DISCUSSION PAPER SERIES

IZA DP No. 8858

Biased Beliefs and Imperfect Information

Eugenio Proto
Daniel Sgroi

February 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Biased Beliefs and Imperfect Information

Eugenio Proto
University of Warwick
and IZA

Daniel Sgroi
University of Warwick

Discussion Paper No. 8858
February 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 8858
February 2015

ABSTRACT

Biased Beliefs and Imperfect Information*

We perform an *incentivized* experiment designed to assess the accuracy of beliefs about characteristics and decisions. Subjects are asked to declare some specific choices and characteristics with *different levels of observability* from an external point of view, and typically formed through *real world experiences*. From the less observable mobile phone purchasing decisions, hypothetical restaurant choices, political views, happiness to the fully observable height and weight; after they are asked to report beliefs on statistics over the same items concerning other individuals living in the same environment. We test two main hypotheses: *(i)* whether for items not perfectly observable, individuals suffer of some type of bias in these beliefs; *(ii)* whether this bias would disappear for weight and height, when the information is perfectly available. We find a powerful and ubiquitous bias in perceptions that is "self-centered" in the sense that those at extremes tend to perceive themselves as closer to the middle of the distribution than is the case. Albeit weaker, this bias does not disappear when the information is readily available as in height and weight. We present evidence from our experiment that *limited attention* and *self-serving deception* can provide explanations for this bias and present important economic applications.

JEL Classification: D03, C83, D84

Keywords: biased beliefs, information, attitudes, characteristics, self-centered bias

Corresponding author:

Eugenio Proto
University of Warwick
Department of Economics
Coventry CV4 7AL
United Kingdom
E-mail: e.proto@warwick.ac.uk

* The authors would like to thank Robert Akerlof, Gordon Brown, Dirk Engelmann, David Gill, Peter Hammond, Graham Loomes, Heracles Polemarchakis, Andrew Oswald, Francesco Squintani and John Stovall for useful comments. Funding was provided by the Leverhulme Trust. We are grateful to Alex Dobson, Stephen Lovelady, Andrew Taylor and Neil Gatti for expert research assistance.

IZA Discussion Paper No. 8858
February 2015

NON-TECHNICAL SUMMARY

How do you compare to others? Are you taller or heavier than average? Are your political views in the mainstream? Are you as happy as your peers? Are your decisions the same ones others make in similar situations?

The beliefs people hold about their positions in relation to others are important in everyday life. People assume that their beliefs are correct, but recent experiments we conducted suggest otherwise. Our research has uncovered a subtle form of self-deception: who we are determines what we regard as the truth. This insight underscores how subjectivity affects objectivity.

We asked University of Warwick students a series of questions about themselves and then a series of questions about themselves in relation to other students. First, we asked students about observable characteristics (their height and weight), attitudes (rating their political views and their level of happiness) and behaviour (their choice of mobile phone brand, and a hypothetical question about which of two recommended restaurants they would choose). Then, we asked students to give estimates about averages among all university students: the average heights and weights; average political beliefs and happiness rankings; the preferred brand of mobile phone; the most common restaurant choice in the hypothetical question. The participants in the experiment were incentivised to be as accurate as possible; by giving correct answers, they potentially could earn higher payments in the experiment.

In every case, we found biases that were remarkable, systematic and pronounced. “Self-centred” perceptions are ubiquitous, in the sense that an individual’s beliefs about the rest of the population depend on his or her own position in that distribution. Those at the extremes tend to perceive themselves as closer to the middle of the distribution than is the case. Accordingly, taller and heavier individuals think that there are more tall and heavy people in the population. Individuals in the political fringes perceive themselves as more representative, as do those who are very happy or sad. Students believed that their mobile phone brands were the most popular, and the tie-breaking choice they made between two virtually indistinguishable restaurants also was the most popular choice.

"Let them eat cake." (Commonly attributed to) Marie Antoinette (1755 – 1793), Archduchess of Austria and Queen of France.[1]

1 Introduction

Early work attempting to address the question of whether individuals hold biased beliefs linked priors with information imperfection.[2] Morris (1995) provides a survey of the considerable literature which followed. He argues that in order to avoid using subjective heterogeneous priors to justify any result *ex post* given the sensitivity of many economic models to assumptions over beliefs, it is important to identify systematic regularities concerning how beliefs are biased to provide more discipline for the use of heterogeneity in prior beliefs.[3] This provides a strong motivation for our study: to provide empirical support for the existence of biased beliefs, catalog systematic regularities and better understand the underlying causes.

To that end we design an experiment in order to asses the role of the information availability in shaping biases on beliefs. We perform an incentivized experiment and consider beliefs on attitudes (specifically relating to happiness and political stance), choices (relating to mobile phone purchasing decisions, and hypothetical restaurant choices) and observable characteristics (height and weight), characterised by a different degree of observability, from very incomplete to perfect. Characteristics like height and weight *among students at the same university* are in fact readily observable in day-to-day life, certainly

[1] While this remark was most certainly not actually made by Marie Antoinette, it has been used to epitomize the apparent inability of the ruling classes in France just prior to the French Revolution to appreciate the difficulties of those significantly poorer than themselves, possibly because of a lack of concern, empathy, of awareness or a completely self-centred set of perceptions."

