this is slightly higher above the threshold of 2.0 in CONTROL, compared to the other treatment groups, but not statistically significant (K-Wallis, p=0.5429). Compared to the baseline, regimen distress decreased in all groups.



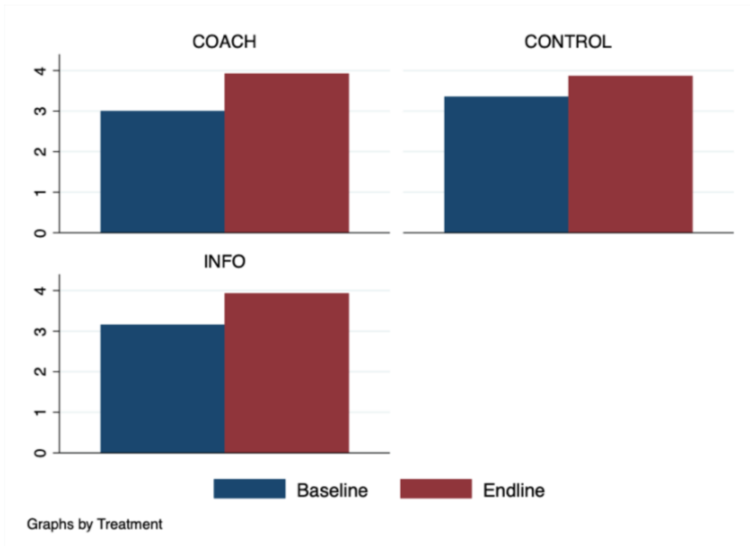
Notes. Bar charts show the mean of the total scale and subscales for the Diabetes Distress Scale across treatment groups at baseline and endline. The red line is a distress threshold of 2.0, which is considered clinically significant. <2.0 means little to no distress. K-Wallis test shows no significant difference between the three groups in Total index score at endline.

FIGURE 9

Diabetes Distress Scale and subscales at baseline and endline

3.4. Subjective Wellbeing pre/post intervention

Figure 10 shows bar graphs of the change in mean subjective wellbeing (SWB) measured at baseline and endline. The SWB scale is scored on a 5-point Likert scale. All three groups are similar at baseline and report an increase in SWB, on average. There is no significant difference in SWB between treatment groups at endline (K-Wallis, p=0.8062).



Notes. Bar charts show subjective wellbeing at baseline and endline across treatment groups. Subjective wellbeing is scored from 1 = not satisfied, to 5 = very satisfied.

FIGURE 10

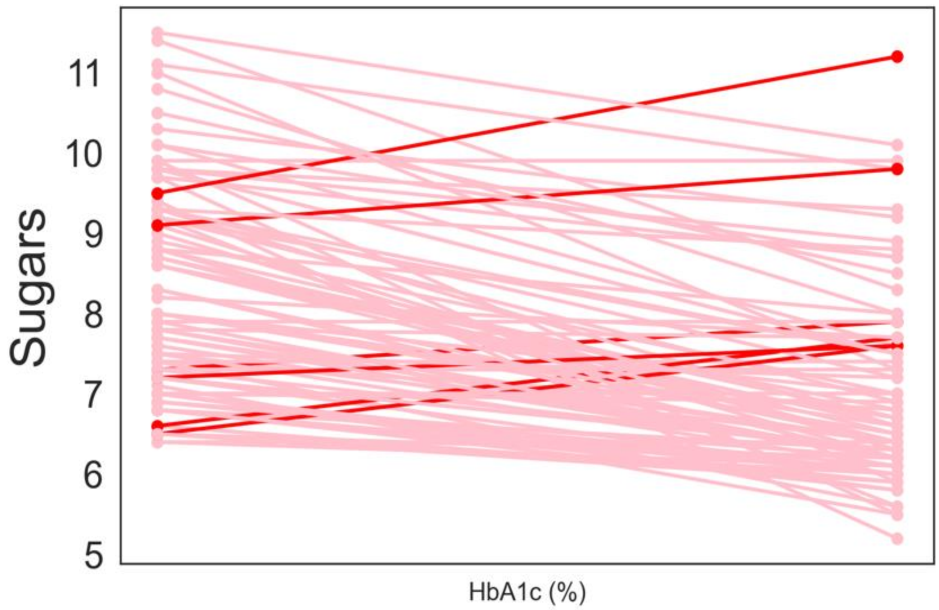
Subjective Wellbeing at baseline and endline

3.5. Glucose control: HbA1c blood test pre/post intervention

Figures 11-13 describe the blood test results pre/post-intervention. Figure 11 shows the change in HbA1c from baseline to endline for the pooled sample, which provides a within-subject comparison. Red lines indicate a minority of participants who do not manage to improve their HbA1c compared to their baseline level, while pink lines show the majority of participants improve. This figure provides the first evidence of the positive impact of the interventions on glucose control at 3 months, the primary outcome of the trial. Only six participants fail to improve HbA1c in the pooled sample. Next, in Figure 12, we show the bar graphs of the percentage change in all the blood markers we measured. In particular, total cholesterol, LDL, HDL, triglycerides, HbA1c and fasting insulin. Of interest is our pre-registered hypotheses, which include that HbA1c is expected to fall after the intervention compared to baseline. Contrary to our hypotheses, however, the INFO group does not see a greater reduction than CONTROL. The INFO group sees a decrease of 11% on average compared to the CONTROL's 14%. Consistent with our hypotheses, the COACHING group performed the best with a reduction of 19% in HbA1c, on average. In the regression analysis below, we test if the difference between groups is significant and robust to additional control variables. In the pooled sample, 42.71% achieved diabetes reversal (HbA1c < 6.5) in 3 months; and by treatment group CONTROL: 43.75%, INFO: 37.5% and COACH: 48.39% with no significant difference between groups (K-Wallis, p=0.6840).

In Figure 13, the box plots summarise the pre/post distributions of fasting insulin and HbA1c, respectively, by treatment group. There is no significant difference in the pre/post distributions for fasting insulin in all three groups. However, compared to baseline, all three treatment groups see a highly significant decrease in HbA1c in the direction expected. The COACH group achieves better glucose control more consistently, as the weight of the distribution is shifted lower towards 6.5% (diabetes threshold), and the median post-HbA1c is the lowest of all three groups.

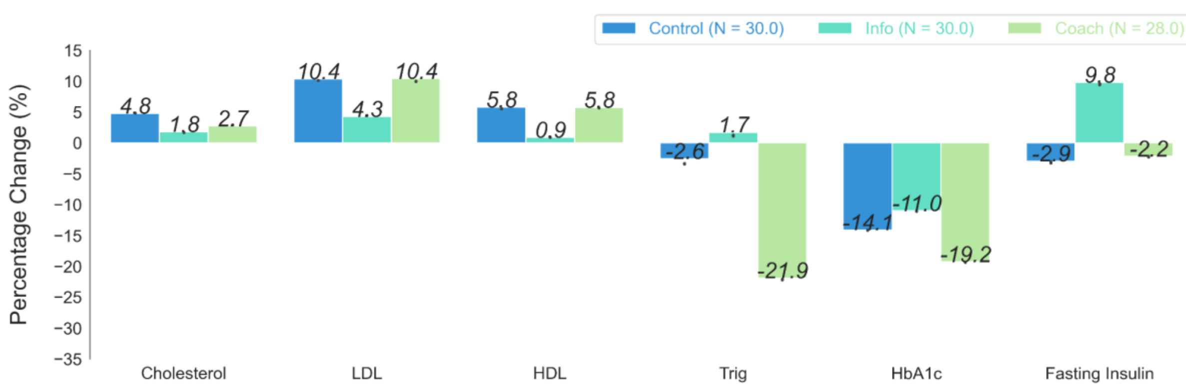
For completeness, we include the pre/post change in the lipid panel distributions by treatment group in Figure 14. The CONTROL has a significant increase in total cholesterol and LDL, and no difference in HDL and triglycerides compared to baseline. The INFO group has no significant differences in the lipid panel compared to baseline. The COACH group has no significant change in total cholesterol or LDL but sees a significant increase in HDL and a significant reduction in triglycerides. This is consistent with following a diet restricting carbohydrates and suggests that the COACH group was the most compliant with the physician's nutrition advice. A significant reduction in blood triglycerides and an increase in HDL is considered an improvement in metabolic health (Alberti et al., 2005). On these blood markers, COACH is significantly better than the other two groups.



Notes. This figure shows the absolute change in HbA1c (%) level from baseline (left) to endline (right) for individual patients. All patients had a common goal of bringing their HbA1c below 6.5% (*i.e.* reversing diabetes). Pink lines show patients that improved or stayed the same. Red lines show individual patients that worsened.

FIGURE 11

Absolute change in HbA1c level from baseline to endline (pooled sample)



Notes. Percentage change after 3 months in total cholesterol, LDL, HDL, triglycerides, HbA1c and fasting insulin is shown by treatment group, (1) CONTROL, (2) INFO, (3) COACH.

FIGURE 12

Percentage change in blood markers by treatment group

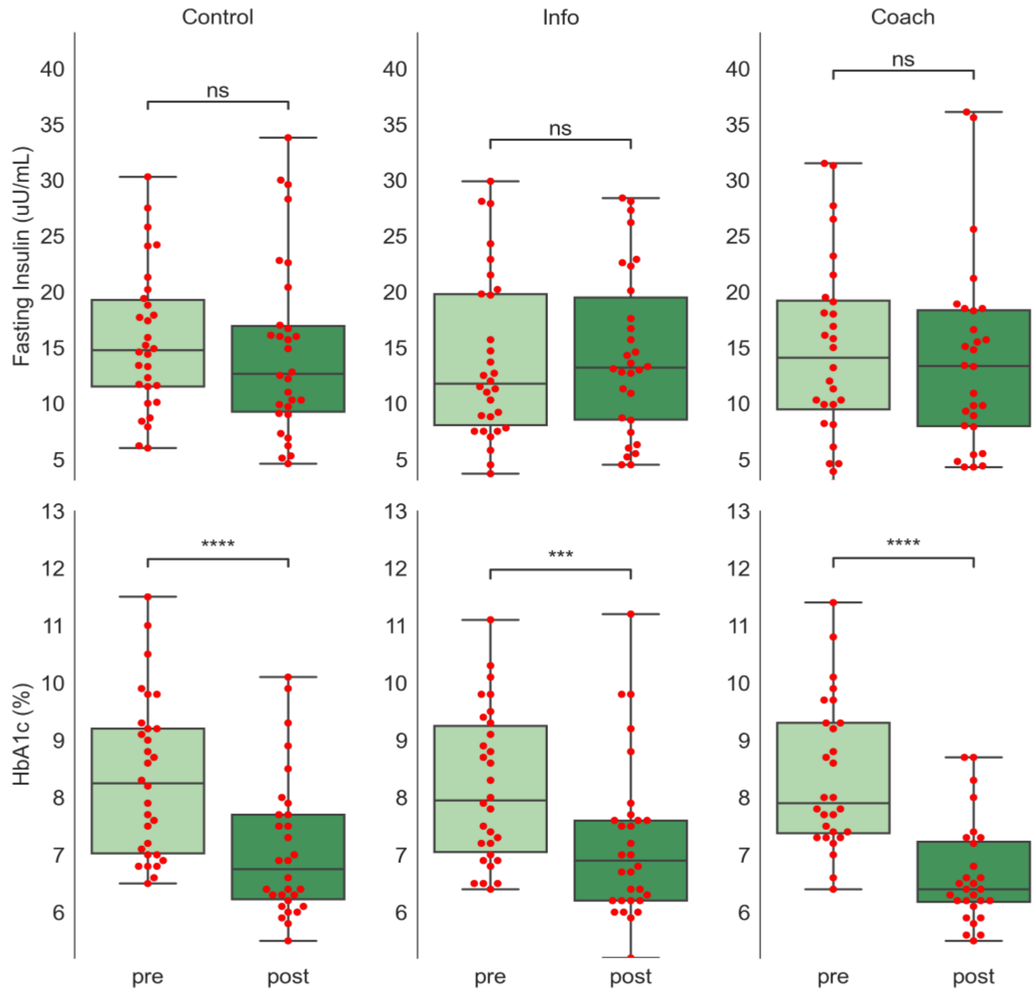
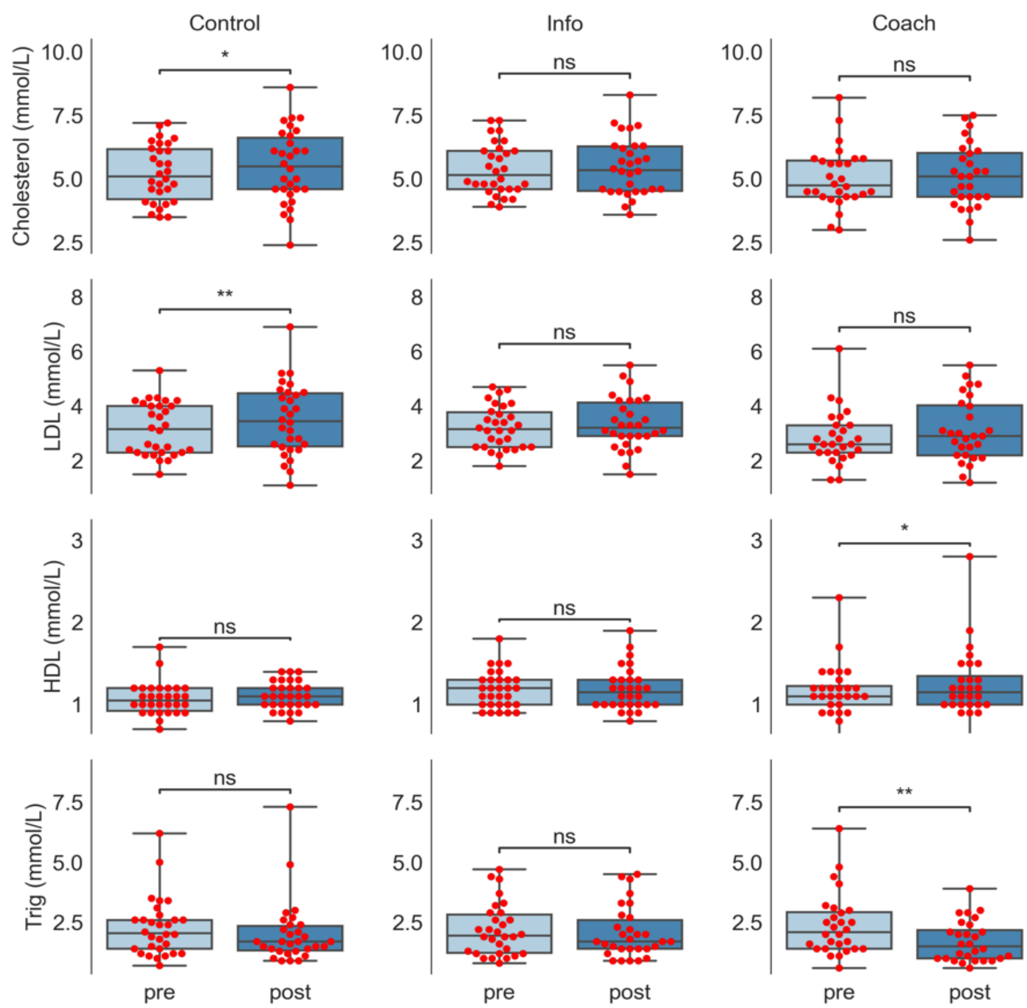


FIGURE 13

Box plots of fasting insulin and HbA1c by treatment group



Notes. This figure shows the change in total cholesterol, LDL, HDL and triglycerides pre/post intervention by treatment group CONTROL, INFO and COACH.

FIGURE 14

Box plots of lipid panel by treatment group

Table 3 shows the results of Ordinary Least Squares (OLS) regressions of the treatment effect of the INFO and COACH group on glucose control, compared to CONTROL. We report clustered standard errors at the individual level. Column (1) reports the basic model, which includes only the baseline HbA1c and treatment dummies. Column (2) adds controls for female, age, education and household income. Column (3) adds recent diagnosis (in the past year) and reported weekly consumption from the various traffic lights food lists given to patients during nutrition education by the physician. Column (4) adds the Global Preference Survey (GPS) measures. In particular, willingness to take risks in general, where 1 is "completely unwilling to take risks" and 10 is "very willing to take risks"; willingness to delay gratification in order to benefit more in the future, where 1 is "completely unwilling to do so" and 10 is very willing to do so; procrastination, rating the statement "I tend to postpone tasks, even if I know it would be better to do them straight away", where 1 is "does not describe me at all" and 10 is "describes me perfectly"; and finally the Short-Grit Score where 1 is "not gritty" and 5 is "very gritty". Finally, column (5) controls for baseline confidence to make a change to one's diet.

A higher baseline HbA1c is a consistently significant predictor of higher endline HbA1c. A 1 percentage point higher baseline HbA1c is associated with a 0.42 percentage point higher endline HbA1c, significant at the 1% level (column 5). The INFO treatment is not significantly different from CONTROL in columns (1)-(2) but becomes significant when we control for recent diagnosis and food choices in column (3) and remains significant in columns (4)-(5) with additional baseline controls. On average, patients in INFO ended with a 0.47 percentage point higher HbA1c than CONTROL, significant at the 5% level. This is in contrast with our pre-registered hypotheses. It means that real-time CGM patients in INFO do not do better on average than CONTROL after the sustainability period of two months without real-time monitoring (*i.e.* only fingerprick testing available). We infer that there is no additional benefit of real-time CGM compared to intermittent CGM after a sustainability period in which the effects are allowed to wash out. We speculate that the INFO group is less familiar with the before/after meals fingerprick testing since they are given costless access to their glucose levels during the main intervention period, and this is removed during the 2-month sustainability period. The CONTROL group is consistently fingerprick testing from the beginning of the intervention period and is only intermittently shown the CGM results at the doctor's appointments. During the sustainability period, they simply continue as before with their structured monitoring programme (before/after meals), fingerprick testing at their discretion.

On average, the COACH patients achieve 0.45 percentage points lower HbA1c than CONTROL at endline, significant at the 10% level. This is in comparison to the health coaching literature, which reports 0.32 percentage points. This is consistent with our pre-registered hypotheses. At endline, there is a greater reduction in HbA1c in COACH compared to groups with no health coaching. These patients also had the real-time CGM removed during the sustainability period but maintained the support and accountability of the health coach every two weeks after the main intervention period.

TABLE 3

Treatment effects on endline average glucose (HbA1c)

	Dependent variable: HbA1c endline				
	(1)	(2)	(3)	(4)	(5)
HbA1c baseline	0.54***	0.56***	0.41***	0.41***	0.42***
	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
INFO	0.18	0.21	0.50**	0.48**	0.47**
	(0.26)	(0.27)	(0.25)	(0.22)	(0.22)
COACH	-0.46**	-0.50*	-0.29	-0.42*	-0.45*
	(0.22)	(0.25)	(0.20)	(0.25)	(0.26)
Female		0.55**	0.54***	0.68***	0.66***
		(0.24)	(0.20)	(0.21)	(0.20)
Recent diagnosis (<1 year)			-0.42*	-0.33	-0.31
			(0.24)	(0.24)	(0.25)
Weekly RED intake (endline)			0.021***	0.018***	0.018**
			(0.01)	(0.01)	(0.01)
Weekly ORANGE intake (endline)			0.030	0.029	0.025
			(0.03)	(0.03)	(0.03)
Weekly GREEN intake (endline)			-0.011**	-0.0094**	-0.0092**
			(0.00)	(0.00)	(0.00)
GPS Willingness to take risks (baseline)				0.10**	0.10**
				(0.05)	(0.05)
GPS Patience (baseline)				-0.095*	-0.080
				(0.05)	(0.05)
GPS Procrastinate (baseline)				-0.071*	-0.072*
				(0.04)	(0.04)
S-Grit Scale (baseline)				-0.36*	-0.33*
				(0.20)	(0.19)
Constant	2.64***	2.03**	2.72***	4.24***	4.34***
	(0.65)	(0.98)	(0.74)	(1.00)	(1.00)
Controls for age, education, income	No	Yes	Yes	Yes	Yes
Control for confidence to change	No	No	No	No	Yes
N	89	81	77	76	76
R-squared	0.39	0.44	0.63	0.68	0.69

Notes. OLS regressions of treatment effects on HbA1c (%). CONTROL is omitted category. Weekly RED, ORANGE, GREEN intake refers to weekly consumption of standard portions of foods from the traffic lights food lists given to patients during nutrition education. Patients self-reported their dietary intake in the preceding 4 weeks at endline. Global Preference Survey (GPS) and S-Grit Scale are non-incentivised survey questions. Significance is starred. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 provides an alternative OLS model with clustered standard errors at the individual level of the treatment effects to those shown in Table 3. Instead of treatment dummies for INFO and

COACH, we include binary dummy variables = 1 if real-time information was provided or coaching was provided, and 0 otherwise. The omitted category is CGM feedback for one month reviewed intermittently at doctor's visits only. The effect of coaching is clinically and statistically significant. A change in 0.5 percentage points is considered clinically significant. In Column 5, with all controls included, the effect of coaching is to lower HbA1c by 0.91 percentage points. This is a substantial effect. The effect of coaching is robust to additional controls. The independent variables included in Table 4, besides real-time and coaching, are identical to Table 3. We compare our results to Athinarayanan et al. (2019), which also advised a ketogenic diet as part of continuous care in the United States and tested blood results pre/post. Similar to their observations, we find baseline HbA1c is a significant predictor of endline HbA1c. However, we do not find that recent diagnosis is a robust predictor of better HbA1c. Interestingly, female participants have higher HbA1c, but age, education and income are never significant. Our results provide additional insights into other significant correlates of lower HbA1c. Namely, self-reporting higher consumption from the healthy GREEN list (meat, low-carb dairy, leafy vegetables) and lower consumption from the RED list (processed foods and foods high in carbohydrates and containing vegetable oil), and grit score, where an increase of 1 unit in the 5-point grit score is associated with a lower HbA1c of 0.33 percentage points. Patience is not a significant predictor of lower HbA1c, and the coefficient is very small (0.080). The sign of the -0.071 coefficient on procrastination, which is significant at the 10% level, is quite surprising but the effect is very small.

Result 2. *Providing real-time CGM for one month compared to intermittent supervised CGM reports is associated with a 0.47 percentage points higher HbA1c (worse control) after a 2-month sustainability period where the effects of CGM are allowed to wash out. Coaching lowers HbA1c by 0.91 percentage points which is clinically and statistically significant (better control).*

TABLE 4

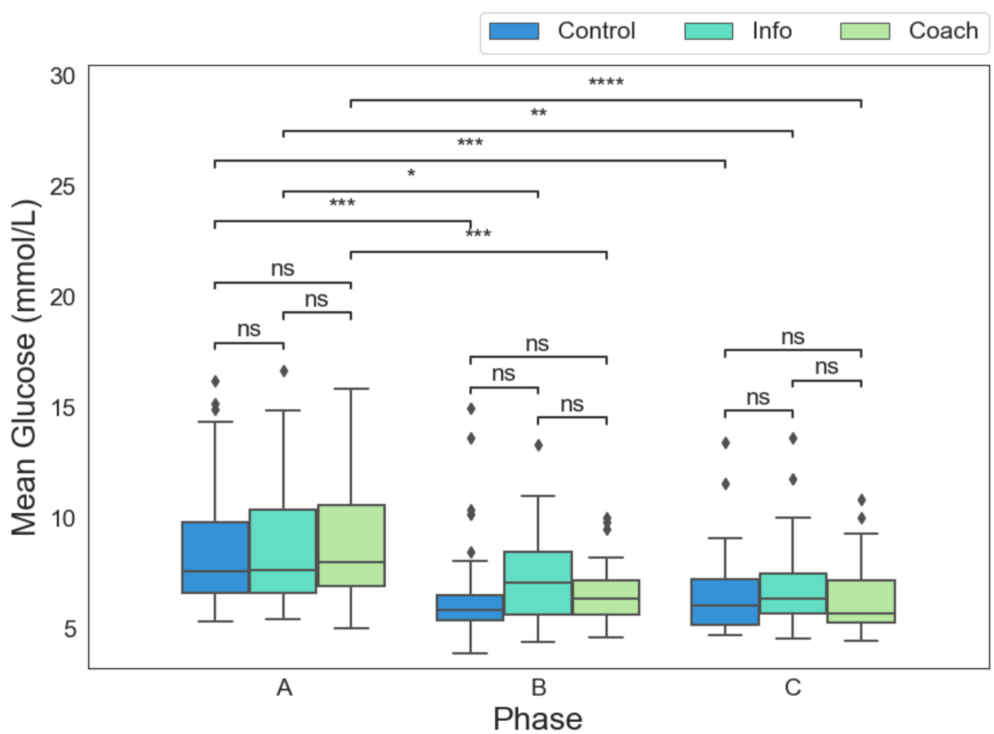
Real-time feedback and coaching effects on average glucose (HbA1c)

	Dependent variable: HbA1c endline				
	(1)	(2)	(3)	(4)	(5)
HbA1c baseline	0.54***	0.56***	0.41***	0.41***	0.42***
	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
Real-time	0.18	0.21	0.50**	0.48**	0.47**
	(0.26)	(0.27)	(0.25)	(0.22)	(0.22)
Coaching	-0.64**	-0.71***	-0.79***	-0.90***	-0.91***
	(0.25)	(0.25)	(0.21)	(0.24)	(0.25)
Female		0.55**	0.54***	0.68***	0.66***
		(0.24)	(0.20)	(0.21)	(0.20)
Recent diagnosis (<1 year)			-0.42*	-0.33	-0.31
			(0.24)	(0.24)	(0.25)
Weekly RED intake (endline)			0.021***	0.018***	0.018**
			(0.01)	(0.01)	(0.01)
Weekly ORANGE intake (endline)			0.030	0.029	0.025
			(0.03)	(0.03)	(0.03)
Weekly GREEN intake (endline)			-0.011**	-0.0094**	-0.0092**
			(0.00)	(0.00)	(0.00)
GPS Willingness to take risks (baseline)				0.10**	0.10**
				(0.05)	(0.05)
GPS Patience (baseline)				-0.095*	-0.080
				(0.05)	(0.05)
GPS Procrastinate (baseline)				-0.071*	-0.072*
				(0.04)	(0.04)
S-Grit Scale (baseline)				-0.36*	-0.33*
				(0.20)	(0.19)
Constant	2.64***	2.03**	2.72***	4.24***	4.34***
	(0.65)	(0.98)	(0.74)	(1.00)	(1.00)
Controls for age, education, income	No	Yes	Yes	Yes	Yes
Control for confidence to change	No	No	No	No	Yes
N	89	81	77	76	76
R-squared	0.39	0.44	0.63	0.68	0.69

Notes. OLS regressions of treatment effects on HbA1c (%). Real-time is a binary variable equal to 1 if real-time CGM feedback was provided (i.e. INFO/COACH treatments), 0 otherwise. Coaching is a binary variable equal to 1 if coaching was provided, 0 otherwise. Weekly RED, ORANGE, GREEN intake refers to weekly consumption of standard portions of foods from the traffic lights food lists given to patients during nutrition education. Patients self-reported their dietary intake in the preceding 4 weeks at endline. Global Preference Survey (GPS) and S-Grit Scale are non-incentivised survey questions. Significance is starred. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6. Continuous Glucose Monitoring patterns

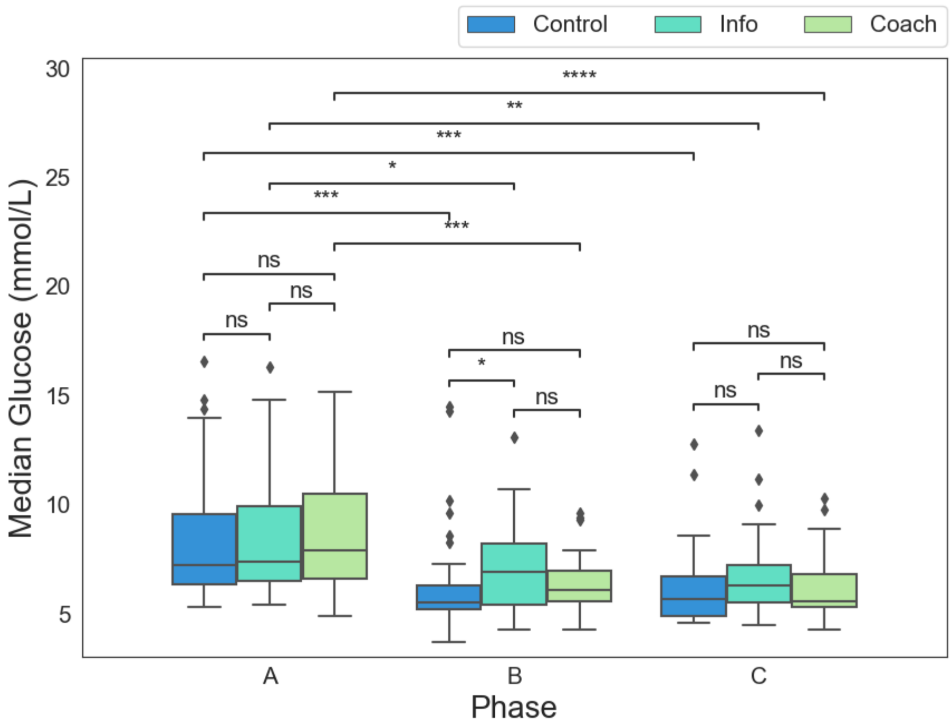
Figure 15 Panel A shows box-and-whisker plots of the mean glucose levels as measured by the CGM. Panel B shows median levels. The box plots summarise the data for CONTROL, INFO, and COACH from baseline (Phase A) to intervention (Phase B and C). At baseline, there is no significant difference between treatment groups in mean glucose levels. This shows that the randomisation worked. The mean significantly drops for all three treatment groups from Phase A to Phase B. However, the INFO group's change from baseline to Phase B is only significant at the 10% level. By Phase C, the INFO group has significantly reduced glucose levels at the 1% level compared to baseline. There is no significant difference between treatment groups in Phase C. After C, patients stopped wearing a CGM for the two-month sustainability period in which the CGM effect should wash out.



Notes. Phase A is the baseline, Phase B and C are intervention. K-Wallis tests for differences between groups. Significance is starred. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Not significant is indicated by ns.

FIGURE 15 Panel A

Mean glucose levels in baseline and intervention by treatment group

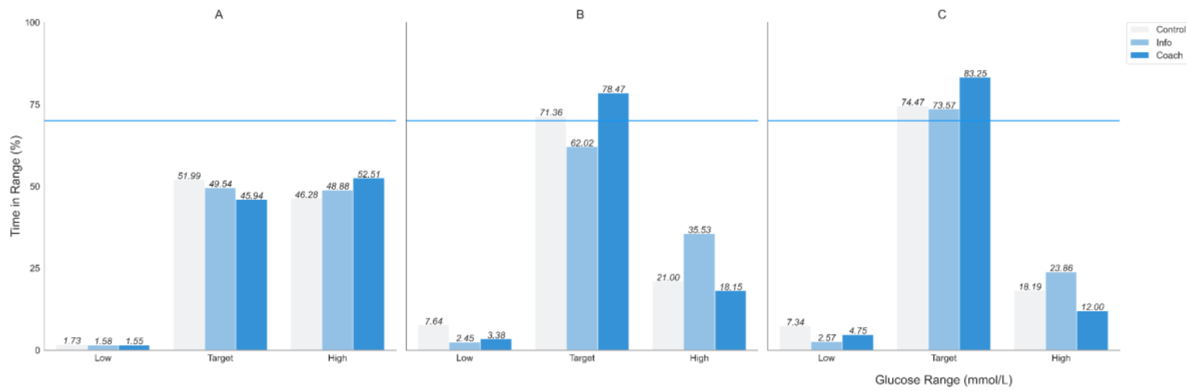


Notes. Phase A is the baseline, Phase B and C are intervention. K-Wallis tests for differences between groups. Significance is starred. * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001. Not significant is indicated by ns.

FIGURE 15 Panel B

Median glucose levels in baseline and intervention by treatment group

Figure 16 below shows time in target range during the baseline (A) and 1-month main intervention period (B and C) in which the CGM is worn. Reassuringly, all three groups start below the clinical goal of 70% time in range (3.9-8 mmol/L), see A: Time in Target 64-52%. There is significant room for all three groups to improve. After wearing CGM A, patients review the report with the doctor and receive ketogenic dietary advice when they are instructed to follow until the end of the trial. We see this reflected in increases in time in target range in B and C in Figure 16.

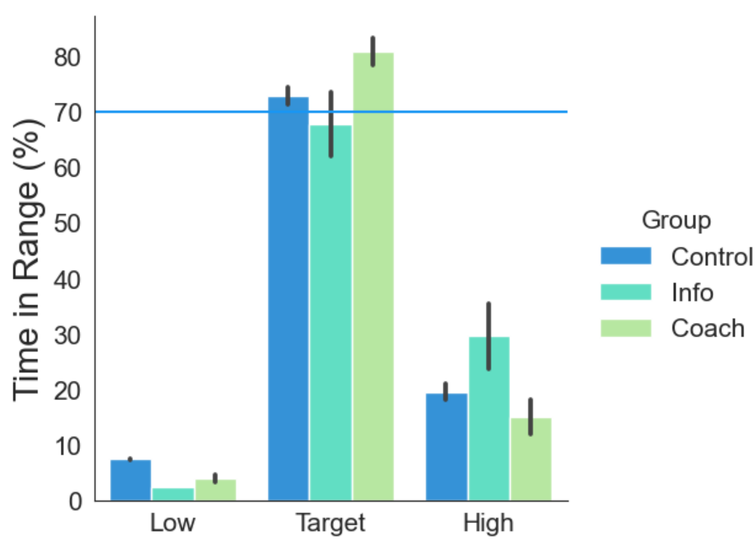


Notes. Target range in the trial is 3.9-8 mmol/L, and the goal is to spend at least 70% of the time in target. Patients in INFO and COACH had access to real-time feedback and could view their time in target range between appointments. All three groups started below the goal, and all three groups achieved the glucose control goal of 70% after 1 month.

FIGURE 16

Time in Target Range by treatment group

Figure 17 excludes the baseline period (A) and collapses the two Phases B and C shown above in Figure 16 by treatment group. All groups achieved their 70% time in target goal. Confidence intervals are shown on the bar graphs. COACH group achieves the greatest glucose control as measured by the CGM's time in target range.



Notes. This figure shows time in target range during Phase B and C only, *i.e.* it excludes the baseline measurements from the CGM (*i.e.* Phase A). 95% confidence intervals illustrated on bar graph.

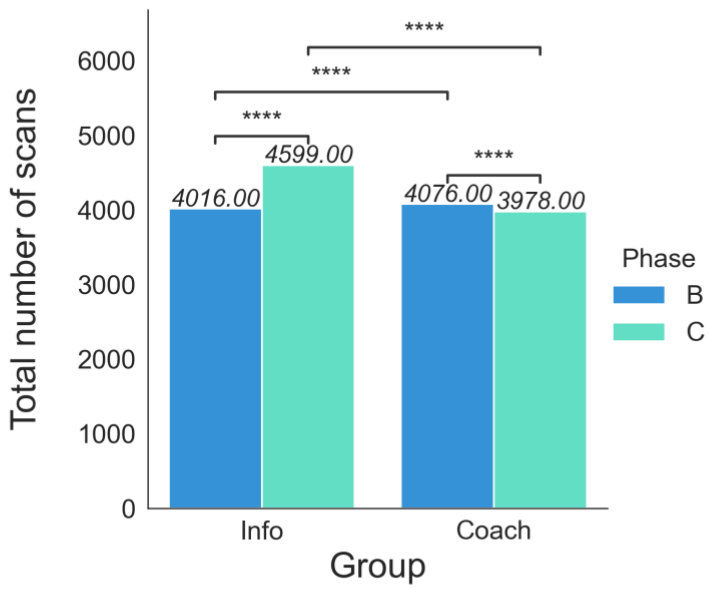
FIGURE 17

Time in Target after ketogenic nutrition education, excluding baseline CGM data.