[2] For example, Geanakoplos (1989) is a classic reference. It is interesting to note that biased beliefs may not be inconsistent with a rational process of Bayesian updating, but merely reflective of the imperfect information available to the observers.

[3] As Morris puts it "We should resort to unmodelled heterogeneities of prior beliefs only when we can imagine an origin for the differences in beliefs and we can perform comparative static exercises, comparing the predictions of heterogeneous prior models with alternative explanations."

among the subject-pool in our experiment. Another special feature of our experiment is that (unlike the large part of the economic literature on biased beliefs) we explicitly avoid analysing performance-related beliefs, where overconfidence usually plays a strong a role in distorting the process of belief updating (e.g. Grossman and Owens, 2012, and Burks et al., 2013) representing a potential source of confounding.[4]

In all but the hypothetical restaurant choice problem, the beliefs are formed through experiences in the real world, outside of the laboratory. As a result we have a measure of external validity for our findings which we fortified by designing the questions in our experiment to be more natural than those which make sense only in the context of an entirely laboratory-based experiment. Our questions are generally of two types. The first type of questions essentially ask what percentage of peers a subject thinks lie below him or her in a cumulative distribution. By asking this in an extremely simple way (for example "What percentage of students at Warwick of your gender do you think are shorter than you?") we avoid the need to explain cumulative distributions or require any knowledge of probability. In the second type of questions we ask our subjects to estimate the averages for the characteristic, attitude or behavior from among their peers (for example"What do you think is the average weight for a Warwick student of your gender?") and again the focus is on keeping the language natural and simple.

We hypothesise that rational individuals with unlimited attention are unlikely to display any bias for observable characteristics like weight and height of other people who they are likely to see on a day-to-day basis in the same environment. However, for elements that are not readily observable like, Happiness, Political Stance, Mobile Phone owned, the best they can do is to use the observation of the realization of the parameter

[4]Burks et al. (2013) present a useful discussion of this literature. Of related interest, this literature has shown that biased beliefs may not be inconsistent with a rational process of Bayesian updating, but merely reflective of the imperfect information available to the observers: Benot and Dubra (2011) show that, if individuals have imperfect knowledge of their own ability, even individuals performing correct Bayesian updating starting from a common prior may report what seems to be an overconfident belief.

in a sample that they directly observe. Since they are more likely to have acquaintances that are similar to them, they are more likely to observe values of the parameters that are closer to their own.[5]

Accordingly, in this paper we test two main hypotheses: *(i)* for characteristics and choices on items not perfectly observable, individuals should be biased in the sense of being "self-centered". Hence those at the extremes should perceive themselves as closer to the middle of the distribution than is the case, while individuals with values closer to the mean tend hold more accurate perceptions about their position; *(ii)* this bias should disappear for weight and height, when information is perfectly available.

We find support for *(i)*, but only partial support for *(ii)*. Accordingly, individuals in the political fringes perceive themselves to be more representative, as do those who are very happy or sad. Individuals with less popular mobile phone brands think that there are more individuals using the same brand, but the ones with popular mobile brands correctly estimate its distribution. Finally, when asked to decide on a tie-breaking rule, in the hypothetical restaurant choice problem those who choose in a given way tend to see their choice as being more popular than is the case.[6] Put simply, individuals tend to see themselves as more "average" than is the case. However, albeit to a smaller extent, the self-centered bias still holds for weight and height: taller and heavier individuals think that there are more tall and heavy individuals in the population, similarly shorter and lighter individuals believe to be less extreme than in reality.

Why might the self-centered bias still hold for the observable characteristics? Individuals are not endowed with an unlimited amount of attention, a point recently emphasized elsewhere (e.g. DellaVigna 2009). Hence they may put *undue* weight on easily available

[5]This echoes the literature on assortative matching (Becker, 1973) or homophily (McPherson, Smith-Lovin and Cook, 2001; Golub and Jackson, 2011) through which people may associate with those who are similar to themselves.

[6]For example, an individual who opts for a more deferential tie-breaking rule sees others as similarly deferential.

data (the so-called *availability heuristic*, Tversky and Kahneman, 1973), their initial estimate is their own individual value and then they update by using easily attainable data from others to adjust.[7] Using data deriving from the same experiment we will argue that this is plausible, but it is unlikely to represent the only explanation. If subjects make inferences by using information from their own similar peers, when they are asked to estimate averages of the characteristics, these estimates should be increasing in the subject's own characteristics. Furthermore, the average individuals, who should draw observations from other average individuals, should not display biased perception. Data from our experiment seem to show this pattern only in part. For example, puzzingly, male subjects tend to perceive average height to be systematically lower than reality and subjects in general tend to perceive a lower level of general happiness.

We argue that a possible explanation of this pattern is that subjects' beliefs are self-centred not only because of limited attention, rather some biases are "self-serving" because individuals in some instances may enjoy the feeling of being different with respect to others than is the case. For example male subjects might want to think that they are taller with respect to the average, or individuals might want to think that they are happier than others. Accordingly, as Carrillo and Mariotti (2000) and Benabou and Tirole (2002) argue, self-serving beliefs may be due to some form of "strategic ignorance" (or "self-serving bias") and might be "rational" in the sense that it is utility-maximizing to self-delude.