Result 3. *All three treatment groups significantly reduced average glucose levels and increased time in target range to achieve the clinical goal above 70% in target range after 1 month following the ketogenic dietary advice. No significant difference in average glucose levels between groups while the CGM is still being worn (midline). Time in target range is the highest for COACH and not significantly different for CONTROL and INFO.*

Figure 18 shows the total number of scans of the real-time feedback CGM with the mobile phone app by treatment group. CONTROL is excluded because they could not engage with their CGM monitor. The INFO group increased their CGM scanning activity significantly during the intervention from Phase B to Phase C ($p<0.0001$). This is associated with improved time in target range within INFO over time. However, greater scanning behaviour is not sufficient to overtake COACH in time in target range. This goes against our hypothesis that COACH would have greater engagement with the real-time CGM and that this could be a mechanism for their greater glucose control. We do not find evidence of this. Instead, this suggests that the COACH group is better able or better supported to use real-time feedback to make healthier food choices with additional support from someone besides their doctor. A critical part of our health coaching is mental contrasting, implementation intentions techniques and motivational interviewing. We speculate that the COACH group was less overwhelmed at the adoption of the novel technology. Their scanning activity was initially higher than INFO in Phase B ($p<0.0001$) and significantly lower than INFO by Phase C ($p<0.0001$). COACH glucose control and time in target is consistently high after they receive nutrition education. Their better glucose control is not associated with more scans of the CGM, *i.e.* more frequent self-monitoring.

The unsupervised real-time CGM caused some patients anxiety for several reasons. Firstly, they noticed the dawn phenomenon, in which glucose spikes in the morning before they have eaten. This is caused by the liver making its own glucose, and patients observed that the uncontrolled morning rise lessened as they followed the dietary advice. Many patients sought reassurance about this. Secondly, because the CGM measures interstitial glucose and not blood glucose, it is less accurate, and there can be a difference between blood glucose and CGM reading. Discrepancies caused confusion because patients were not sure which feedback to follow. It was common for the CGM to report a reading 1 mmol/L lower than the fingerprick. Patients were advised to double-check with a fingerprick if they observed a hypoglycaemic reading to ensure it was a real "low". They were advised to "correct" low blood sugar readings with foods from the healthy GREEN list, rather than sugary items which fall on the RED list. There were also device failures, and patients seemed distressed when the device failed and had to be replaced. Patients reported failures via WhatsApp message to the research team, were invited to come in to have it replaced by the nurse, and received replacements when less than 9/14 expected days of data were stored by their device. We do not input missing data in this analysis. Abbott admitted that we had received a faulty batch of CGMs. This caused more interruptions and additional appointments with the nurse than we anticipated.



Notes. Total number of scans by the INFO and COACH groups, which wore real-time CGM, during the one-month intervention. K-Wallis tests reported. Significance is starred. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

FIGURE 18

Scanning behaviour by real-time CGM groups

Result 4. *The treatment group with real-time feedback alone (no coaching) increased its scanning activity during the first month of the intervention, while those with coaching slightly reduced their engagement with the CGM over time. After one month of the intervention, INFO scans significantly more often than COACH.*

4. CONCLUSION

Behavioural economics is establishing itself as an important discipline in the modifying of self-destructive behaviour, such as lifestyle choices contributing to diabetes (John et al., 2011). This research contributes to a growing literature showing that insights from behavioural economics and psychology can help people engage in behaviours that are consistent with their long-term wellbeing. The experimental design allows us to rigorously evaluate the impact of real-time feedback from CGM with and without the continuous care component of health coaching. We build on previous clinical trials using nutritional ketosis for the management and reversal of T2DM. To our knowledge, we are the first to disentangle standard fingerprick self-monitoring blood glucose management and real-time CGM from a comprehensive continuous care intervention with health coaching. Unpacking some components from the black box of a behavioural intervention allows us to draw conclusions about the minimal cost- and time-effective treatment necessary for reversing T2DM. It is also critical to understand the underlying cognitive and behavioural barriers in this setting, which we explore through surveys. Finally, this study contributes to the active behavioural economics literature on habit formation. Fang et al. (2020) argue that when multiple barriers are present, such as imperfect information and limited attention (and in our setting of food choice and also cultural norms), a single intervention that does not address all barriers simultaneously might fail to fulfil its potential. This motivates our third treatment that combines real-time feedback with health coaching. Indeed, we find that addressing only the issue of real-time information is not sufficient for sustained behaviour change after the CGM is removed for two months.

All groups show significant improvement in mean glucose levels as measured by the CGM. The rich report of the CGM may help to close the information monitoring gap between patient and doctor, and offers an accountability mechanism, but we do not find evidence of the value added by real-time information for the patient, unless accompanied with coaching. All groups achieve the time in target goal of 70%. After a sustainability period of two months without CGM, INFO has significantly worse outcomes than CONTROL, though an improvement compared to baseline. COACH achieves the greatest reduction in Haemoglobin A1c (HbA1c, average blood glucose over the past 2 months used to diagnose diabetes) of about 19%, while CONTROL reduces by 14% and

INFO 11%. OLS regression results show that COACH achieves significantly better glucose control. COACH is the only group to significantly improve other markers of metabolic health HDL (cholesterol) and blood triglycerides. A significant reduction in blood triglycerides and an increase in HDL is considered an improvement in metabolic health (Alberti et al., 2005).

In the HbA1c blood tests, INFO performs significantly worse than CONTROL after the sustainability period in which the CGM effect is allowed to wash out. The success of the coaching group was not due to scanning the real-time CGM more often. COACH scans significantly less often than INFO, but the additional weekly support from someone besides the doctor appears to provide important supervision of the adoption of real-time CGM in this population in South Africa. All treatments see a significant reduction in HbA1c compared to baseline. In the pooled sample, 43% of patients reversed diabetes (HbA1c<6.5%), notable compared to the best practice standard of care reversal rate of <1%. This is comparable to the 55% reversal rate at 3 and 6 months from a non-randomised continuous care intervention study including coaching and nutritional ketosis in the United States (Athinarayanan et al., 2019). Our study differs from previous work because we have a sustainability period in which the effects could wash out, we randomise patients into treatment groups with varying levels of input, and we conduct it in a developing country setting. The high diabetes reversal rate given the potential challenges of new technology adoption and varying levels of health literacy in our setting speak to the efficacy of our more modest intervention.

Reporting higher consumption of unhealthy food from the RED list is associated with higher HbA1c at endline, and greater consumption of food from the healthy GREEN list is associated with lower HbA1c at endline. Scoring higher on grit is significantly associated with greater glucose control at endline. Being more willing to take risks in general is significantly associated with worse glucose control at endline. Identifying economic preferences of patients could potentially help to tailor programmes in the future. For example, to predict which patients would benefit from more frequent accountability mechanisms and remote support. This timely research contributes to the behavioural and medical literatures by providing evidence of the effect of technology-assisted behavioural interventions that incorporate a ketogenic diet to reverse diabetes.

5. APPENDICES

A. Waiting Room Notice

You are invited to participate in research!

“The effect of technology-assisted behavioural interventions in type 2 diabetics”

The purpose of the study
We want to understand how best to support type 2 diabetes patients to build sustainable healthy eating habits. What are the necessary elements of a program to help individuals achieve their diet and glucose control goals? What level of personalised feedback and technology is most helpful? We want to learn from your experience over the next 3-months. Your participation will help us understand important obstacles and cost-effective solutions to support patients’ diabetes reversal.

Would like to support us by participating in our research?

Who may take part in the study?
About 90 patients will be included in the study. Eligible participants must be: men or women; age 30-65 years old; have a type 2 diabetes diagnosis; HbA1c test above 6.5% and below 11%; not using insulin; no previous experience using a continuous glucose monitor, a current patient of Kenilworth Medicross, able to read and speak English; use a smartphone, willing and able to attend the five required study visits, wear three continuous glucose monitors over six weeks and keep a food diary for 8 days in total. You will be excluded if you are pregnant, have anemia or an unmanaged hypothyroid condition.

Do You Have Any Further Questions? Contact the researchers.
Sofia Monteiro Cell: +27 72 253 6420 Email: sofia.g.monteiro@gmail.com
Dr. Kate Larmuth Cell: +27 74 217 0042 Email: kateus65@gmail.com

Ask your Medicross doctor about the study or contact the researchers to find out more. You can also contact Dr. Neville Wellington directly.

Locally responsible study physician/study center
Dr Neville Wellington
Kenilworth Medicross Medical Centre
TEL: 021 683 5867
E-Mail: neville.wellington@medicross.co.za

B. Study Information for Participants and Informed Consent Form

Principal investigator (UCT):
Dr. Jacolene Kroff Tel: +27 21 650 5126 Email: jacolene.kroff@uct.ac.za
Dr. Kate Larmuth (Co-PI) Cell: +27 74 217 0042 Email: kateus65@gmail.com

Principal investigator (University of Cologne, Germany):
Prof. Daniel Wiesen Email: wiesen@wiso.uni-koeln.de

Study director (Max Planck Institute for Research on Collective Goods):
Prof. Matthias Sutter Email: sutter@coll.mpg.de

PhD Researcher:
Sofia Monteiro (MPI, Uni Cologne) Cell: +27 72 253 6420 Email: sofia.g.monteiro@gmail.com

The overall responsibility for this study lies with
Dr. Jacolene Kroff
Tel: +27 21 650 5126 Email: jacolene.kroff@uct.ac.za
Division of Exercise Science and Sports Medicine
Department of Human Biology
University of Cape Town
PO Box 115
Newlands
7725
Cape Town, South Africa

Locally responsible study physician/study center
Dr Neville Wellington
Kenilworth Medicross Medical Centre
67 Rosmead Ave
Kenilworth
7700
TEL: 021 683 5867
E-Mail: **neville.wellington@medicross.co.za**
Information about Participating in this Study

"The effects of technology-assisted behavioural interventions in type 2 diabetics"

Dear Madam, Dear Sir,
We are pleased that you are interested in our research. In it, we look at lifestyle choices with patients who have type 2 diabetes. We would like to invite you to join us and help to generate more valuable research around type 2 diabetes.
In this document we will tell you about our study and why we are doing it. We will tell you what is asked of you and what the risks and benefits of participating are. At the end of the document is a page which you will have to complete and sign with a researcher (remotely online or in person).
Please read the following carefully and then decide whether or not you would like to participate in this study.

Your participation in this study is entirely up to you. You will only be included in this study if you give your consent (your agreement) in writing (in the last section of this document). You may choose to leave the study at any point in time, however, we would really appreciate you taking part, if you choose to sign up.
Your doctor has already given you a lot of information about the study. This document will give you more important information about the study goals and your role. Please read the participant information carefully. Afterwards, a researcher will answer your questions. Please do not hesitate to ask questions or point out any areas that are not clear to you. You may then decide, in your own time, whether or not you wish to take part.

The purpose of the study

We want to understand how best to support type 2 diabetes patients like yourself to build sustainable healthy eating habits. To do this we want to answer the following questions:
What elements of a program are required to help individuals achieve their diet and blood sugar goals?
What level of personalised feedback and which type of technology is most helpful?
We want to learn from your experience over the next 3-months in the program we have designed for you. Your participation will help us understand important obstacles and solutions to support patients' diabetes reversal.

Type of study

This randomized controlled experiment will measure behaviour and health outcomes. There will be about 90 participants. The study has three different groups that receive a slightly different intervention and you will have an equal chance of being assigned to participate in any one of them. You will not be given information about the other groups. The three groups will vary in the amount of personalised feedback and healthcare practitioner contact. We want to find out which program is the most helpful to patients like yourself to bring your HbA1c test below 6.4% (in other words to put your diabetes into remission).
We will use new continuous glucose monitoring or CGM technology to measure your sugar levels that does not require finger prick testing. All patients will participate in a baseline period for two weeks where we measure your current diet and sugar levels. Then all patients will receive a 1-month intervention which includes nutrition education, medical advice and regular check ups with a study doctor to monitor your progress. You will also wear a CGM during the intervention and baseline period. We will explore which type of feedback and support is most useful to patients like yourself to fine tune your diet and help you reach your health goals. We will also use health assessments (like your BMI, weight and cholesterol levels), questionnaires (including one on Covid-19 and your experiences during this time) and mobile apps. To complete the study, all participants must wear three CGM sensors and attend five study visits to your doctor over three months. All other contact will happen remotely. At the beginning and end of the study the doctor will do blood tests to see if you have reached your health goals..
The project will be financed by the Max Planck Institute for Research on Collective Goods.
The Human Research Ethics Committee of the University of Cape Town (HREC) advised on and approved the present research project.

Who may take part in the study?

About 90 patients will be included in the study. Eligible patients must be: men or women; age 30-65 years old; have a type 2 diabetes diagnosis; HbA1c test above 6.5% and below 11%; not using insulin; no previous experience using a continuous glucose monitor, a current patient of Kenilworth Medicross Medical Centre or a referral from the Cape Town catchment area, able to read and speak English; use a smartphone, willing and able to attend the five required study/doctors visits, wear three continuous glucose monitors over six weeks and keep a food diary for 8 days in total. You cannot take part in this study if you are pregnant, breast-feeding or fall pregnant during the study. You cannot take part if you have anemia, an unmanaged hypothyroid condition, renal disease (e.g. poor kidney function or kidney stones) or an eating disorder (e.g. anorexia, bulimia).

What are the possible risks?

The potential dangers in taking part in this study are small. A nurse will take your blood at Visit 1 and 5, and you will be asked to talk about personal topics in the baseline questionnaire. Lastly, wearing the CGM sensors means you will have a small device (size of a 5 Rand coin) attached to the back of your arm

COVID-19: SARS-CoV2 infection is a danger for people with type 2 diabetes. However, this danger is not increased by your participation in our study. Medicross has Personal Protective Equipment (PPE), enforced social distancing and follows good clinical practice guidelines to keep everyone safe. All recommended care will be taken to ensure your health and safety. Patients must contact the researchers if they have any concerns or negative experiences. Please contact your study doctor if you fall ill. Do not attend a study visit to Medicross if you experience shortness of breath, dry cough or a temperature. Consult government websites such as sacoronavirus.co.za. We will use remote monitoring (email, phone calls, or zooms) to limit the need for person-to-person contact in the study.

Blood collection: Your study doctor will order blood tests. This is part of the normal care in diabetes management. Occasionally, blood collection can result in minor infection, some physical pain, mental discomfort, fainting and injury to a nerve or a vein. These risks are small and will be minimized by the use of staff trained to take blood samples and the use of disposable, single-use materials..

CGM: The tiny internal part of the continuous glucose monitor is a sensor. It is inserted under the skin in an instant with a disposable applicator. A nurse will do this for you at Medicross to ensure it is applied correctly. The adhesive sticker securing the external part of the sensor may potentially irritate the skin. It is safe to shower and swim with your CGM on.

Surveys: The baseline survey, COVID-19 and diabetes survey and the food frequency questionnaire have the potential to generate negative emotions since they cover socio-economic issues, health and disease.

What are the possible benefits?

You will take part in a short study of your glucose control that has the potential to empower you with knowledge and confidence to improve your long term health outcomes. For patients like you who have poor glucose control (HbA1c 6.5 - 11%) but are not yet on insulin medication, the potential health benefits of making food choices to stay in recommended glucose ranges and lowering HbA1c to at least 6.4% are substantial. Good health lowers the risk of poor outcomes with COVID-19. You will receive a personalised lifestyle program which aims to put your diabetes into remission. You may enjoy using the CGM (at no personal expense) and learn to manage and reverse your disease from personalised feedback. You will be given a full personal feedback report once we have collected and analysed all the data so you can see how you contributed to the study. You may feel a sense of fulfilment knowing that you are contributing to improving the lives of patients like yourself by taking part in research. You will receive R200 cash for transport and time at study visits. All participants will receive informative doctors consults with one of the study doctors and a limited set of blood tests for free. You will contribute to the scientific understanding of behaviour change and diabetes.

What is required of me?

There are three distinct phases in the study. Briefly, phase 1 includes the informed consent and baseline procedures, phase 2 is the major intervention phase and phase 3 the follow-up and sustainability assessment section. There are five study visits and your participation will take about 3 months.

Phase 1: Baseline (2 weeks): Clinic Visit 1 (1 hour)

After you are enrolled (given your consent to begin the study), you will complete three questionnaires about yourself, what you usually eat, and challenges you may have experienced because of the COVID-19 pandemic. You can complete these online or via telephone call. . Signing the informed consent form will allow us to collect and analyse standard blood tests for diabetes, body composition (eg. BMI and height and weight), individual characteristics, food diary entries and CGM data. You will have a half hour consultation with your study doctor. Your doctor will order blood tests that are standard in diabetes management. For two weeks you will carry on with your normal routine so we can measure your usual habits. You will wear a continuous glucose monitor and keep a detailed food diary for four days using the FatSecret app on your smartphone (or a paper diary if you prefer).

Phase 2: Intervention (4 weeks): Clinic Visits 2-4 (30-minutes)

At Visit 2, you will receive a nutritional education session from your study doctor. Your doctor will give you personalised medical advice and prescriptions based on your current health markers and CGM data at your 30-minute consultation. The nutrition education is standard for diabetes patients in the study doctors practice. A nurse will apply a new CGM for you to wear over the next two weeks while you fine tune your diet. This will be repeated at Visit 3. Your goal will be to put your diabetes into remission by following your study doctor's advice. In the final week of the intervention you will keep a food diary for 4 days. At Visit 4 your CGM will be removed by the nurse. Your study doctor will review your progress and sugar patterns at each consultation.

Phase 3: Sustainability (2 months): Visit 5 (1 hour)

For two months you will continue following the nutrition advice of your doctor. You will self-monitor until your Visit 5 doctor's consultation and blood test. You will not wear the CGM in this period. You will be free to continue self-monitoring using the food diary app or fingerprick glucose tests if you wish. At the end you will complete a closing questionnaire about your experiences and we will ask you about what foods you ate in the last month.

What happens after the study?

Your study doctor will be invested in your long term health and wellbeing. After your participation is completed they will follow up with you about a year later to check up how you are doing. You may contact them immediately after the study if you wish to continue treatment with them.

Data Processing and Data Protection

Your data, including data on health, sex, age, weight, and height, blood pressure and biometric data will be recorded and analysed in coded form so that there is no way to identify you. It will only be possible to link study data to you personally via an identification list kept secure at UCT. Only the study researchers will have access to this list.
The legal basis for data processing is your voluntary, written consent. You must first give your permission.
The data collected during the study will be recorded and evaluated by an electronic data system. After the study ends, all data will be stored in a secure system for a period of 10 years. Afterwards, your data will be deleted unless legal reasons require longer storage.
The processing of the data is the responsibility of the researchers and the locally responsible physician.
You have the right to access and look at your personal data which are collected. This data will not be shared with third parties.
If you discover any errors in your data, you have the right to have them corrected. You also have the right to receive a free copy of these data.
In addition, you have the right to request deletion of your stored data.
Scientific publications of the results are only carried out anonymously, i.e., in a form that does not link to you personally.
The running of the study will be checked by an independent person whose job is to make sure the study is run properly". This person and the researchers may only pass your data in a form that is anonymous and may not make copies of your information.

Are There Any Risks Associated with Data Processing?

Every collection, storage, use, and sending of data involves confidentiality risks (e.g., the possibility of identifying the person concerned). These risks cannot be completely excluded, and they increase the more data can be interlinked. The researchers involved in this study assure you that they will do everything possible to protect your privacy in accordance with the latest technology.

Can I Withdraw My Consent?
You can take away your consent at any time, in writing or on the phone, without giving reasons and without any disadvantage to you. If you decide you want to quit, no further data will be collected.
If you decide to quit, you may be asked for your consent that the data and blood samples already collected may be used further. If you do not agree all your data will be deleted immediately.

Other medical treatment
While you are participating in the study you may only undergo other medical treatment – except in emergencies – after prior consultation with your study doctor. You must inform the study doctor immediately of any emergency treatment you receive or if you have any medical concerns.

Reimbursement of Expenses
There are no costs for you or your health insurance company for participation in the study. Travel costs will be reimbursed to you in the amount of R200 per visit.

What if Something Goes Wrong?
The University of Cape Town (UCT) undertakes that in the event of you suffering any significant deterioration in health or well-being, or from any unexpected sensitivity or toxicity, that is caused by your participation in the study, it will provide immediate medical care. UCT has appropriate insurance cover to provide prompt payment of compensation for any trial-related injury according to the guidelines outlined by the Association of the British Pharmaceutical Industry, ABPI 1991. Broadly-speaking, the ABPI guidelines recommend that the insured company (UCT), without legal commitment, should compensate you without you having to prove that UCT is at fault. An injury is considered trial-related if, and to the extent that, it is caused by study activities. You must notify the study doctor immediately of any side effects and/or injuries during the trial, whether they are research-related or other related complications.

UCT reserves the right not to provide compensation if, and to the extent that, your injury came about because you chose not to follow the instructions that you were given while you were taking part in the study. Your right in law to claim compensation for injury where you prove negligence is not affected. Copies of these guidelines are available on request.
Please feel free to contact the UCT Human Research Ethics Committee if you have any queries: Floor E53-Room 46, Old Main Building, Groote Schuur Hospital, Observatory, 7925
Phone: (021) 406 6338 Email: hrec-enquiries@uct.ac.za

Do You Have Any Further Questions?
Sofia Monteiro Cell: +27 72 253 6420 Email: sofia.g.monteiro@gmail.com
Dr. Kate Larmuth Cell: +27 74 217 0042 Email: kateus65@gmail.com

INFORMED CONSENT

I____(full name) hereby consent to participate in the study visits, nutrition education, glucose monitoring, food diary and questionnaires as part of the study entitled: "The effect of technology-assisted behavioural interventions in type 2 diabetics." - carried out by the investigators of the University of Cape Town, Max Planck Institute for Research on Collective Goods and University of Cologne.

All the procedures, risks and benefits have been explained to me in a language that I understand and all my questions have been answered.	Yes	No

I am aware that I am free to ask more questions about the study procedures at any time	Yes	No
I am aware that I do not have to answer any questions that I don't want to	Yes	No
I am aware that participation in this study is entirely voluntary. I am free to withdraw from the study at any time.	Yes	No
I also understand that the investigators may withdraw me from the study at any time without reason.	Yes	No
By agreeing to participate in this study, I am aware that my information will be stored in a safe and anonymous way at UCT and will then be analysed by the investigators only, in coded form.	Yes	No

DATA PROTECTION: *This scientific study will collect personal data and medical findings about you. The storage, transfer, and analysis of these data is carried out in accordance with all current legal regulations and requires the following voluntary consent prior to participation in the study*		
I agree that data/diagnostic data on questionnaires and electronic data carriers collected within the scope of this study may be recorded in coded form and forwarded to:		
- the principal investigator and primary researchers	Yes	No
- the study physicians and health coach	Yes	No
Medical doctors have to uphold medical confidentiality of their patients. I agree that the study researchers (listed at the beginning of this document) may inspect my personal data held by the study doctor only if it relates to the studyi. For these measures only do I release the study doctor from the medical confidentiality obligation.	Yes	No
I agree that my data may be stored, for up to 10 years, after the end of the study according to scientific standards. After this period, my personal data will be deleted or made anonymous (any lists identifying you to your data code will be deleted)	Yes	No
I have been informed that I can terminate my participation in the study at any time. In this case, I may choose if my data, already collected, will be deleted or made completely anonymous.	Yes	No

Participant Name:____
Date of birth: ____

EMERGENCY CONTACT DETAILS
Name and Surname: _________________________________
Relation: _________________________________
Home phone: _________________________________
Cell phone: _________________________________

Print name of participant	Signature	Date
Print name of investigator	Signature	Date

C. *Baseline Questionnaire, HREC REF: 310/2020*

Survey Flow

Block: Introduction (2 Questions)
Standard: WHO STEPS: Demographics (8 Questions)
Standard: WHO STEPS: Tobacco Use (2 Questions)
Standard: WHO STEPS: Alcohol Consumption (8 Questions)
Standard: WHO STEPS: Physical Activity (21 Questions)
Standard: WHO STEPS: History of Raised Blood Pressure (4 Questions)
Standard: WHO STEPS: History of Diabetes (9 Questions)
Standard: WHO STEPS: History of Cardiovascular Diseases (5 Questions)
Standard: GPS Risk Preferences (1 Question)
Standard: GPS Delay gratification, math ability, procrastination (3 Questions)
Standard: Subjective Wellbeing (NIDS Household Survey) (1 Question)
Standard: Diabetes Distress Scale (DDS) (2 Questions)
Standard: Shopping and cooking (3 Questions)
Standard: Short Grit Scale (1 Question)
Standard: Self-efficacy and importance of behaviour change (2 Questions)
Standard: Stage of change (2 Questions)

Page Break

Start of Block: Introduction

instructions
Congratulations! You are enrolled into the exciting study "The effect of technology-assisted behavioural interventions in type 2 diabetics".

This Baseline Questionnaire will take about 30 minutes.

Please answer the following questions as honestly as you can. Your answers are important to us. Remember that your responses are confidential and there are no right or wrong answers. Your name and personal identifying information will not be analysed or published.

Honest feedback about your personal experience managing type 2 diabetes will help us improve the program for other patients like you.

You can contact the researcher, Sofia Monteiro, if you have any questions about the study:
Mobile: +27 72 253 6420
WhatsApp: +49 152 5542 7686

You can contact your study doctor if you have any concerns about your health during the study.

---

participant_code **Participant code** (enter your unique participant code you received from the researchers):

____

End of Block: Introduction

Start of Block: WHO STEPS: Demographics

education What is the **highest level of education** you have completed?

- ○ No formal schooling (1)
- ○ Some primary school (2)
- ○ Primary school completed (3)
- ○ High school completed (4)
- ○ College/University completed (5)
- ○ Post graduate degree (6)
- ○ Prefer not to answer (7)

ethnicity What is your ethnic/racial **background**?

- ○ Black (5)
- ○ Coloured (6)
- ○ White (4)
- ○ Asian (7)
- ○ Indian (8)
- ○ Other (9) ____
- ○ Prefer not to answer (10)

marital_status What is your **marital status**?

- ○ Never married (1)
- ○ Currently married (2)
- ○ Separated (3)
- ○ Divorced (4)
- ○ Widowed (5)
- ○ Cohabiting (6)
- ○ Prefer not to answer (7)

Page Break

age_years How old are you? (Enter number of years e.g. 65)

____

sex What is your **gender**?

○ Male (1)

○ Female (2)

○ Other (3) ____

○ Prefer not to answer (4)

employment Which statement best describes your **current employment status**?

○ Working (paid employee) (1)

○ Working (self-employed) (2)

○ Not working (temporary layoff from a job) (3)

○ Not working (looking for work) (4)

○ Not working (retired) (5)

○ Not working (disabled) (6)

○ Not working (other) (7)

○ Prefer not to answer (8)

hh_members How many people in total, **including yourself**, live in your household?

____

hh_income Information on income is important to understand. Can you give an estimate your **annual household income**?

We are looking for the TOTAL yearly income generated by all people who live and contribute to your household. If you don't have an exact amount an estimate is fine.

- Up to R20,500 (1)
- More than R20,500 but less than R89,000 (2)
- More than R89,000 but less than R202,500 (3)
- More than R202,500 but less than R412,000 (4)
- More than R412,000 but less than R707,000 (5)
- More than R707,000 but less than R1,512,000 (6)
- More than R1,512,000 but less than R2,414,000 (7)
- More than R2,414,000 (8)

End of Block: WHO STEPS: Demographics

Start of Block: WHO STEPS: Tobacco Use

smoke_daily Do you currently smoke daily any **tobacco products**, such as cigarettes, cigars or pipes?

- Yes (1)
- No (2)

smoke_e-cigarettes Do you currently smoke any **e-cigarettes** daily?

- Yes (1)
- No (2)

End of Block: WHO STEPS: Tobacco Use

Start of Block: WHO STEPS: Alcohol Consumption

instructions The next questions ask about the **consumption of alcohol**.

alcohol_ever Have you ever consumed any **alcohol** such as beer, wine, spirits?

- Yes (1)
- No (2)

Skip To: End of Block If Have you ever consumed any alcohol such as beer, wine, spirits? = No

alcohol_past_year Have you consumed any alcohol within the past 12 months?

- Yes (1)
- No (2)

Skip To: alcohol_freq_past_yr If Have you consumed any alcohol within the past 12 months? = Yes

alcohol_stopped Have you stopped drinking due to health reasons, such as a negative impact on your health or on the advice of your doctor or other health worker?

- Yes (1)
- No (2)

Skip To: End of Block If Have you stopped drinking due to health reasons, such as a negative impact on your health or on t... = Yes

Skip To: End of Block If Have you stopped drinking due to health reasons, such as a negative impact on your health or on t... = No

alcohol_freq_past_yr During the past 12 months, **how frequently** have you had at least one alcoholic drink?

- Daily (1)
- 5-6 days per week (2)
- 3-4 times a week (3)
- 1-2 days per week (4)
- 1-3 days per month (5)
- Less than once a month (6)
- Never (7)

alcohol_past_mo Have you consumed an alcoholic drink within the **past 30 days**?

- Yes (1)
- No (2)

Skip To: End of Block If Have you consumed an alcoholic drink within the past 30 days? = No

alcohol_freq_past_mo During the past 30 days, on how many **occasions** did you have at least one alcoholic drink? (Enter number)

____

alcohol_intensity During the past 30 days, when you drank alcohol, on average, how many **standard alcoholic drinks** did you have during one drinking occasion?

____

End of Block: WHO STEPS: Alcohol Consumption

Start of Block: WHO STEPS: Physical Activity

instructions **Physical Activity**:Next we are going to ask you about the time you spend doing different types of **physical activity** in a typical week. Physical activity is anything that you do that's active from sports to cleaning the house to gardening. Please answer these questions even if you do not consider yourself to be a physically active person.
In answering the following questions '**vigorous-intensity activities**' are activities that require hard physical effort and cause large increases in breathing or heart rate and some sweating. '**moderate-intensity activities**' are activities that require moderate physical effort and cause small increases in breathing or heart rate.

instructions **Physical activity at work (paid employment or unpaid work at home)**
The following set of questions ask about your physical activity doing work.

work_vigorous
Does your work involve **vigorous-intensity activity** that causes large increases in breathing or heart rate like [carrying or lifting heavy loads, digging or construction work] that causes some sweating for at least 10 minutes continuously?

- Yes (1)
- No (2)

Skip To: work_moderate If Does your work involve vigorous-intensity activity that causes large increases in breathing or he... = No work_vigorous_freq In a typical week, on how many days do you do **vigorous-intensity activities** as part of your work?

- Daily (1)
- 4-6 times a week (2)
- 2-3 times a week (3)
- Once a week (4)

work_vigorous_time How much time do you spend doing **vigorous-intensity activities** as part of your work on a typical day? (Hours: minutes)

____

work_moderate Does your work involve **moderate-intensity activity**, that causes small increases in breathing or heart rate such as brisk walking [or carrying light loads] for at least 10 minutes continuously?

- Yes (1)
- No (2)

Skip To: instructions If Does your work involve moderate-intensity activity, that causes small increases in breathing or h... = No

work_moderate_freq In a typical week, on how many days do you do **moderate-intensity activities** as part of your work?

- Daily (1)
- 4-6 times a week (2)
- 2-3 times a week (3)
- Once a week (4)

work_moderate_time How much time do you spend doing **moderate-intensity activities** as part of your work on a typical day? (Hours: minutes)

____

instructions **Travel to and from places**:The next questions exclude physical activities at work that you have already mentioned earlier.
Now we would like to ask you about the usual way you travel to and from places. For example to work, for shopping, or to place of worship.

travel_walk Do you walk or use a bicycle for at least 10 minutes continuously to get to and from places?

- Yes (1)
- No (2)

Skip To: instructions If Do you walk or use a bicycle for at least 10 minutes continuously to get to and from places? = No

travel_walk_freq In a typical week, on how many days do you walk or bicycle for at least 10 minutes continuously to get to and from places?

- Daily (1)
- 6 times a week (2)
- 5 times a week (3)
- 4 times a week (4)
- 3 times a week (5)
- twice a week (6)
- once a week (7)

travel_walk_time How much time do you spend walking or bicycling for travel on a typical day? (Hours: minutes)

instructions **Recreational/leisure activities:**The next questions exclude the work and transport activities that you have already mentioned.
Now we would like to ask you about sports, fitness and recreational activities (leisure).

sports_vigorous Do you do any **vigorous-intensity** sports, fitness or recreational (leisure) activities that cause large increases in breathing or heart rate like [running or football] for at least 10 minutes continuously?

- Yes (1)
- No (2)

Skip To: sports_moderate If Do you do any vigorous-intensity sports, fitness or recreational (leisure) activities that cause... = No sports_vigorous_freq In a typical week, on how many days do you do **vigorous-intensity** sports, fitness or recreational (leisure) activities?

- Daily (1)
- 6 times a week (2)
- 5 times a week (3)
- 4 times a week (4)
- 3 times a week (5)
- twice a week (6)
- once a week (7)

sports_vigorous_time How much time do you spend doing **vigorous-intensity** sports, fitness or recreational activities on a typical day? (Hours: minutes)

____

sports_moderate Do you do any **moderate-intensity** sports, fitness or recreational (leisure) activities that cause a small increase in breathing or heart rate such as brisk walking, [cycling, swimming, volleyball] for at least 10 minutes continuously?

- Yes (1)
- No (2)

Skip To: instructions If Do you do any moderate-intensity sports, fitness or recreational (leisure) activities that cause... = No sports_moderate_freq In a typical week, on how many days do you do **moderate-intensity** sports, fitness or recreational (leisure) activities?

- Daily (1)
- 6 times a week (2)
- 5 times a week (3)
- 4 times a week (4)
- 3 times a week (5)
- twice a week (6)
- once a week (7)

sports_moderate_time How much time do you spend doing **moderate-intensity** sports, fitness or recreational activities on a typical day? (Hours: minutes)

instructions **Sedentary behaviour**:
The next question is about sitting or reclining at work, at home, getting to and from places, or with friends including time spent sitting at a desk, sitting with friends, traveling in car, train, reading, playing cards or watching television, but does not include time spent sleeping.

sedentary_time How much time do you usually spend **sitting or reclining** on a typical day? (Hours: minutes)____

End of Block: WHO STEPS: Physical Activity

Start of Block: WHO STEPS: History of Raised Blood Pressure

instructions The next questions ask about your **blood pressure**.

BP_raised Have you ever been told by a doctor or other health worker that you have **raised blood pressure** or hypertension?

- Yes (1)
- No (2)

Skip To: End of Block If Have you ever been told by a doctor or other health worker that you have raised blood pressure or... = No

BP_told_past_year Have you been **told** you have raised blood pressure in the **past 12 months**?

- Yes (5)
- No (6)

BP_meds_past_2wks In the **past two weeks**, have you taken any **medication** for raised blood pressure prescribed by a doctor?