Finally, our data seem to rule out that the effect is due to salience of the individuals with extreme characteristics (such as being very tall, or politically at the far right or left). We asked subjects to estimate the averages for subjects belonging to the top and bottom 10% of the different distributions. If there is an effect due to salience, individuals should

[7]This process is called anchoring in cognitive psychology (Tversky and Kahneman, 1974) and the failure to realize that inferences are biased by a non-representative sample can be categorized as irrational and there is a long literature which follows this line of thinking.

systematically overestimate the averages of the top and underestimate the averages at the bottom. We cannot observe this pattern from our data.

To relate our findings back to other existing work, there is a growing economic literature on individuals' biased beliefs, and we can distinguish three main form of bias: individuals tend to be *overconfident* about their own abilities (e.g. Moore et al., 2008; Burks et al., 2013); the so-called *law of small numbers*, when individuals expect random draws to be excessively representative of the distribution from which they are drawn (e.g. Clotfelter and Cook, 1993; Rabin, 2002); *projection bias*, where individuals expect future utility is too close to todays (e.g. Loewenstein, ODonoghue, and Rabin, 2003). We want to analyse the effect of information over biases and avoid possible confounding. Hence we focussed on a different form of bias, concerning other people's characteristics and choices that, unlike ability, skills or future preferences can be assumed as completely known to the subjects featuring the characteristic or making the choice and easily measured in a laboratory environment.

The psychology literature has already emphasised that Individuals typically suffer of the "false consensus bias" (Ross, Greene and House, 1977): subjects are generally asked to indicate their attitude or behavior on a dichotomous measure (yes or no, agree or disagree) often to a matter of opinion. They are then asked to estimate the percentage of their peers who would respond one way or the other and their estimate of consensus for their own position exceeds the estimate for it made by those who endorsed the opposite position.[8] Engleman and Strobel (2012) show that in a laboratory experiment, false consensus arises only when information is not readily available, but requires some processing by individuals in order to be of use. However, beside our interest in assessing the effect of the incomplete information, there is a second important point of departure between our paper and the false consensus literature. In this experimental literature

[8]See page 612 in the Handbook of Experimental Economics (Kagel and Roth, 1995) for more on false consensus

individuals are generally asked to indicate their attitude or behavior on a dichotomous measure (yes or no, agree or disagree) often to a matter of opinion. On the contrary, in the current paper we are concerned with matters of fact, and hence we investigate how beliefs differ with respect to the true state of the world and not with respect to the opinion of others. The difference is fundamental for this paper since we are concerned with the accuracy of beliefs with respect to the truth (or departures from rational expectations).[9]

Another concept usually associated with false consensus is "assumed similarity" (Cronbach, 1955). This is generally measured as the absolute difference between the position attributed to oneself and that ascribed to a benchmark individual known to everyone. An assumed similarity score is generally computed as the absolute difference between the position attributed to oneself and that ascribed to the target, and generally each individual tends to position target closer to themselves. Again, these procedures do not provide information on the accuracy of perception, the aim of the current paper, since they are instead interested in comparing individuals' relative perceptions.

The self-centered bias identified in this paper has important implications for economics, for example informational herding theory and auction theory, and for economic modelling more generally. There are also direct implications for policies directed towards tackling problems that are at least partly based on beliefs about where people lie in distributions. A good example would be policies designed to tackle obesity in a world in which those who are very overweight do not perceive themselves as overweight stress-

[9]In one seminal paper on false consensus, Marks and Miller (1987) note: "...The false consensus hypothesis has no direct bearing on whether subjects will overestimate, underestimate, or accurately estimate the actual consensus for their own behavior." For example, consider the view of a right-wing individual who is asked what percentage of people are right-wing. She might say 60%. A left-wing person who is asked the same questions might say 50%. The"false consensus bias" is concerned with the difference between 60% and 50%, while our concern is the difference between 60% and the true percentage of right-wingers and how this difference between beliefs and the truth changes as the type of individual changes. Notice that we can derive one observation from false consensus which closely links to our work: if individuals do differ in their opinions then it is immediate that at least one person must have an opinion that is at odds with the truth. However, this does not help us to see which individual is more likely have such a view.

ing the harm of being overweight should perhaps take second place to educating people about their own position in the distribution. We discuss many application of both the general and specific type and corresponding policy ramifications in section 5.

In the next section we summarize the experimental design and the key variables. Section 3 examines the experimental results both in terms of the perceived cumulative distributions. Section 4 discusses possible explanations also by analysing the errors made by different individuals when asked to estimate averages. Section 5 presents some applications of the self-centred beliefs. Session 6 concludes. Tables and figures are provided in the Appendix, and the experimental instructions are reproduced in the Supplementary Information.

2 Experimental Design and Key Variables

A key feature of our design is the merger of field elements in a laboratory-based design by conducting a laboratory experiment which draws on our subjects real-world experiences. We want to make use of real-world choices and characteristics as far as possible to lend external validity to our findings. We also want to be able to exert as much control over our subjects as possible and to incentivize them for accuracy in their estimations which makes a laboratory ideal.