- Yes (1)
- No (2)

End of Block: WHO STEPS: History of Raised Blood Pressure

Start of Block: WHO STEPS: History of Diabetes

instructions The next few questions ask about your diabetes.

diabetes_years How long have you had your type 2 diabetes diagnosis? [Enter number of years and/or months]

____

diabetes_test_method What **method** do you mainly use for **testing your own sugar level**?

- Blood or urine tests at the doctors office only (1)
- Blood glucose test strips read by eye at home (2)
- Blood glucose test strips read by meter at home (3)
- Urine glucose test strips at home (4)
- None (5)

Skip To: diabetes_own_BGmeter If What method do you mainly use for testing your own sugar level? = None diabetes_test_freq **How often have you tested** your blood sugar levels in the last month (4 weeks)?

- Never (1)
- Once a week or less (2)
- About 2-6 times a week (3)
- Once a day (4)
- 2 - 4 times a day (5)
- More than 4 times a day (6)

diabetes_own_BGmeter Do you own a blood glucose meter (for fingerprick tests)?

- Yes (1)
- No (2)

diabetes_care_freq Approximately how often do you visit a doctor for your diabetes?

- 5 or more times a year (1)
- 3 to 4 times a year (2)
- 1 or 2 times a year (3)
- Once very 2 or 3 years (4)
- Once every 5 years (5)
- Never (6)

diabetes_sick_days If you are employed, how many days have you taken off work due to illness in the last 3 months?

____

diabetes_cant_work Did your diabetes cause you to lose your job, quit or retire early?

- Yes (1)
- No (2)

diabetes_insurance How do you pay for your doctor prescribed medications?

- My insurance plan covers all my prescribed medications. (1)
- My insurance plan covers some of my prescribed medications. (2)
- I only have a hospital plan. I pay for all my medications privately. (3)
- Other, please say more (e.g. I don't have private health insurance) (4) ____

End of Block: WHO STEPS: History of Diabetes

Start of Block: WHO STEPS: History of Cardiovascular Diseases

instructions The next few questions ask about cardiovascular disease.

CVD_heart_attack Have you ever had a **heart attack or chest pain** from heart disease (angina) or a stroke (cerebrovascular accident or incident)?

- Yes (1)
- No (2)

CVD_disprin Are you currently taking **disprin** (aspirin) regularly to prevent or treat heart disease?

- Yes (1)
- No (2)

CVD_statins Are you currently taking **statins** which is medicine for high cholesterol (Lovastatin/Simvastatin/Atorvastatin or any other statin) regularly to treat heart disease?

- Yes (1)
- No (2)

Skip To: End of Block If Are you currently taking statins which is medicine for high cholesterol (Lovastatin/Simvastatin/A... = No

CVD_statins_years For how many years have you been prescribed statins?

____

End of Block: WHO STEPS: History of Cardiovascular Diseases

Start of Block: GPS Risk Preferences

GPS_willingness_risk In the following questions, drag and drop the slider to move it.
Please tell us, in general, how willing or unwilling you are to take risks, using a sliding scale, where **0 means you are “completely unwilling to take risks”** and **10 means you are “very willing to take risks.”**

	0	1	2	3	4	5	6	7	8	9	10
1 ()											

End of Block: GPS Risk Preferences

Start of Block: GPS Delay gratification, math ability, procrastination

GPS_patience We now ask you for your willingness to act in a certain way. Dragging the slider to **0 means “completely unwilling to do so,”** and moving the slider to **10 means “very willing to do so.”**
How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?

	0	1	2	3	4	5	6	7	8	9	10
1 ()											

GPS_good_at_math How good are you at math? Please indicate your answer on the sliding scale, choosing a letter from Grade A to Grade F. Drag and drop the slider to move it.
The Grade **A means “I am excellent at math,”** and a **F means “I am bad at math”**

1 (1)
2 (2)
3 (3)
4 (4)
5 (5)

GPS_procrastinate How well does the following statement describe you as a person? Drag and drop the slider to answer the question.
A **0 means “does not describe me at all,”** and a **10 means “describes me perfectly.”**

I tend to postpone tasks even if I know it would be better to do them right away.

	0	1	2	3	4	5	6	7	8	9	10
1 ()											

End of Block: GPS Delay gratification, math ability, procrastination

Start of Block: Subjective Wellbeing (NIDS Household Survey)

SWB Drag the slider on the right hand side of the face down or up to select how you feel about your life as a whole right now from **very unsatisfied** to **very satisfied**.

1 (1)
2 (2)
3 (3)
4 (4)
5 (5)

End of Block: Subjective Wellbeing (NIDS Household Survey)

Start of Block: Diabetes Distress Scale (DDS)

instructions Living with diabetes can sometimes be tough. There may be many problems and hassles concerning diabetes and they can vary greatly in severity. Problems may range from minor hassles to major life difficulties. Listed below are 17 potential problem areas that people with diabetes may experience. Consider the degree to which each of the 17 items may have distressed or bothered you DURING THE PAST MONTH and select the appropriate number. Please note that we are asking you to indicate the degree to which each item may be bothering you in your life DURING THE PAST MONTH, NOT whether the item is merely true for you. If you feel that a particular item is not a bother or a problem for you, you would select "1. Not a problem". If it is very troublesome to you, you might select "6. A very serious problem".

DDS To what degree has any of these things bothered you in the past month?

	1. Not a problem (1)	2. A slight problem (2)	3. A moderate problem (3)	4. A somewhat serious problem (4)	5. A serious problem (5)	6. A very serious problem (6)
1. Feeling that my doctor doesn't know enough about diabetes and diabetes care. (1)	○	○	○	○	○	○
2. Feeling that diabetes is taking up too much of my mental and physical energy every day. (2)	○	○	○	○	○	○
3. Not feeling confident in my day-to-day ability to manage diabetes. (3)	○	○	○	○	○	○
4. Feeling angry, scared and/or depressed when I think about living with diabetes. (4)	○	○	○	○	○	○
5. Feeling that my doctor doesn't give me clear enough directions on how to manage my diabetes. (5)	○	○	○	○	○	○
6. Feeling that I am not testing my blood sugar frequently enough. (6)	○	○	○	○	○	○

7. Feeling that I will end up with serious long-term complications, no matter what I do. (7)	○	○	○	○	○	○
8. Feeling that I am often failing with my diabetes routine. (8)	○	○	○	○	○	○
9. Feeling that friends or family are not supportive enough of self-care efforts (e.g. encouraging me to eat the "wrong" foods). (9)	○	○	○	○	○	○
10. Feeling that diabetes controls my life. (10)	○	○	○	○	○	○
11. Feeling that my doctor doesn't take my concerns seriously enough. (11)	○	○	○	○	○	○
12. Feeling that I am not sticking closely enough to a good meal plan. (12)	○	○	○	○	○	○
13. Feeling that friends or family don't appreciate how difficult living with diabetes can be. (13)	○	○	○	○	○	○

14. Feeling overwhelmed by the demands of living with diabetes. (14)	○	○	○	○	○	○
15. Feeling that I don't have a doctor who I can see regularly enough about my diabetes. (15)	○	○	○	○	○	○
16. Not feeling motivated to keep up my diabetes self-management. (16)	○	○	○	○	○	○
17. Feeling that friends or family don't give me the emotional support that I would like. (17)	○	○	○	○	○	○

End of Block: Diabetes Distress Scale (DDS)

Start of Block: Shopping and cooking

instructions The next few questions ask you about food and where and why you buy it.

buy_groceries Are you the one who usually buys groceries for your household?

○ Yes (1)

○ No (2)

hh_cook Are you usually the one who cooks for your household?

○ Yes (1)

○ No (2)

End of Block: Shopping and cooking

Start of Block: Short Grit Scale

S-Grit-Scale Please respond to the following 8 items from the Short Grit Scale. Be honest - there are no right or wrong answers!

	Very much like me (1)	Mostly like me (2)	Somewhat like me (3)	Not much like me (4)	Not like me at all (5)
1. New ideas and projects sometimes distract me from previous ones. (1)	◯	◯	◯	◯	◯
2. Setbacks don't discourage me. (2)	◯	◯	◯	◯	◯
3. I have been obsessed with a certain idea or project for a short time but later lost interest. (3)	◯	◯	◯	◯	◯
4. I am a hard worker. (4)	◯	◯	◯	◯	◯
5. I often set a goal but later choose to pursue a different one. (5)	◯	◯	◯	◯	◯
6. I have difficulty maintaining my focus on projects that take more than a few months to complete. (6)	◯	◯	◯	◯	◯
7. I finish whatever I begin. (7)	◯	◯	◯	◯	◯
8. I am diligent. (8)	◯	◯	◯	◯	◯

End of Block: Short Grit Scale

Start of Block: Self-efficacy and importance of behaviour change

confidence_to_change How confident are you that you will be able to make a change to your diet?

Answer on the sliding scale from 1 to 10, where
1 means "not confident at all" ... and **10 means "extremely confident".**

	1	2	3	4	5	6	6	7	8	9	10
How confident are you that you can make a change? ()											

importance_of_change How important is making this change to you?

Answer on the sliding scale from 1 to 10, where **1 means "not important at all"** ... and **10 means "extremely important".**

	1	2	3	4	5	6	6	7	8	9	10
How important is making this change to you? ()											

End of Block: Self-efficacy and importance of behaviour change

Start of Block: Stage of change

instructions The next question will ask about the stage of behaviour change you feel you are in. We really are interested in your honest answer. Remember there are no right or wrong answers and your responses are confidential.

stage_of_change Select the phase that most applies to you:

- ○ I am not doing a Low Carb High Fat (Banting) diet and have no intention of starting. (1)
- ○ I am not doing a Low Carb High Fat (Banting) diet yet but am committed to taking action within the next 6 months (2)
- ○ I am seriously considering starting a Low Carb High Fat (Banting) diet and will start within 30 days (3)
- ○ I am in the first month of following a Low Carb High Fat (Banting) diet (4)
- ○ I am on a Low Carb High Fat (Banting) diet and have been doing it for 3 months (5)
- ○ I am on a Low Carb High Fat (Banting) diet and have been doing it for 6 months and longer (6)

End of Block: Stage of change

D. Food Frequency Questionnaire

Survey Flow

Block: Introduction (2 Questions)
Standard: Breads (7 Questions)
Standard: Cereals and grains (7 Questions)
Standard: Fruit and Veg (9 Questions)
Standard: Animal protein (15 Questions)
Standard: Dairy (9 Questions)
Standard: Added fats (11 Questions)
Standard: Legumes, nuts and seeds (8 Questions)
Standard: Drinks (9 Questions)
Standard: Junk food (5 Questions)
Standard: Dessert, biscuits, sweets (10 Questions)

Page Break

Start of Block: Introduction

This is a Food Frequency Questionnaire (FFQ) asking about foods you have eaten in the past month.
It is important that you only answer about foods that you actually ate and just in the past month.
I will read out a list of foods and the portion units for each food. You will be asked to say on average how many of these portions you usually eat at a time and how times you have eaten them. Try to think about all the times that you have eaten the particular food, including in take-aways, as snacks, or as part of other dishes.
For example, I may say: 'A piece of chicken'. If you normally eat only one piece whenever you eat chicken and you eat chicken twice a week, you would answer: 'One piece, twice a week'. If you normally eat 3 pieces but only on one day of the week, you would answer: 'Three pieces, once a week'.
Another more complicated example may be 'A cup of non-starchy vegetables' and I would give examples such as 'broccoli, spinach, cabbage, peppers, salad greens etc'. Try to think of all the times you have eaten any type of non-starchy vegetables like these in the past month. So you would add up all the times you ate any of them in salads, soups, as snacks on their own, in curries etc. You would then think how many cups you usually ate each time. So some answers may be: 'half a cup, once a day' or '2 cups, 3 times a week' or 'none' if no starchy vegetables were eaten in the past month.
Try to be as honest as possible. Please try not to let your answers be affected by what you think you should have eaten. We want to know what you actually ate and will keep your answers strictly confidential. Everyone else besides me will only see a code next to your answers so they won’t know they are yours.
Take as much time as you like to think about your answers. Once we get going, it should be more clear what to do but please feel free to ask me questions at any time.
Shall we start?
There are 18 questions in this survey.

id Enter the code that was given to you by the researchers.

____

End of Block: Introduction

Start of Block: Breads

instructions Thinking back over just the past month, on average, how often did you eat the following foods? Remember to state the number of portions you usually ate at a time and how often you ate them.

---

breads[white] A slice of white bread or roll.

- None
- 1 or less per month
- 2 to 3 per month
- 1 - 2 per week
- 3 - 4 per week
- 5 - 6 per week
- 1 per day
- 2 - 3 per day
- More than 3 per day

Notes: Unless otherwise listed all the dropdown choices are the same for each of the following questions. We omit them here for brevity.

breads[brown] A slice of brown, wholewheat or rye bread or roll.

breads[roti] A roti.

breads[lowcarb] A slice of low carb bread (e.g. Heba or other)

breads[crackers] A savoury cracker or biscuit.

breads[pastry] A savoury pastry (e.g. croissant).

End of Block: Breads

Start of Block: Cereals and grains

cereals[cereal] A bowl of regular breakfast cereal or granola (e.g. All Bran, musli, Cornflakes etc.)

cereals[oats] A bowl of cooked porridge or oats.

cereals[lowcarb] A bowl of cooked (low carb) Heba pap.

cereals[pap] A bowl of regular cooked pap.

cereals[pasta] A cup of cooked pasta or noodles.

cereals[rice] A cup of cooked rice.

cereals[millet] A cup of other cooked grains (e.g. millet, cous cous, sorghum).

End of Block: Cereals and grains

Start of Block: Fruit and Veg

instructions For the following fruit and vegetables, try to think about all the times you have eaten them either alone, as salads or as parts of other dishes.

fruit&veg[potatoes] A medium white potato (excluding hot chips and crips).

fruit&veg[starchy] A cup of starchy vegetables (e.g. sweet potato, butternut, pumpkin, carrots, corn, peas etc.)

fruit&veg[nonstarchy A cup of non-starchy vegetables (e.g. broccoli, spinach, cabbage, peppers, salad greens etc.).

fruit&veg[olives] A small handful of olives.

fruit&veg[avocado] An avocado.

fruit&veg[berries] A small handful of berries.

fruit&veg[banana] A piece of any other type fruit (e.g. an apple, naartjie, orange, banana, grapes, mango etc.)

fruit&veg[driedfruit A small handful of dried fruit.

End of Block: Fruit and Veg

Start of Block: Animal protein

protein[red_meat] A palm sized piece of fresh red med e.g. beef, lamb or pork (including sausage but excluding dried and cured meats below)

protein[poultry] A piece of fresh poultry e.g. chicken, turkey etc.

protein[biltong] A small handful of biltong or droewors.

protein[processed] A slice of any type of cured/cold meats e.g. bacon, salami, sandwich ham, chicken roll, polony, viennas etc.

protein[offal] A cup of organ meats e.g. liver or kidneys.

protein[freshfish] A palm sized piece of fresh or frozen fish.

protein[canned_fish] A can of tinned fish e.g. pilchards, sardines, tuna etc.

protein[shellfish] A palm sized portion of shellfish (e.g. mussels, oysters).

protein[egg] An egg.

protein[shakes] A heaped table spoon of protein powder.

protein[bonebroth] A cup of bone broth or homemade stock.

Display This Question:

If A palm sized piece of fresh red med e.g. beef, lamb or pork (including sausage but excluding drie... != None

meat_fat When you ate red meat (beef, lamb, pork, game etc.) in the past month, how often did you also eat the fat with the meat?

- Never ate the fat
- Some of the time
- About half the time
- Most of the time
- Always ate the fat

Display This Question:

If A piece of fresh poultry e.g. chicken, turkey etc. != None

chicken_skin When you ate poultry (e.g. chicken), how often did you also eat the skin?

- Never ate the skin
- Some of the time
- About half the time
- Most of the time
- Always ate the skin

Display This Question:

If A piece of fresh poultry e.g. chicken, turkey etc. != None

chicken_breaded When you ate poultry (e.g. chicken) in the past month, how often was it breaded (crumbed), or battered (dipped in flour)?

- Never breaded or battered
- Some of the time
- About half the time
- Most of the time
- Always breaded or battered

Display This Question:

If A palm sized piece of fresh or frozen fish. != None fish_breaded When you ate fish, how often was it breaded (crumbed), or battered (dipped in flour)?

- Never breaded or battered
- Some of the time
- About half the time
- Most of the time
- Always breaded or battered

End of Block: Animal protein

Start of Block: Dairy

instructions In the past month, how often did you have:

dairy[wholemilk] A cup of milk (add up all milk in hot drinks, cereal, on own etc.).

diary[cream] A tablespoon of cream or sour cream.

dairy[yoghurt] A cup of yoghurt.

dairy[hard_cheese] A matchbox sized portion of hard cheese e.g. gouda, cheddar, edam, feta etc.

dairy[soft_cheese] A matchbox sized portion of soft cheese e.g. cottage cheese, cream cheese etc.

milk_full_cream When you drank milk in the past month, how often was it full cream?

- Never full cream
- Some of the time
- About half the time
- Most of the time
- Always full cream

yoghurt_double_cream When you ate yoghurt in the past month, how often was it full or double cream?

- Never full cream
- Some of the time
- About half the time
- Most of the time
- Always full cream

yoghurt_flavoured When you ate yoghurt in the past month, how often was it flavoured?

- Never flavoured
- Some of the time
- About half the time
- Most of the time
- Always flavoured

End of Block: Dairy

Start of Block: Added fats

instructions For the following added fats, try to think of how much you ate altogether including when used for cooking meals, on their own, or when added to prepared food.

fats[butter] A table spoon of real butter or ghee.

fats[margarine] A table spoon of margarine.

fats[lard] A table spoon of lard or animal fat (excluding fat as part of meat).

fats[olive_oil] A table spoon of olive oil.

fats[coconut_oil] A table spoon of coconut oil.

fats[canola] A table spoon of any other vegetable oil like sunflower, canola, soybean, corn, or palm oil.

fats[coconut_milk] A cup of coconut milk or coconut cream.

fats[salad_dressing] A table spoon of salad dressing (only if not included in the oils mentioned above)

fats[mayonnaise] A table spoon of mayonnaise

fats[condiments] A table spoon of condiment sauces added to food e.g. tomato ketchup, chutney, BBQ, chilli, mustard etc.

End of Block: Added fats

Start of Block: Legumes, nuts and seeds

legumes[baked_beans] A cup of baked beans.

legumes[beans_other] A cup of soy beans.

legumes[kidney_beans A cup of other beans e.g. kidney, butter etc

legumes[lentils] A cup of lentils.

legumes[peanuts] A small handful of peanuts or cashew nuts

legumes[nuts_other] A small handful of other nuts e.g. macadamia, pecans, walnuts, almonds, brazil etc.

legumes[nut_butter] A table spoon of nut butter e.g. peanut butter

legumes[seeds] A small handful of seeds

End of Block: Legumes, nuts and seeds

Start of Block: Drinks

drinks[coke] A can of regular soft drink e.g. coke, sprite, jive (340 ml)

drinks[diet_coke] A can of diet soft drink (340 ml)

drinks[juice] A glass of fruit juice or squash (300 ml)

drinks[energy] A can of energy or sports drink (340 ml)

drinks[milo] A cup of hot chocolate, milo or other milk drink (250 ml)

drinks[beer] A can of beer (340 ml)

drinks[wine] A glass of wine (175 ml)

drinks[cider] A can of alcoholic cooler or cider (340 ml) drinks[spirits] A shot of spirits e.g. whiskey, vodka, gin etc. (25ml)

End of Block: Drinks

Start of Block: Junk food

snacks[pizza] A medium pizza

snacks[samosa] A small samosa

snacks[hot_chips] A cup of hot chips or french fries

snacks[pie] A regular sized pie

snacks[crisps] A small packet of crisps or popcorn

End of Block: Junk food

Start of Block: Dessert, biscuits, sweets

sugar[dessert] A bowl of any kind of pudding dessert e.g. ice cream, chocolate mouse etc. but not cakes below.

sugar[cake] A medium sized piece of any kind of cake or sweet pastry e.g. cake, muffin, brownie, waffle, pancake, milk tart, doughnuts, koeksisters, chelsea bun etc.

sugar[biscuit] A sweet biscuit

sugar[rusk] A rusk

sugar[sweets] A packet of sweets (50g)

sugar[milk_choc] A small bar of milk chocolate (50 g)

sugar[dark_choc] A small bar of dark chocolate (50 g)

sugar[tsp_sugar] A teaspoon of sugar - on its own or added to foods and drinks e.g. in tea, coffee, cooked meals, cereal etc

sugar[honey_jam] A teaspoon of honey, syrup, jam, or nectar eg coconut nectar - on own or added to foods and drinks

sugar[non_caloric] A teaspoon of non-caloric sweeteners added to foods or drinks e.g. aspartemine, stevia, sucralose, xylitol etc. (excluding in food products such as diet sodas)

End of Block: Dessert, biscuits, sweets

E. COVID-19 and T2D Questionnaire, HREC REF: 310/2020

Survey Flow

Block: Introduction (2 Questions)
Standard: Contact with COVID-19 (8 Questions)
Standard: COVID-19 and access to medical care (6 Questions)
Standard: COVID-19 impact on productivity and lifestyle (16 Questions)
Standard: General feedback (3 Questions)

Page Break

Start of Block: Introduction

Q1.1
Dear study participant,

There is little information about how the COVID-19 pandemic has affected the type 2 diabetes community. Please answer the following questions. We want to know whether the COVID-19 pandemic has affected your access to medical care, any COVID-19 related challenges you may have faced, and any COVID-19 related fears you may have.

Remember your responses will be kept strictly confidential and will be analysed anonymously. Your honest answers are valuable to our research and may help other patients with diabetes adapt to lifestyle challenges of COVID-19.

This survey will take about 10 minutes. You can take as much time as you need.

If you have any concerns or questions please contact one of the researchers or your study physician. You can also consult government websites such as: https://sacoronavirus.co.za/

Sofia Monteiro (PhD Candidate): sofia.g.monteiro@gmail.com (072 253 6420)
Dr Kate Larmuth: kateus65@gmail.com

Q1.2 Please enter your Participant code:

____

End of Block: Introduction

Start of Block: Contact with COVID-19

Q2.1 The following questions ask about your contact with COVID-19.

Page Break

Q2.2 Have you ever been diagnosed with COVID-19?

- Yes (1)
- No (2)

Display This Question:

If Have you ever been diagnosed with COVID-19? = YesQ2.3 What COVID-19 symptoms did you have? Select all that apply.

- Cough (1)
- Fever (2)
- Sore throat (3)
- Shortness of breath (4)
- Tiredness (5)
- Aches and pains (6)
- Runny nose (7)
- Nasal congestion (8)
- Diarrhoea (9)
- Other (10) ____
- Asymptomatic (no symptoms) (11)

Display This Question:

If Have you ever been diagnosed with COVID-19? = Yes

Q2.4 When were you diagnosed with COVID-19? [Enter date]

Q2.5 Has anyone you know personally been diagnosed with COVID-19? (e.g. family, friend, neighbour, close colleague)

- Yes (1)
- No (2)

Q2.6 Has anyone you know personally died from COVID-19?

- Yes (1)
- No (2)

Q2.7 Have you had contact with anyone with diagnosed COVID-19 in the last 14 days?

- Yes (1)
- No (2)

Q2.8 Have you had symptoms such as cough, fever, sore throat, or shortness of breath in the last 14 days?

- Yes (1)
- No (2)

End of Block: Contact with COVID-19

Start of Block: COVID-19 and access to medical care

Q3.1 The following questions ask about whether COVID-19 has affected your access to medical care, or medications.

Q3.2 Has fear of COVID-19 ever prevented you from seeking medical care for your diabetes?

- Yes (1)
- No (2)

Q3.3 Has fear of COVID-19 ever prevented you from seeking medical care for other illnesses or dental problems?

- Yes (1)
- No (2)

Q3.4 Have you had any remote/online consultations with a doctor or healthcare provider in the last six months (e.g. phone call or video call)?

- Yes (1)
- No (2)

Q3.5 Has fear of COVID-19 ever prevented you from picking up your prescribed medications from the pharmacy?

- Yes (1)
- No (2)

Q3.6 Do you know someone who you could call to pick up your medications if you were not able to go yourself?

- Yes (1)
- No (2)
- Not applicable, the pharmacy delivers my medications to my house. (3)

End of Block: COVID-19 and access to medical care

Start of Block: COVID-19 impact on productivity and lifestyle

Q4.1 The following questions ask whether COVID-19 has affected your work, social life, family life and lifestyle choices.

Q4.2 Are you able to work from home?

- ○ Yes (1)
- ○ No (2)
- ○ Not applicable (3)

Q4.3 Was your occupation/job classified as an essential service during lockdown March 2020?

- ○ Yes (1)
- ○ No (2)
- ○ Not applicable (3)

Q4.4 Has the COVID-19 pandemic and/or lockdown affected your productivity?

- ○ Not at all (1)
- ○ It made me a lot less productive (2)
- ○ It made me a little less productive (3)
- ○ It made me a little more productive (4)
- ○ It made me much more productive (5)

Q4.5 Has the COVID-19 pandemic and/or lockdown led to a loss of employment for you?

- ○ Yes (1)
- ○ No (2)
- ○ Not applicable (e.g. retired) (3)
  - ○ Other (4) ____

Q4.6 Have you struggled financially since the COVID-19 pandemic started?

- Not more than usual (1)
- A bit more than usual (2)
- A lot more than usual (3)

Q4.7 Improving your blood sugar control, and reversing your diabetes is associated with better immunity and health outcomes. Does fear of COVID-19 play a role in your motivation to make healthy lifestyle choices?

- Not at all (1)
- COVID-19 plays a small role in my motivation to make healthy choices (2)
- COVID-19 plays a big role in my motivation to make healthy choices (3)

Q4.8 How stressed are you because of the COVID-19 pandemic?

- Not more stressed than usual (1)
- A little more stressed than usual (2)
- Much more stressed than usual (3)

Q4.9 Do you live alone?

- Yes (1)
- No (2)

Q4.10 Have you felt isolated, lonely or disconnected from other people in your life because of the COVID-19 pandemic?

- Not at all (1)
- A little more than usual (2)
- Much more than usual (3)

Q4.11 Are you currently being treated by a medical professional for anxiety or depression?

- ○ Yes (1)
- ○ No (2)

Q4.12 Has the COVID-19 pandemic and/or lockdown caused you to change your food choices?

- ○ Not at all (1)
- ○ I eat more unhealthy food than usual (2)
- ○ I eat less unhealthy food than usual (3)

Q4.13 Has the COVID-19 pandemic and/or lockdown affected how you or your household do your grocery shopping? Tick all answers that apply to you.

- ▢ Not at all (1)
- ▢ More online shopping (2)
- ▢ Less frequent shopping at supermarkets (3)
- ▢ More frequent shopping at supermarkets (4)
- ▢ Less frequent grocery shopping in general (5)
- ▢ More frequent grocery shopping in general (6)

Q4.14 Do you order food from restaurants or cafes to be delivered to your home?

- Never (1)
- Less than once per month (2)
- 1 to 3 times per month (3)
- Once or twice per week (4)
- 3 or more times per week (5)

Q4.15 Have you eaten at a restaurant or cafe in the last two weeks?

- Yes (1)
- No (2)

Q4.16 Has the COVID-19 pandemic and/or lockdown affected your physical activity?

- Not at all (1)
- I am more physically active than usual (2)
- I am less physically active than usual (3)

End of Block: COVID-19 impact on productivity and lifestyle

Start of Block: General feedback

Q5.1 Please tell us about any other challenges that COVID-19 has caused for you personally or professionally.

____

Q5.2 What are your fears or concerns related to COVID-19?

____

Page Break

Q5.3 **What happens if I have contact with someone with diagnosed COVID-19 during the study?**

You must self-quarantine. This means that you should stay at home for 14 days and monitor yourself for symptoms such as a cough, fever, sore throat and difficulty breathing. Stay away from other older people and people with a compromised immune system or underlying condition such as high blood pressure or diabetes.

You must please contact one of the researchers and your study doctor immediately via phone call, text or email. This will let us monitor your health closely. We want to ensure that you get the best possible care. Your wellbeing is our priority.

If a health worker thinks that you may have COVID-19 they may refer you to a health facility to be tested.

For more information contact government websites such as https://sacoronavirus.co.za/

End of Block: General feedback

F. Low Carbohydrate Dietary Recommendations by Dr N Wellington

[Phase 2: Physician reviews Libre Pro data from past two weeks with the patient and introduces them to low carb eating guidelines for next three months]

Introduction

All carbohydrates are digested to basic building blocks of glucose and fructose, and these are absorbed into the bloodstream. High levels of blood sugar (called hyperglycaemia), of levels over 5mmol/L, stimulate the pancreas to release insulin. Insulin helps to keep blood sugar (glucose) levels as close to the normal range as possible (4-6mmol/L before eating and up to 7.8mmol/L after meals). Insulin stimulates the liver and muscles to absorb glucose, converting it to glycogen, and converting the extra glucose to triglycerides for storage as fat. Eventually, glucose (and fructose) is stored as fat and excess fat can eventually cause insulin resistance and diabetes. The lifestyle treatment of diabetes revolves around an understanding of carbohydrates and how to control their intake. It is also important to understand that fat is the primary source of energy for our bodies, and this means that it is necessary to preferentially eat dietary fat (mainly as saturated fat) i.e. to eat a diet low in carbohydrates and higher in fat. Fat is satiating - it fills you up for longer - and generally means the overall calorie intake is lower than on a high carbohydrate low fat diet, where people get hungry quickly and tend to eat more.

Common carbohydrates and their approximate values

1 tsp sugar = 5g of carbohydrates
1 slice bread ≈ 15g
1 plate cereal (eg 2 Weetbix biscuits) ≈ 30g
250ml (1 cup) Coca cola = 28g
250ml fruit juice = 32g
100g spaghetti = 70g
100g potato/sweet potato = 20g/32g
1 cup cooked rice = 20g
1 medium apple = 20g
1 large banana = 30g (note this is equivalent to about 2 slices of bread!!)
100ml sweetened yoghurt = 15g
250ml sweetened yoghurt = 37.5g
100g slab chocolate = 60g

More lists of carbohydrates can be found at www.carbohydrate-counter.org the book 'The drinking man's diet', and on all nutrition labels (see attached picture for Weetbix, oats and peanut butter.)

For those with trying to lose weight it is recommended that you eat no more than 50g/day. **For patients with diabetes you should aim to eat less than 25g per day.** Each person needs to find their level. For patients with diabetes it is critical to regularly **check their glucose levels before and about 90 minutes after meals** and to adjust insulin and medication accordingly. This will need to be done in conjunction with your doctor or diabetes educator. For dietary advice and help with planning meals it is recommended that you consult with a dietician who is familiar with low carbohydrate lifestyles.

[Show examples of typical nutrition labels]

What should you avoid and what can you eat?
As suggested above, avoid all sugars, sweets, confectionaries, breads, cereals and low fat foods. Low fat foods tend to have sugar added to make them taste better. Avoid all processed foods which generally mean foods that come in a box. While whole foods like fruit and vegetables contain micronutrients that your body can use, it is advisable to cut them down or avoid if it means exceeding your daily target of carbohydrates.

Foods to eat include the following:
Eggs
Full fat meats
Chicken with the skin
Fish
Leafy green vegetables
Avocado
Olive oil
Coconut and coconut oil
Fatty nuts like almonds and macadamia nuts
Cheese (unprocessed)
Full cream milk
Butter
Mushrooms

There are many resources for this way of eating:
You can visit www.dietdoctor.com, or www.realmealrevolution.co.za, or www.helgavan.com and click on the LCHF drop down tab.
Good books to read include '**Real Meal Revolution**', 'The Art and Science of Low Carbohydrate Living', 'The Diet Delusion', and 'New Atkins, New You'. For patients with diabetes I highly recommend Dr Richard Bernstein's book 'The Diabetes Solution' and 'Low carb solutions for diabetics'

Information for patients with diabetes

It is possible to reverse type 2 diabetes, but it will take a concerted effort and motivation. The following points are critical to this.

1. Know your carbohydrates i.e. which foods contain them and how much
2. Reduce your carbohydrates to under 25g per day
3. Monitor, monitor, monitor! Ensure that you monitor your blood glucose levels before and after meals to ensure you are eating correctly, and you are taking the correct medication (esp if you are taking insulin)
4. Eat whole foods (unprocessed) with healthy fats
5. Stay motivated and don’t give up.

G. Traffic lights lists for dietary advice

G

THE GREEN LIST

THE GREEN FOOD LIST IS THE ONLY LIST THAT YOU CAN EAT FROM ON A DAILY BASIS. THESE ARE THE FOODS THAT ARE NUTRITIOUS, LOW IN CARBS PER PORTION AND EXTREMELY HEALTHY. PRACTICING PORTION CONTROL IS STILL IMPORTANT WHEN EATING FROM THIS LIST

THE BANTING POCKET GUIDE QUICK REFERENCE LISTS

For a more extensive list of foods, including the macro nutrient breakdown, please refer to our book The Banting Pocket Guide.

EAT BETTER SOUTH AFRICA!

THE BANTING POCKET GUIDE

ANIMAL PROTEIN
All eggs
Beef / veal
Mutton / Lamb
Pork
Venison/game
Ostrich
All Poultry :
Chicken
Duck
Turkey
Offal:
Brain
Brawn
Tripe
Trotters
Liver, heart, kidneys
Tongue
Chicken feet/heads/ gizzards
Naturally cured meats and sausage
Bacon
Chorizo
Pancetta
Salami
Sausage

SEAFOOD
Fish – fresh and canned in brine
Calamari
Crab
Oysters
Prawns

DAIRY*
Amasi
Buttermilk
Coconut milk
Cow's milk – full cream
Cheese hard and soft
Cottage cheese
Cream cheese
Cream – fresh/sour
Yoghurt full cream/Greek
*Using dairy products may stall weight loss in some people.

FATS AND OILS
Beef tallow
Butter
Duck fat
Ghee
Lard
Almond oil
Avocado oil
Coconut oil
Olive oil
Macadamia nut oil

FLAVOURING & CONDIMENTS
All natural herbs and spices are acceptable if they do not contain sugars and chemical additives. Includes
Aniseed, Basil, Capers, Caraway seed, Cardamom, Chillies, Cinnamon, Coriander, Curry powder, Dill, Fennel, Garlic, Ginger, Horseradish, Marjoram, Masala, Organum, Paprika, Parsley, Pepper, Peppermint, Rosemary, Sage, Thyme, Turmeric. Vinegar, including Apple cider.

BEVERAGES
Coffee (100% pure coffee)
Tea- including green tea and Rooibos
Water, soda water, sparkling mineral water.

NUTS & SEEDS
Almond, Brazil nuts, Coconut, Macadamia nut, Pecans, Pine nuts, Pistachio nuts, Walnuts.
Chia seed, Flax seed, Linseed, Pumpkin seed, Sesame seed, Sunflower seed
HEBA, Psyllium husk

SWEETENERS
Xylitol granules
Erythritol granules
Stevia powder
NOTE: We do not recommend artificial sweeteners of any kind. It is our opinion that if you want to stay lean and healthy for the rest of your life you need to avoid all foods that taste sweet. The desire to eat sweet foods is the addiction that drives poor food choices leading to obesity and ill health.