Our data was collected using a series of computerized tasks and questions presented in a controlled experiment at the University of Warwick.[10] The text of the questionnaires and the accompanying instructions are provided in the Supplementary Information. The experiment was conducted in a laboratory, however many of the choices and characteris-

[10]The experiment took place in 19 sessions with about 8 students per session, and was conducted on 27 May, 30 May and 29 June 2011. There was also an earlier non-incentivized pilot experiment which consisted of 120 participants drawn from the same experimental pool, held on 17 March, 5 May and 11 May, 2010. The main results for this paper will be drawn from the fully-incentivized experiment, though the data from the pilot study will be used when calculating the average height, weight, happiness and political stance of the Warwick student body.

tics are drawn from the real-world experiences of our participants as well as information given to them in the laboratory. The participants were 154 students drawn from the university-wide experimental pool of over 1500 subjects.[11]

Subjects were given a £2.50 show-up fee, plus a bonus of £5 pounds if a randomly drawn answer was within 10% of the correct answer in rounds 2 to 6 as described below. For example, if participants were asked to state the average height of the student body in Warwick and this was the randomly allocated bonus question they received a £5 bonus if and only if their answer was within 10% of the true average.[12] The payment scheme was fully transparent to all participants and highlighted in the instructions during the experiment. No participant was allowed to participate more than once.[13] The experiment itself typically lasted 20 minutes and the average payment was a little over £5, producing an hourly rate of around 25 US dollars. The experimental time-line is summarized below but a full transcript of the instructions and tasks faced by the participants is provided in the Supplementary Information.

2.1 Experimental Time-line

Participants arrive at the laboratory, are registered and taken to a screened computer terminal where they receive on-screen instructions as detailed in the Supplementary Information. The payment scheme is explained and they are informed that everything

[11]Ensuring a varied pool of experimental participants selected on as randomized a basis as possible was especially important for this experiment. The University of Warwick keeps a register of those available for use as experimental participants and a research assistant (rather than the experimenters) drew from this large pool of potential applicants on a random basis. Participants were recruited without any knowledge of the nature of the experiment and so could not self-select into or out of the experiment. The times and dates of the sessions were varied to avoid discriminating against participants from any demographic, and in the event we had large variety in terms of subject, year-group and gender.

[12]For our purposes the true average was based on the numbers generated within this experiment and and from the earlier non-incentivized pilot experiment. For one question, denoted (5e) below the scheme was changed slightly as participants had to select from an interval and so they were told that an answer in the correct interval or the one to either side would be sufficient to win the prize.

[13]Participation in the pilot experiment also ruled out participation in the full experiment.

they do is fully anonymous. To that end they receive randomly generated usernames and passwords to use as logins for the terminals. After entering usernames and passwords the experiment proper begins in round 1 by asking participants to report their gender (1a), height (1b), weight (1c), happiness (1d), political beliefs (1e), and current brand of mobile phone (1f). They are also given a hypothetical restaurant choice as follows: "Imagine that you have to decide between two restaurants in which to have dinner alone. They are called restaurant A and B. You have some private information that A is better, but you know that an equally well-informed colleague has information suggesting that B is better. Would you choose to eat at A, B or are you indifferent?" (1g).

For (1b) and (1c) they were allowed to enter their heights and weights in metric or imperial measurements as they wished and this freedom was maintained in all rounds. For (1d) participants are instructed to use a 7-point Likert scale as follows: "Please use a 7-point scale where 1 is completely sad, 2 is very sad, 3 is fairly sad, 4 is neither happy nor sad, 5 is fairly happy, 6 is very happy and 7 is completely happy." For (1e) they are instructed to use a similar scale: "Please use a 7-point scale where 1 is far left, 2 is left, 3 is centre left, 4 is centre, 5 is centre right, 6 is right and 7 is far right." (1g) might be of special interest to those interested in rational herding and informational cascade literature (see Banerjee, 1992, and Bikhchandani, Hirshleifer and Welch, 1992) and essentially asks what participants would do in a situation of theoretical indifference when processing information. There is no clear right or wrong answer to (1g) though "A" points to a measure of confidence in the participant's private signal over that of their colleague, whereas "B" perhaps implies a measure of deference towards others (or a lack of confidence in the individual's own signal). Once they answered these questions and hit a "submit answers" button they were taken to the round 2 questions.

At this point they were informed that one question would be chosen at random in round 2 or the later rounds (rounds 3, 4, 5 and 6) as a "prize question" for which they

would receive a bonus payment of £5 if their answer was within 10% of the correct answer. They were then asked to report the percentage of students at Warwick they thought were less happy than they were (2a), less right-wing (2b), shorter (2c) and lighter (2d). For (2c) and (2d) they were asked to consider only their own gender. In (2e) they were asked to consider the mobile phone brand listed in round 1 and asked what percentage of students at Warwick they thought also used the same brand of mobile phone as their main mobile phone. For (2f) they were asked to consider the hypothetical restaurant choice and report what percentage of their fellow Warwick students they thought chose the same answer that they did (they were reminded of the entirety of the question and the possible answers).

In round 3 they were asked to report the average height for someone in the 10% tallest Warwick students of their gender (3a), the average weight for someone who is in the 10% heaviest Warwick students of their gender (3b), the average happiness for someone who is in the 10% happiest students at Warwick (3c), and the average political belief for someone who is in the 10% most right-wing students at Warwick (3d). For (3c) and (3d) they were asked to use a 7-point Likert scale as before.