VEGETABLES
Amaranth/marog
Artichokes - globe
Asparagus
Aubergine
Broccoli
Brussels sprouts
Cabbage
Calabash / gourd
Cauliflower
Celery
Chives
Collards
Cucumber
Endive
Gherkins (dill, sugar free)
Green beans
Kale
Kohlrabi
Leek - boiled
Lettuce
Mixed frozen vegetables (cauliflower, carrot, green beans)
Mushrooms
Okra
Onion
Pepper- green, red, yellow
Pumpkin
Radish
Sauerkraut
Seaweed
Sousou/ chayote
Spinach
Spring onion
Squash - gem, hubbard,
Squash – baby marrow
Sugarsnap peas
Tomato
Turnip
Waterblommetjies
Wild rocket

FRUITS
Avocado
Olives

THE NOAKES FOUNDATION

THE ORANGE LIST

THE ORANGE FOOD LIST IS FOR PEOPLE WHO HAVE REACHED THEIR GOAL WEIGHT AND WANT TO INCLUDE SOME VEGETABLES AND BERRIES ON THIS LIST, OR FOR THOSE WHO ARE NOT SENSITIVE TO CARBOHYDRATES AND CAN TOLERATE THESE VEGETABLES AND FRUITS. THIS LIST IS ALSO FINE FOR AN OCCASIONAL SWEET TREAT, **BUT ONLY ONCE YOU HAVE REACHED YOUR GOAL WEIGHT.** WE HAVE INSERTED THE CARB COUNT HERE SO YOU CAN BE AWARE OF THE HIGHER CARB VALUES:

THE BANTING POCKET GUIDE QUICK REFERENCE LISTS

For a more extensive list of foods, including the macro nutrient breakdown, please refer to our book The Banting Pocket Guide.

VEGETABLES per 100g

Artichoke 14.3g
Beetroot 7.96g
Carrot boiled 5.3g
Carrot raw 6.4g
Leek – raw 12.4g
Parsnip 13.01g
Squash – Butternut 10.2g
Sweet potato – orange 17.4g
Sweet potato – white 15.1g
Tomato – sundried (per 25g) 10.9g

PROTEINS

Abalone (per 125g) 14.6g
Mussel (per 100g) 7.4g
Perlemoen (per 125g) 14.6g
Snails (per 75g) 11.6g

FRUIT per 50g

Apple 6.5g
Apricot 6.5g
Banana 9.4g
Blackberries 4.3g
Blueberries 6.1g
Cranberries 3.8g
Figs 6.8g
Gooseberries 6.0g
Granadilla 6.5g
Grape 7.4g
Guava 7.7g
Kiwifruit 6.5g
Kumquat 4.7g
Lemon 7.0g
Lime 7.7g
Litchi 8.6g
Melon green flesh 4.5g
Melon orange flesh 4.1g
Naartjie 5.0g
Nectarine 5.2g
Orange 4.6g
Papaya 4.6g
Pawpaw 4.3g
Peach 4.3g
Pear 7.2g
Pineapple 6.1g
Plum 5.5g
Raspberries 2.6g
Strawberries 3.0g
Watermelon 3.0g
Youngberries 2.15g

SWEETENERS

Honey (per 5g) 4g

NUTS per 30g

Betel nut 16.1g
Chestnut 13.3g
Cashew nut 8.9g

THE BANTING POCKET GUIDE

EAT BETTER SOUTH AFRICA!

THE NOAKES FOUNDATION

10 BASIC RULES OF BANTING

1. Banting is about eating when hungry and stopping when satisfied.

2. Eat clean, fresh, real food. Real food goes off and has a very short shelf life. Do not eat processed or pre-packaged foods.

3. Make sure that you include fats, proteins and healthy carbs in all your meals, whether you are eating three meals a day or only two. Meals must be nutrient dense and well balanced.

4. Do not eat more than three meals a day; there is no rule dictating which time of the day you should eat or that you have to eat all three meals.

5. Do not have sweeteners in your coffee or tea; go cold turkey if you want to see results.

6. Drink water throughout the day, but only when you are thirsty.

7. Make sure you are getting enough vitamins and minerals. If you experience energy loss in the beginning, you may supplement.

8. Do not drink any fizzy drinks, fruit juices or 'slimming' drinks, not even if they claim to be sugar free. They all contain artificial sweeteners and additives that can have a negative effect on your health and weight.

9. Do not snack between meals unless you are really hungry. Snacking between meals can lead to weight gain.

10. What works for you may not necessarily work for others. We are all unique.

THE RED LIST

THE RED FOOD LIST ITEMS MUST BE AVOIDED AT ALL COSTS. WE DON'T EVEN RECOMMEND THESE FOODS AS A ONCE-IN-A-WHILE TREAT, AS THEY ARE HIGHLY PROCESSED AND CONTAIN UNHEALTHY ADDITIVES AND CHEMICALS.

RED ITEM FOODS WILL ALMOST ALWAYS CONTAIN INGREDIENTS THAT ARE HARD TO PRONOUNCE

THE BANTING POCKET GUIDE QUICK REFERENCE LISTS

For a more extensive list of foods, including the macro nutrient breakdown, please refer to our book The Banting Pocket Guide.

ALL PRODUCTS CONTAINING ANY OF THESE INGREDIENTS

Atta (chapatti flour)
Breaded or battered foods
Cake flour, Chickpea flour
Corn flour, Durum (wheat)
Malt, Matzo meal, Modified wheat starch
Oatmeal, Oat bran, Whole oats
Potato starch, Rice flour
Semolina, Sorghum, Soy flour
Dried beans, Couscous
Lentils, Pasta, Polenta
Rice, Samp
Split peas, Stampkoring
Wheat germ, Wheat starch

BEVERAGES

Canned coffee – generally containing other ingredients like dextrose, etc
Tea with added artificial ingredients
Fizzy drinks including diet or lite drinks
Cordials, Fruit drinks, Fruit juice
Shakes of any kind
Energy drinks

ALCOHOL

Beer
Ciders
Dessert wine
Liqueurs & Shooters

DAIRY

All low fat/ fat free products
Cheese spreads, Processed cheese
Canned cream, Dessert cream
Coffee creamer
Condensed milk
Custard
Flavoured yoghurt
Ice cream
Powdered milk, Rice milk, Soy milk

FATS AND OILS

All commercial fat spreads/ margarine
Flavoured butters
Canola oil, Corn oil
Cottonseed oil, Grapeseed oil,
Soybean oil,Sunflower oil

SAUCES AND DRESSINGS

All commercial sauces and dressings
Barbeque sauce, Cook in sauce, Marinades, Mustard sauce, Peri-peri sauce, Pasta sauce, Salad creams and dressings
Tomato sauce
Sweet sauces

FAST FOOD AND TAKEAWAYS

Burgers, Hot dogs, Spare ribs,
Crumbed chicken or fish
Fries, Wraps, Pizza, Hotdogs

MEAT AND FISH

All meat that has been cured with sugar and/or marinated meats with added ingredients
Corned meat
Cold processed meats, e.g. sandwich ham/ham/chicken/beef, etc generally found at the deli
Crumbed/battered meat, e.g. crumbed chicken, hamburger patties, chicken nuggets, meat pies, readymade meals, meat free products (soy), fish bakes, crumbed fish fingers
Pilchards in tomato sauce
Tuna in vegetable oil

FRUIT AND VEGETABLES

Dried fruit – all varieties
Legumes
Corn
Potatoes

SWEETENERS

Agave
Aspartame
Blackstrap molasses
Cane sugar, Beet sugar
Castor sugar
Coconut sugar, Date sugar
Carob syrup, Corn syrup, Maple syrup
Dextrose
Fructose
Glucose
Maltitol
Saccharin
Sorbitol
Sucralose
Table sugar
Tapioca sugar
Treacle

H. Coaching Agreement

OVERVIEW

Welcome to Lauren Kim Wellness, a professional Coaching practice. This document constitutes a contract between us (the "Agreement"). You should read it carefully and raise any questions and concerns that you have before you sign it.

SERVICES

The services provided by Lauren Kim Wellness include Coaching on topics decided jointly with you, the client. Coaching is a partnership (defined as an alliance, not a legal business partnership) between the Coach and the Client in a thought-provoking and creative process that inspires the client to maximize personal potential. It is designed to facilitate the creation/development of personal health and wellness goals and to develop and carry out a strategy/plan for achieving those goals. Coaching utilizes strategic planning, values clarification, brainstorming, motivational interviewing, and other coaching techniques.

FEEDBACK

If, at any time, you feel that your needs are not being met, or you are not getting what you want out of the coaching please tell me, so we can discuss your needs and adjust your coaching program as needed. We will continue to work on the goals that you define unless you want to stop, which we will do whenever you ask.

SESSION TIME

Coaching is scheduled at the mutual convenience of the Coach and the Client. The day and time for the next call will be scheduled at the close of each coaching session.

CANCELLATIONS

I ask that you give 24 hours prior notice if you need to cancel or change the time of an appointment. The Coach will make reasonable efforts to reschedule sessions that are cancelled in a timely manner.

TERMINATION

Either party may end the coaching relationship by providing the other party with a one-week written notice, which may be transmitted by email.

CONFIDENTIALITY

I protect the confidentiality of the communications with my clients as described by the International Coach Federation code of ethics. I will only release information about our work to others with your written permission, or if I am required to do so by a court order or similar mandate. It is impossible to fully protect the confidentiality of information which is transmitted electronically. This is particularly true of email and information stored on computers connected to the Internet and if you use a cell phone.

NATURE OF THE RELATIONSHIP

The client is solely responsible for creating and implementing his/her own physical, mental and emotional well-being, decisions, choices, actions, and results arising out of or resulting from the coaching relationship and his/her coaching calls and interactions with the Coach. As such, the Client agrees that the Coach is not and will not be liable or responsible for any actions or inaction or for any direct or indirect result of any services provided by the Coach. The Client has been made aware that the coaching relationship is in no way to be construed as psychotherapy, psychological counselling, or any type of therapy. In the event the client feels the need for professional counselling or therapy, it is the responsibility of the Client to seek a licensed professional. The Client also understands coaching does not prevent, cure, or treat any mental disorder or medical disease.
The coach is available on WhatsApp and email and will respond to correspondence within normal business hours Monday- Friday or at the coach's discretion outside of those hours.

MUTUAL NON-DISCLOSURE

The Coach and Client mutually recognize that they may discuss future plans, business affairs, customer lists, financial information, job information, goals, personal information, and other private information. The Coach will not voluntarily communicate the Client's information to a third party. In order to honour and protect the Coach's intellectual property, the Client likewise agrees not to disclose or communicate information about the Coach's practice, materials, or methods to any third parties.

DISPUTE RESOLUTION

Any controversy or claim arising out of or relating to this agreement, or the breach of this agreement, shall be settled by arbitration, which will occur via telephone by an arbitrator that we mutually agree upon. The costs of the arbitration shall be borne by the losing party.

AUTHORIZATION FOR RELEASE OF INFORMATION FOR PROFESSIONAL CREDENTIALIALING

I____________________ give permission to ______________________ to use my name, contact information, and hours of coaching towards any of his credentialing processes as a coach. I further give permission for those credentialing agencies to contact me for verification of information he provides. This permission will remain in effect until specifically cancelled by me.

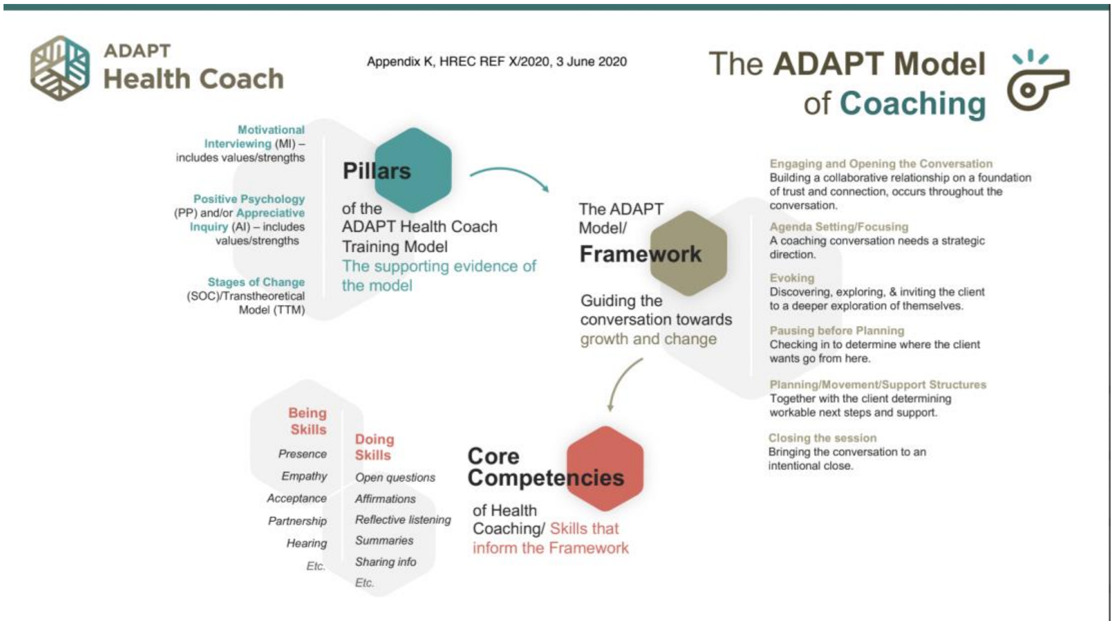
____

Your signature below indicates that you have read the information in this document ("Coaching Agreement and Informed Consent) and any Attachments, and agree to abide by its terms during our professional Coaching relationship.

Client_______________________________ Date _____________

Coach_______________________________ Date _____________

I. ADAPT Model of Health Coaching



J. Endline Questionnaire HREC REF: 310/2020

Survey Flow

Block: Introduction (2 Questions)
Standard: Self-report changes in habits (5 Questions)
Standard: Subjective Wellbeing (NIDS Household Survey) (1 Question)
Standard: Diabetes Distress Scale (DDS) (2 Questions)
Standard: Feedback (1 Question)

Page Break

Start of Block: Introduction

Q1.1 **Now that we are reaching the end of the trial, we would like to follow up with you about your experience participating in the diabetes behaviour change study.**

Please answer the following questions as honestly as you can. The questionnaire will take approximately 15 minutes but you can take more time if you want to. There is an unstructured text box at the end of this survey which you can use to share feedback with the researchers. Your answers are important to us.

Q1.2 Please enter your Participant code:

____

End of Block: Introduction

Start of Block: Self-report changes in habits

Q2.1 We would like to ask you about any changes in your habits during the last month. Compared to before, were there any changes in your habits?

Q2.2 Compared to before, did you **smoke** less, about the same or more tobacco products, such as cigarettes, cigars or pipes during the past 4 weeks?

○ Less (1)

○ About the same (2)

○ More (3)

○ Not applicable. I don't smoke. (4)

Q2.3 Compared to before, did you **drink** less, about the same, or more during the past 4 weeks?

- ○ Less (1)
- ○ About the same (2)
- ○ More (3)
- ○ Not applicable. I don't drink. (4)

Q2.4 Compared to before, was your **physical activity** less, about the same, or more during the past 4 weeks?

- ○ Less (1)
- ○ About the same (2)
- ○ More (3)

Skip To: End of Block If Compared to before, was your physical activity less, about the same, or more during the past 4 we... = About the same

Q2.5 How did your **physical activity** change? Use the space provided to note down any changes you made.

____

End of Block: Self-report changes in habits

Start of Block: Subjective Wellbeing (NIDS Household Survey)

Q3.1 Using a scale of 1 to 10 where 1 means “very dissatisfied” and 10 means “very satisfied”, how do you feel about your life as a whole right now?

- 1 "Very dissatisfied" (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)
- 8 (8)
- 9 (9)
- 10 "Very satisfied" (10)

End of Block: Subjective Wellbeing (NIDS Household Survey)

Start of Block: Diabetes Distress Scale (DDS)

Notes: Refer to Baseline Questionnaire for DDS – this is a repeat measurement.

End of Block: Diabetes Distress Scale (DDS)

Start of Block: Feedback

Q5.1 Please write down any feedback about your experience participating in the study that you would like the researchers to know. This might be about the CGM, your diet, challenges you face, tips for success, appointment feedback, support feedback, your doctor, your nurse or anything else you'd like to tell us. Your comments may be positive or negative. We are really interested in your honest feedback. This will help other patients like you in the future.

____

End of Block: Feedback

6. REFERENCES

Alberti, K. G. M., Zimmet, P., & Shaw, J. (2005). The metabolic syndrome—a new worldwide definition. *The Lancet, 366*(9491), 1059-1062.