Round 4 was phrased identically to round 3 except that in each case in the four questions they were asked to report the average for the 10% shortest (4a), 10% lightest (4b), 10% most sad (4c) and 10% most left-wing (4d), again for the population of students at Warwick, using a 7-point Likert scale for (4c) and (4d), and considering only their own gender for (4a) and (4b).

Round 5 focused on overall averages rather than extremes in the distribution. They were asked to report the average height (5a) and weight (5b) for a Warwick student of their gender, and using a 7-point Likert scale the average happiness (5c) and political belief (5d) for a Warwick student. In question (5e) they were asked to estimate the percentage of their fellow Warwick students who used each of a selection of mobile phone

brands. They were presented with a tabulated list of the most popular brands in the UK, and they were informed that the list was presented in alphabetical order (except for the "other" category which was presented last). They were asked to include an entry for every brand (including "other"). For (5f) they were asked again about the hypothetical restaurant choice: "Think again about the restaurant question you were asked earlier in the session. To remind you, you had to decide between two restaurants in which to have dinner alone. They were called restaurant A and B. You had some private information that A is better, but you knew that an equally well-informed colleague had information suggesting that B was better. What percentage of your fellow Warwick students do you think would have chosen to eat at restaurant A if they were asked the same question? Remember that the other options were indifferent and B."

For round 6 the participants were asked to answer a single question designed as a check on their ability to understand and manipulate expectations and probability: "Consider the following gamble. You have a 20% chance of winning £100, a 40% chance of winning £10 and a 40% chance of winning £0. If you played this gamble many times what would you expect to be your average winnings per gamble? (in pounds)"

Round 7 is a final questionnaire and, as is conventional, was not incentivized (participants were informed that the incentivized part of the experiment had ended) since there was no way of checking right or wrong answers. They were asked to report their age (7a), nationality (7b), degree subject (7c), whether they studied mathematics up to their final year at school (7d) and also comment on their methods, if any, during the incentivized parts of the experiment (7e).

Once each round was completed participants could not go back and change earlier answers, nor did they know the content of later rounds upon entering answers to earlier rounds. This was important as it prevented any attempt to retroactively alter their answers to make winning the bonus payment easier.

3 Results

Before moving on to the analysis we present the main variables in table 1 below. The variable "Happiness" is coded from completely sad (1) to completely happy (7) and taken from the answers to question (1d), and "Political Stance", from extreme left (1) to extreme right (7), taken from question (1e). "Weight" is converted to kg from the answers in question (1c) and "Height" to cm from the answers in question (1b).

In this section we first present an analysis of individuals' beliefs about their own positions in the true distribution, starting first with a model. Thereafter we analyze estimated averages. In both cases our findings indicate that beliefs about position in the distribution and about averages are a function of individual's own position in the distribution.

3.1 Modeling Beliefs

Let Θ be the set of characteristics, attitudes and choices considered in this paper, for example height, reported happiness or the market share of a brand of mobile phone. $\theta \in \Theta$ is a particular characteristic, attitude or choice, for example, height. We index the value of each characteristic for each individual i, so θ_i might be individual i's height. Let $F(\theta)$ be the cumulative distribution of characteristic θ. In a mild abuse of notation, denote $F(\theta_i)$ as individual i's true position in the cumulative distribution and then $E_i(F(\theta_i))$ is i's belief about his position in the cumulative distribution.[14] Given the nature of incentives in the experimental design a typical subject i solves the following programme:

[14]We preferred to ask "less that", rather than "at least as" because we judged it a more natural question that individuals are likely to have faced in their everyday life. We calculated the CDFs using the same definition.

$$Min_{E_i(F(\theta_i))} \{E_i (F (\theta_i)) - F (\theta_i)\} \tag{1}$$

We assume that individuals form their expectations by drawing from a sample M_i from the total population N. However we cannot rule out the possibility that the sample of size M_i may be biased. In particular is may be that the sample is taken from members of the population who are similar to the individual i. An unbiased estimator would be

$$E_i (F (\theta_i)) = F (\theta_i) + error_i \tag{2}$$

where $error_i$ is i.i.d., with mean 0 across the population. In order to asses the existence of a systematic bias, we will therefore consider the model:

$$E_i (F (\theta_i)) = G(F (\theta_i)) + \sigma(\theta_i)\epsilon_i, \tag{3}$$

Where $\sigma(\theta_i)$ is a general function to allow for heteroscedasticity with respect to θ_i, and ϵ_i is a white noise error.

From the data, we can observe the beliefs $E_i (F (\theta_i))$, the real distribution $F (\theta_i))$ hence we estimate $G(F (\theta_i))$ both non parametrically using local polynomial smoothing[15] and by making parametric assumptions on this functional form.

This formulation allows us to examine directly the difference between $G(F (\theta_i))$ and the true distribution, $F (\theta_i)$ which represents the bias made by each individual, allowing for non-linearities in beliefs as we change the own values of each individual. One way to add structure to this bias might be to define $G(F (\theta_n))$ as $G_{M_i}(F (\theta_n))$ with the G

[15]Kernel-Weighted Local polynomial smoothing involves fitting the response to a polynomial form of the regressor via locally weighted least squares. In the Kernel-weighted regression, $G(F (\theta_n))$ is calculated without assuming a functional form, as a constant term of a regression weighted by the kernel function of $E_i (F (\theta_i))$ on the polynomial terms $G(F (\theta_n)) - G(F (\theta_i))$, $(G(F (\theta_n)) - G(F (\theta_i)))^2, ..., (G(F (\theta_n)) - G(F (\theta_i)))^p$, for each point $G(F (\theta_n))$. The definitive reference is Fan and Gijbels (1996), see also the Stata 12 base reference manual, p. 1001.

function acting as a way to choose a subsample $M_i \subseteq N$ where the statistic $F(\theta_n)$ is defined. In particular we might then consider $M_i = N$, when information is perfect and individuals have unlimited attention, hence $G_N(F(\theta_i)) = E(F(\theta_i))$, with no implied bias. Alternatively $M_i = N_i$, with $N_i =$ individuals belonging to her closer group when information is not perfect and outsiders are not observable and/or when individuals have limited attention and can only remember other individuals who can observe more often (thus if there is homophily, individuals in N_i are more similar to i). Or we might consider $M_i = M_i^*$, with M_i^* as a sample selectively chosen by the individual, when memory is self-serving in the sense of Benabou-Tirole (2002).[16] We purposefully leave the potential nature of the bias open to allow us to look for (and hopefully differentiate between) several different sources, which we will return to in section 4. In the following subsection we present the empirical evidence with respect to the characteristics weight, height, happiness and political stance. We will also analyze the beliefs about mobile distribution and deference in information processing. These last are slightly different because $F(\theta_i)$ represent frequencies and not CDFs.

3.2 Empirical Evidence on Perceived Distributions

Given the nature of the data, we present the results in three separate groups. The CDFs of the observable characteristics, height and weight, are in figure 1; the CDFs of the unobservable characteristics, happiness and political stance, are in figure 2.[17] Finally, the frequencies of answers to the restaurant choice question (1g) and the mobile choice question (1f) can be read from the histogram in figure 3.

[16] Along similar lines, see Compte and Postlewaite (2004) who motivate their paper with the story of a lawyer who observe a full history of his successes and failures but who may dismiss his losses as stemming from biases in the judicial system and instead act as though he has only ever won each case. They then go on to show how equilibria in which this sort of reasoning takes place can exist in a model of rational overconfidence.

[17] When calculating the population averages we also used the 120 data from the pilot experiment to increase the size of the sample to 274.

We start by analyzing weight, height, happiness and political stance. From the right panels of figures 1 and 2. We can observe the real CDF, $F(\theta_i)$ and the scatter-plot of the perceived positions, $G(F(\theta_i))$ with their locally weighted scatter-plot smoothing or "Lowess".[18] The scatters in the right panels of figure 1 represent the real and the perceived CDFs in θ_i for each subject i, θ_i for height and weight, and in figure 2 the same observations for political stance and happiness. The dashed line represents the estimated function $G(F(\theta_i))$ with the 95% confidence interval; In order to compare $G(F(\theta_i))$ and $F(\theta_i)$, we also plot the 45 percent degree line. There is a clear pattern: individuals at the extreme tend to overestimate the number of individuals who are equal to or more extreme than themselves, while those at the center of the distribution seem better informed. This is true for happiness and political stance, female and male weight, and male height. For political stance and happiness we note that the two extreme numbers (1 and 7) seem to revert to the 45 degree line. This makes sense since these two characteristics are bounded above and below, therefore extreme individuals will be able to calculate more accurately their position (and have lower scope for self-deception). This effect appears even more clearly in the parametric analysis to follow.

In the 4 columns of table 2, where we estimated $G(F(\theta_i))$ parametrically for happiness, political stance, height and weight, we find comparable results. In this parametric analysis, we assume a linear model, so that a positive and significant constant term would signal an overestimation of sadder, more left-wing, shorter and lighter individuals and a coefficient less than 1 would signal a overestimation of happier, more right-wing, taller and heavier. A coefficient closer to one would signal a more accurate estimation in average.

[18]This displays for each value of the independent variable, θ_i a smoothed value of the dependent variable, $G(F(\theta_i))$. The smoothed values are obtained by running a linear regression using only the data $(x_i; y_i)$ and a subset of the data with the x values close to x_i. Data are weighted so that the central point $(x_i; y_i)$ receives the highest weight and points that are farther away receive smaller weight. The estimated regression line is then used to predict the smoothed value for y_i only; a separate weighted regression is performed for every point in the data.

We find all coefficients less than 1 ($p - value < 0.01$), the closest to unity is the coefficient related to height, then in decreasing order for weight, political stance and happiness. This seems consistent with a pattern of decreasing observability among those characteristics. We also find positive and significant intercepts, with the only exception being the coefficient on female weight, which is not significantly different from 0. This indicates that females towards the lowest extremes are the most accurate in their perceptions about this characteristic.