Allen, N. A., Fain, J. A., Braun, B., & Chipkin, S. R. (2009). Continuous Glucose Monitoring in Non–Insulin-Using Individuals with Type 2 Diabetes: Acceptability, Feasibility, and Teaching Opportunities. *Diabetes Technology & Therapeutics*, *11*(3), 151–158. https://doi.org/10.1089/dia.2008.0053

Athinarayanan, S., Adams, R., Hallberg, S., McKenzie, A., Bhanpuri, N., & Campbell, W. et al. (2019). Long-Term Effects of a Novel Continuous Remote Care Intervention Including Nutritional Ketosis for the Management of Type 2 Diabetes: A 2-Year Non-randomized Clinical Trial. *Frontiers In Endocrinology*, *10*. https://doi.org/10.3389/fendo.2019.00348

Battelino, T., Danne, T., Bergenstal, R. M., Amiel, S. A., Beck, R., Biester, T., … Phillip, M. (2019). Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range. *Diabetes Care*, *42*(8), 1593–1603. https://doi.org/10.2337/dci19-0028

Bazzano, L. A., Hu, T., Reynolds, K., Yao, L., Bunol, C., Liu, Y., … He, J. (2014). Effects of Low-Carbohydrate and Low-Fat Diets. *Annals of Internal Medicine*, *161*(5), 309. https://doi.org/10.7326/M14-0180

Block, J. M. (2008). Continuous Glucose Monitoring: Changing Diabetes Behavior in Real Time and Retrospectively. *Journal of Diabetes Science and Technology*, *2*(3), 484–489. https://doi.org/10.1177/193229680800200320

Cadario, R., & Chandon, P. (2019). Which Healthy Eating Nudges Work Best? A Meta-Analysis of Field Experiments. *Marketing Science*, mksc.2018.1128. https://doi.org/10.1287/mksc.2018.1128

Cade, J., Thompson, R., Burley, V., & Warm, D. (2002). Development, validation and utilisation of food-frequency questionnaires–a review. *Public Health Nutrition*, *5*(4), 567–587. Retrieved from https://www.cambridge.org/core/journals/public-health-nutrition/article/development-validation-and-utilisation-of-foodfrequency-questionnaires-a-review/463EFE9970053E8BD922CC88F52E6244

Carlson, A. L., Mullen, D. M., & Bergenstal, R. M. (2017). Clinical Use of Continuous Glucose Monitoring in Adults with Type 2 Diabetes. *Diabetes Technology & Therapeutics*, *19*(S2), S-4-S-11. https://doi.org/10.1089/dia.2017.0024

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. https://doi.org/10.1016/J.JEBO.2011.08.009

Charness, G., Cobo-Reyes, R., Eyster, E., Sánchez, Á., & Sutter, M. (2021). Improving healthy eating in children: Experimental evidence. *MPI Collective Goods Discussion Paper.*

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, *138*, 271–281. https://doi.org/10.1016/J.DIABRES.2018.02.023

Dunn, T. C., Xu, Y., Hayter, G., & Ajjan, R. A. (2018). Real-world flash glucose monitoring patterns and associations between self-monitoring frequency and glycaemic measures: A European analysis of over 60 million glucose tests. *Diabetes Research and Clinical Practice*,

137, 37–46. https://doi.org/10.1016/j.diabres.2017.12.015
Ehrhardt, N., & Al Zaghal, E. (2019). Behavior Modification in Prediabetes and Diabetes: Potential Use of Real-Time Continuous Glucose Monitoring. *Journal of Diabetes Science and Technology*, *13*(2), 271–275. https://doi.org/10.1177/1932296818790994
Falk, A. ;, Becker, A. ;, Dohmen, T. ;, Huffman, D. B. ;, & Sunde. (2016). *The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences*. Retrieved from http://hdl.handle.net/10419/130357www.econstor.eu
Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. https://doi.org/10.1093/qje/qjy013
Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S., & Staake, T. (2020). *Complementarities in Behavioral Interventions: Evidence from a Field Experiment on Energy Conservation* (No. crctr224_2020_149). University of Bonn and University of Mannheim, Germany.
Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.
Finley, C. E., Barlow, C. E., Greenway, F. L., Rock, C. L., Rolls, B. J., & Blair, S. N. (2007). Retention rates and weight loss in a commercial weight loss program. *International Journal of Obesity*, *31*(2), 292–298. https://doi.org/10.1038/sj.ijo.0803395
Gouda, H., Charlson, F., Sorsdahl, K., Ahmadzada, S., Ferrari, A., & Erskine, H. et al. (2019). Burden of non-communicable diseases in sub-Saharan Africa, 1990–2017: results from the Global Burden of Disease Study 2017. The Lancet Global Health, 7(10), e1375-e1387. https://doi.org/10.1016/s2214-109x(19)30374-2
Hall, H., Perelman, D., Breschi, A., Limcaoco, P., Kellogg, R., McLaughlin, T., & Snyder, M. (2018). Glucotypes reveal new patterns of glucose dysregulation. *PLOS Biology*, *16*(7), e2005143. https://doi.org/10.1371/journal.pbio.2005143
Hallberg, S. J., Gershuni, V. M., Hazbun, T. L., Athinarayanan, S. J., Hallberg, S. J., Gershuni, V. M., … Athinarayanan, S. J. (2019b). Reversing Type 2 Diabetes: A Narrative Review of the Evidence. *Nutrients*, *11*(4), 766. https://doi.org/10.3390/nu11040766
Hallberg, S., McKenzie, A., Williams, P., Bhanpuri, N., Peters, A., & Campbell, W. et al. (2018). Author Correction: Effectiveness and Safety of a Novel Care Model for the Management of Type 2 Diabetes at 1 Year: An Open-Label, Non-Randomized, Controlled Study. Diabetes Therapy, 9(2), 613-621. https://doi.org/10.1007/s13300-018-0386-4
Hallsworth, M. (2017). Rethinking public health using behavioural science. *Nature Human Behaviour*, *1*, 612. Retrieved from https://www.nature.com/articles/s41562-017-0188-0
Hashimoto, Y., Fukuda, T., Oyabu, C., Tanaka, M., Asano, M., Yamazaki, M., & Fukui, M. (2016). Impact of low-carbohydrate diet on body composition: meta-analysis of randomized controlled studies. *Obesity Reviews*, *17*(6), 499–509. https://doi.org/10.1111/obr.12405
Hendrickx, H., McEwen, B., & Ouderaa, F. (2005). Metabolism, mood and cognition in aging: The importance of lifestyle and dietary intervention. Neurobiology Of Aging, 26(1), 1-5. https://doi.org/10.1016/j.neurobiolaging.2005.10.005
International Diabetes Federation. (2017). *IDF Diabetes Atlas, 8th edn. International Diabetes Federation, Brussels, Belgium*. Retrieved from http://diabetesatlas.org/resources/how-to-cite.html
Jeffery, R. W., Epstein, L. H., Wilson, G. T., Drewnowski, A., Stunkard, A. J., & Wing, R. R.

(2000). Long-term maintenance of weight loss: Current status. *Health Psychology*, *19*(1, Suppl), 5–16. https://doi.org/10.1037/0278-6133.19.Suppl1.5

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux. Retrieved from https://content.taylorfrancis.com/books/download?dac=C2018-0-82622-6&isbn=9781912453207&format=googlePreviewPdf

Kelly, M., & Barker, M. (2016). Why is changing health-related behaviour so difficult?. Public Health, 136, 109-116. https://doi.org/10.1016/j.puhe.2016.03.030

List, J., & Samek, A. (2017). The behavioralist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption. *Journal of Health Economics*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167629614001398

McKenzie, A., Hallberg, S., Creighton, B., Volk, B., Link, T., & Abner, M. et al. (2017). A Novel Intervention Including Individualized Nutritional Recommendations Reduces Hemoglobin A1c Level, Medication Use, and Weight in Type 2 Diabetes. JMIR Diabetes, 2(1), e5. https://doi.org/10.2196/diabetes.6981

Meng, Y., Bai, H., Wang, S., Li, Z., … Q. W.-D. research and, & 2017, U. (2017). Efficacy of low carbohydrate diet for type 2 diabetes mellitus management: a systematic review and meta-analysis of randomized controlled trials. *Diabetes Research and Clinical Practice*, *131*, 124–131. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168822717304023

Nansel, T. R., Lipsky, L. M., & Liu, A. (2016). Greater diet quality is associated with more optimal glycemic control in a longitudinal study of youth with type 1 diabetes. *The American Journal of Clinical Nutrition*, *104*(1), 81–87. https://doi.org/10.3945/ajcn.115.126136

Noakes, T., & Windt, J. (2017). Evidence that supports the prescription of low-carbohydrate high-fat diets: a narrative review. *British Journal of Sports Medicine*, *51*(2), 133–139. Retrieved from https://bjsm.bmj.com/content/51/2/133?utm_content=buffer6196e&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

Nojilana, B., Bradshaw, D., Pillay-van Wyk, V., Msemburi, W., Laubscher, R., Somdyala, N., … Dorrington, R. (2016). Emerging trends in non-communicable disease mortality in South Africa, 1997-2010. *South African Medical Journal*, *106*(5). Retrieved from https://www.ajol.info/index.php/samj/article/view/138196

Nojilana, B., Bradshaw, D., Pillay-van Wyk, V., Msemburi, W., Somdyala, N., Joubert, J., … Dorrington, R. (2016). Persistent burden from non-communicable diseases in South Africa needs strong action. *South African Medical Journal*, *106*(5), 436–437. Retrieved from https://www.ajol.info/index.php/samj/article/viewFile/138173/127742

O'Donoghue, T., & Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, *89*(1), 103–124. https://doi.org/10.1257/aer.89.1.103

Paoli, A., Rubini, A., Volek, J. S., & Grimaldi, K. A. (2013). Beyond weight loss: a review of the therapeutic uses of very-low-carbohydrate (ketogenic) diets. *European Journal of Clinical Nutrition*, *67*(8), 789. Retrieved from https://www.nature.com/articles/ejcn2013116

Park, C., & Le, Q. A. (2018). The Effectiveness of Continuous Glucose Monitoring in Patients with Type 2 Diabetes: A Systematic Review of Literature and Meta-analysis. *Diabetes Technology & Therapeutics*, *20*(9), 613–621. https://doi.org/10.1089/dia.2018.0177

Patel, D., Lambert, E. V., da Silva, R., Greyling, M., Kolbe-Alexander, T., Noach, A., … Gaziano, T. (2011). Participation in Fitness-Related Activities of an Incentive-Based Health Promotion Program and Hospital Costs: A Retrospective Longitudinal Study. *American Journal of Health Promotion*, *25*(5), 341–348. https://doi.org/10.4278/ajhp.100603-QUAN-172

Salkind, B. (2019). 2019_09_27-Wellness-Vision-with-Betsy-Salkind.pdf. Presentation, Kresser Institute.

Simopoulos, A. (2016). An increase in the omega-6/omega-3 fatty acid ratio increases the risk for obesity. *Nutrients*, *8*(3), 128. Retrieved from https://www.mdpi.com/2072-6643/8/3/128

Stanhope, K. L. (2016). Sugar consumption, metabolic disease and obesity: The state of the controversy. *Critical Reviews in Clinical Laboratory Sciences*, *53*(1), 52–67. https://doi.org/10.3109/10408363.2015.1084990

Sturm, R., An, R., Segal, D., & Patel, D. (2013). A cash-back rebate program for healthy food purchases in South Africa: results from scanner data. *American Journal of Preventative Medicine*, *44*(6), 567–572. Retrieved from https://www.sciencedirect.com/science/article/pii/S0749379713001682

Vale, M. J., Jelinek, M. V., Best, J. D., Dart, A. M., Grigg, L. E., Hare, D. L., ... & McNeil, J. J. (2003). Coaching patients On Achieving Cardiovascular Health (COACH): a multicenter randomized trial in patients with coronary heart disease. Archives of internal medicine, 163(22), 2775-2783.

Thomas, M., Elliott, J., Rao, S., Fahey, K., Paul, S., & Miaskowski, C. (2011). A Randomized, Clinical Trial of Education or Motivational-Interviewing-Based Coaching Compared to Usual Care to Improve Cancer Pain Management. Oncology Nursing Forum, 39(1), 39-49. https://doi.org/10.1188/12.onf.39-49

Volek, J., Fernandez, M., & Feinman, R. (2008). Dietary carbohydrate restriction induces a unique metabolic state positively affecting atherogenic dyslipidemia, fatty acid partitioning, and metabolic syndrome. *Progress in Lipid Research*, *47*(5), 307–318. Retrieved from https://www.sciencedirect.com/science/article/pii/S0163782708000167

Volek, J. S., Phinney, S. D., Forsythe, C. E., Quann, E. E., Wood, R. J., Puglisi, M. J., … Feinman, R. D. (2009). Carbohydrate Restriction has a More Favorable Impact on the Metabolic Syndrome than a Low Fat Diet. *Lipids*, *44*(4), 297–309. https://doi.org/10.1007/s11745-008-3274-2

Volek, J., Sharman, M., Gómez, A., Judelson, D., Rubin, M., Watson, G., … Kraemer, W. (2004). Comparison of energy-restricted very low-carbohydrate and low-fat diets on weight loss and body composition in overweight men and women. *Nutrition & Metabolism*, *1*(1), 13. https://doi.org/10.1186/1743-7075-1-13

Wang, L., Mårtensson, J., Zhao, Y., & Nygårdh, A. (2018). Experiences of a health coaching self-management program in patients with COPD: a qualitative content analysis. International Journal Of Chronic Obstructive Pulmonary Disease, Volume 13, 1527-1536. https://doi.org/10.2147/copd.s161410

Webster, C. C., Murphy, T. E., Larmuth, K. M., Noakes, T. D., & Smith, J. A. (2019). Diet, Diabetes Status, and Personal Experiences of Individuals with Type 2 diabetes Who Self-Selected and Followed a Low Carbohydrate High Fat diet. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 12*, 2567.

Wing, C., Simon, K., & Bello-Gomez, R. A. (2018). Designing difference in difference studies: best practices for public health policy research. Annual review of public health, 39.

Wolever, R., Simmons, L., Sforzo, G., Dill, D., Kaye, M., & Bechard, E. et al. (2013). A Systematic Review of the Literature on Health and Wellness Coaching: Defining a Key Behavioral intervention in Healthcare. Global Advances In Health And Medicine, 130712120243000. https://doi.org/10.7453/gahmj.13.042

Wu, X., Guo, X., & Zhang, Z. (2019). The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. *JMIR mHealth and uHealth, 7*(1), e12297.

Affidativ
according to § 8 paragraph 3 of the doctoral degree regulations of February 17, 2015

"I hereby affirm in lieu of oath that I have prepared the submitted work independently and without using any other than the specified aids. Statements, data and concepts taken directly or indirectly from other sources are identified with reference to the source. In the selection and evaluation of the following material, the persons listed below helped me in the manner described, for a fee / free of charge (underline as appropriate):

Other people, in addition to the co-authors listed in the introduction to the work, were not involved in the content-related preparation of the present work. In particular, I did not make use of the paid help from mediation or advisory services. Nobody has received direct or indirect monetary benefits from me for work that is related to the content of the submitted dissertation.

The thesis has not yet been submitted to another examination authority in the same or a similar form, either in Germany or abroad.

I assure you that to the best of my knowledge, I have told the pure truth and have not concealed anything.

I confirm that the submitted electronic version corresponds in full to the submitted printed version.

I am aware of the criminal liability of a false affidavit, namely the threat of punishment according to § 156 StGB with an imprisonment of up to three years or with a fine in the case of intentional commission of the act or according to § 161 Paragraph 1 StGB up to one year imprisonment or fine in the case of negligent commission.

Cape Town, 7 March 2022 Signature

SOFIA MONTEIRO
Postal address: 22 Shiraz Crescent, Constantia, Cape Town, South Africa • +27 (0)72 253 6420 •
sofia.g.monteiro@gmail.com

EDUCATION

PhD Candidate in Economics — October 2017-Present
University of Cologne, WiSo Faculty, Cologne, Germany
Courses: Adv. Experimental Economics, Adv. Experimental Methods, Behavioural Ethics, Health Economics, Social Identity and Narrative, Field Experiments in Education (Norwegian School of Economics), German (Max Planck Institute Bonn), Health Advisor Training (Nutrition Network)

Master of Social Science in Applied Economics — January 2015-June 2017
University of Cape Town, School of Economics, Cape Town, South Africa
Distinctions: Microeconomics-Class Medal, Adv. Econometrics, Micro-econometrics, Labour Economics

IAPO graduate academic exchange — September – December 2016
University of Massachusetts, Amherst, United States
Distinctions for Applied Econometrics, Political Economy & German

J-PAL Africa workshop series on randomised evaluations — August 2015-November 2015
University of Cape Town

Italian Language, Renaissance Art History, Life Drawing — January 2014-November 2014
The British Institute in Florence, Italy | Charles H. Cecil Studios, Florence, Italy

Bachelor of Social Science (Honours) in Economics, First Class — January 2013-December 2013
University of Cape Town, School of Economics, Cape Town, South Africa
GPA: 75% | Distinctions for Advanced Analysis of Household Survey Data, Experiments in Economics, Labour Economics (Class medal), Microeconomics, Economics of Conflict, War and Peace

Bachelor of Social Sciences (Economics, Psychology), With Distinction — February 2010-December 2012
University of Cape Town, Cape Town, South Africa
GPA 75% for both majors & graduated with distinction in psychology

GRADUATE RESEARCH

MASTERS DISSERTATION

'The Prince and the Pauper': The effect of inherited-wealth status on productivity in the lab.

BACHELORS (HONOURS) DISSERTATION

The 'Decent Work' Debate: Rival Narratives of the Wage-Employment Relationship in South Africa.

AWARDS

University of Cape Town Merit Scholarship for Masters study	2015-2016
Economic Research Southern Africa (ERSA) Masters scholarship	2015-2016
NOPOOR funding for experimental research	2015
Louis Wiener Scholarship for excellent coursework results	2015
Class Medal for Microeconomics II (Masters)	2015
NRF Merit Scholarship for Honours study	2013
Class Medal for Labour Economics (Honours)	2013
Dean’s Merit List	2011-2012
Faculty Scholarship	2011-2012
Golden Key International Honours Society	2011
University entrance scholarship on the basis of exceptional Matric results	2009

WORK EXPERIENCE

Max Planck Institute for Research on Collective Goods — October 2017-November 2021
Research fellow in Prof. Matthias Sutter’s Experimental Economics Group

South African Labour Development Research Unit (SALDRU) — March-July 2017
Research Assistant to Prof. Patrizio Piraino
Output: *Nopoor* policy briefs & review of international estimates of intergenerational earnings mobility.

School of Economics, University of Cape Town — February 2013-June 2016
Teaching Assistant for graduate Microeconomics (2016) |
Tutor undergraduate Microeconomics (2012, 2015)
& Natural Resource Economics & Macroeconomics I
(2013) Research Assistant to Dr Justine Burns
Output: baseline interviews for experimental evaluation of *Activate!* leadership program

Road Traffic Management Corporation & 2Roads, South Africa — December 2015-February 2016
Research Assistant to consultants Anton Groenewald & Solange Rosa
Output: National Road Safety Strategy 2016-2030 for South Africa
Research: Economic impact of road injuries and deaths & behavioural interventions

Centre for the Study of African Economies, University of Oxford — February-December 2014
Remote research Assistant for Dr Elwyn Davies and Dr Andrew Kerr