Mobile phone ownership and attitudes in the hypothetical restaurant choice are not ordered variables, it is therefore impossible to determine a perceived CDF and a real CDF. We therefore collected data on beliefs about frequencies and compared these with the true frequencies. Accordingly, we define our $F(\theta_i)$, as the true frequency of the choice θ_i and $G(F(\theta_i))$, the beliefs' function. In the left panels figure 3 we can compare the histogram representing the distribution on mobile phones and attitudes toward in the hypothetical restaurant choice question, $F(\theta_i)$, with the subjective beliefs $G(F(\theta_i))$, where the different θ_i have been ordered increasing with their frequencies. Although we are comparing now frequencies rather than CDFs, we note a similar pattern to the one emphasized in figures 1 and 2. The Lowess tends to stay above the real frequencies for less popular brands and for less frequent tendencies in the restaurant choice question.

As before we estimate the model 3, and present the estimated $G(F(\theta_i))$, in the right panels of figure 3. The reading of the panels is slightly different from before, with the average subjective beliefs above the real level for less frequent θ and are non-significantly different for more frequent θ. The interpretation of the results are the same as with happiness, political stance, height and weight: individuals with more common brands or with the most common attitude in the restaurant question have beliefs that are on average correct, on the contrary individuals making less common choices overestimate the frequency of their choices.

In table 2, we estimate parametrically $G(F(\theta_i))$ for the mobile phone distribution question (2e) and attitudes towards information processing from the restaurant question (2f). This confirms the main finding emphasized by figure 3.[19]

In essence subjects more at the extremes of the distributions think there are more individuals in the same position (or indeed in a more extreme position) than themselves, while more average subjects tend to have more correct beliefs. So, a very tall person really does perceive him or herself more "normal". Perhaps more economically significant, someone who has purchased a less popular mobile phone believes it to be more popular than is the case. Note that this bias is true for different frames. Either when subject are asked about their position and when their asked about the frequency of their choices.

One remarkable point that stands out is the ubiquitous nature of this bias: only for females who are among the shortest and only in the parametric analysis is there any deviation from the simple rule that those at the extremes do not see themselves as being as extreme as they truly are.

More formally, we find that individuals with a θ_i closer to the average tend to estimate better their position than those with a θ_i closer to the extremes of the distribution. Denote E_i as the expectation operator for individual i. We can then express this as a simple conjecture: for each $\theta \in \Theta$, $\exists \theta^*$ s.t. $\theta_i < \theta^* \Rightarrow E_i\left(F\left(\theta_i\right)\right) > F\left(\theta_i\right)$ and $\theta_i > \theta^* \Rightarrow E_i\left(F\left(\theta_i\right)\right) < F\left(\theta_i\right)$. Put simply, we can consider a point in the cumulative distribution θ^* such that individual's with value $\theta_i < \theta^*$ believe they are positioned higher in the distribution than is the case, and for $\theta_i > \theta^*$ they believe they are lower.

More compactly, $E_i\left(F\left(\theta_i\right)\right)$ stochastically dominates in the second order sense $F\left(\theta_i\right)$. This last characterization can be observed from the left panels of figures 1 and 2, by

[19]We also note a slight tendency for the owners of the very smallest market share brands to better understand their own-brand market-share as compared with those with slightly higher market share brands. This could be due to the fact that the owners of these least popular brands were looking for niche products with specific features and so spent more time and effort on research or may see themselves as different.

comparing the actual distribution and the Lowess of the beliefs.

This is a form of mean-reversion of beliefs. Since beliefs seem too heavily dependent on θ_i we label them "self-centered".

4 Discussion: What Generates Self-centered Beliefs?

The above analysis suggests that we can rule out the possibility that individuals are fully Bayesian and endowed with unlimited attention, something that we can define as a perfect "statistical model" of learning. When incentivized subjects are asked to estimate their position in the distribution of height and weight it seems likely that they would refer to causal observation gathered in their everyday lives. Consider for instance, height, and how many individuals our typical subject might have seen in the few hours prior to the experimental session alone. In a large campus university with several thousand students, an individual subject is likely to have passed several hundred other students just getting from her accommodation to the laboratory. If an individual, when asked to assess his or her position in the dimension of an easily observable characteristic, considers only his or her group of close acquaintances, it seems much more likely that they might be (rationally or irrationally) anchoring themselves to this group rather than having only seem individuals from this group.

There are many possible routes through which the self-centered beliefs might emerge. Here we discuss then and how well they fit our data. We can broadly distinguish three different theories: (i) limited attention; (ii) self-serving beliefs; and (iii) salience of the extremes.

i) Limited Attention

Consider first an explanation based on limited attention. We might argue that individuals can be perfectly Bayesian in the way they update their beliefs but only use

the observation of the realization of the parameter in a limited sample that they observe more often because they have limited attention capacity. The fact that the extremes in the distribution suffers the worst from the bias lends some support to the *availability heuristic* explanation identified by Tversky and Kahneman (1973), according to which individuals put *undue* weight on easily available data, and may then draw biased inferences.[20] Closely related is the concept of anchoring in cognitive psychology: individuals base their estimate by first looking at their own individual value and then use the evidence from others to adjust (Tversky and Kahneman, 1974). The failure to realize that inferences are biased by a non-representative sample because of anchoring can be categorized as irrational.

We can test the plausibility of this explanation in figure 4, where we plotted the beliefs about the averages perceived by each subject against the individual characteristic of the same subject, and a line representing the real average, calculated by using the subjects in our sample. If when asked about the averages, individuals only sample their close peers and if individuals associate with similar peers, we should observe a positive association and that the regression line should cross the real average line close to the middle, so that average subjects that match with other average subjects make unbiased estimation.

Figure 4 suggests that a limited attention explanation can explain self-centred beliefs to some extent, but it is unlikely to be the only determinant. We can observe that for weight, both male and female subjects in the middle are the most correct, consistently with the unlimited attention story. Furthermore, we can observe a positive relationship between the characteristics and the perception in all but in female height. For height, male subjects in the upper end of the scale seem the most precise, which is consistent with

[20]Linked to this, is the idea that observing others is one of the cheapest ways to acquire information and is the key form of learning in the social learning and herding literatures (see Banerjee, 1992, and Bikhchandani, Hirshleifer and Welch, 1992).

the results in figure 1, where we can observe they are also the ones that more correctly estimate their positions in the distribution. Looking at the bottom right-hand panel of figure 5, consistently with the above observation, we also note that in general male subjects significant underestimate real average height. For female height the relationship is negative and insignificant, which posses an interesting puzzle, and female subjects seem also to slightly overestimate the average. Furthermore, average happy subjects seem to slightly underestimate the level of happiness, which seems to be true also in general, as we can note from the top left-hand panel of figure 5, where average happiness is significantly below the value reported by the subjects. Finally, we note that in political stance subjects believe the true distribution to be slightly more to the right than is the case in our data.[21] In what follows next, we will try to explain these anomalies.

(ii) Self-Serving Beliefs

When forming beliefs, focussing on peers that are similar can be a strategy rather than due to a limited amount of attention. Costly learning can explain this form of deception since we can argue that where falsely believing yourself to be in a particular position is useful, it might cost more (in terms of final utility) to learn otherwise.[22] To explain why our subjects fail to learn the truth, we might even appeal to the optimal experimentation literature, beginning with Robbins (1952) and brought to economics by Rothschild (1974). This literature demonstrates that optimizing decisions while simultaneously attempting to learn can result in learning stopping before the truth is known. Since our subjects need to make decisions on a daily basis using the beliefs they form, this process of decision-making and learning in a simultaneous setting may be appropriate. In relation to the concept of "self-deception", Carrillo and Mariotti (2000) and Benabou and Tirole (2002) provide numerous mechanisms (including self-delusion and memory

[21]Of course this could be due to the fact that subjects participating to the experiment might be slightly more politically left-wing than the typical student at Warwick!

[22]Coate and Loury (1993) and Farmer and Terrell (1996) look at costly learning in the context of discrimination.

manipulation) to foster rational ignorance of the truth (indeed Carrillo and Mariotti call this type of behavior "strategic ignorance").

Therefore, we can argue that individuals, in some cases, may want to report a bias estimation to to feel better. This would also explain why average individuals are not always correct in estimating average happiness and height, as we argued above using figure 4. The reason the average subjects tend to underestimate the level of general happiness and male height and overestimate female height might be a self-serving belief. Short males might want to think that there are more short individuals than in reality, this would also explain why tall males do not typically suffer from this bias (see figure 4, but also figure 1, where we note that the self-centered bias is very small for male subjects in the upper end of the height distribution). At the same time, tall females might want to think that there are more tall females than in reality, but being a small female is not a particular concern (we note from figure 1 that the self-centered bias is very small or non-existing in female subjects in the lower end of the height distribution).[23] Finally, we can argue that individuals might want to generally think that their levels of happiness is higher than the average, this can explain the fact that average happy subjects perceive a slightly smaller average level of happiness and that perceived happiness is in general smaller than in reality (as we noted from the top left-hand panel of figure 5).

(iii) Salience of the Extremes

Notice also that extreme characteristics (for instance, the very tall or extreme left wing) are generally thought to be more salient than those with average values, which could provide another possible bias. We can reasonably rule out this possibility for mobile brands, in this case there is no reason why the extreme, i.e. the owners of a fringe brand, should be more salient than the others, and there is no evidence to support salience in this case. However, for political stance, happiness, weight and height this is a real possibility.

[23]Consistently with this results, Oswald (2008) finds that perceived height is increasing but concave in real height.

To test this we collected data on perceived averages for the potentially most salient (the top and bottom 10%) in rounds 3 and 4 of the (incentivized) experiment as described in section 2. For example, we asked subjects in question (2b) to report their average for the top 10% heaviest individuals (of their gender). In figure we plot the histograms of the perceived averages in the top 10%, in the bottom 10% of the distributions and the averages in the entire distributions together with the values estimates using our sample, for Political Stance, Happiness, Weight and Height. If there is an effect due to the salience of the extremes we should observe that the perceived averages in the top should be constantly higher than the "real" values and the ones in the bottom should be constantly lower than the real values.

From figure 5, we cannot observe this pattern and these estimations seem in general remarkably accurate. The only cases consistent with the effect due to salience is: the perceived average in the top 10%, which is higher that the real values in female weight; and in the bottom 10% of female height, we observe that the perceived value is smaller than the real value. However, for male height this pattern is reversed: the real average of the top 10% is actually higher than the perceived one, and for the other 9 cases there is no significant difference between the real and the perceived values. From figure 5, we can then argue against the salience explanation for self-centred beliefs.

Closely related to salience is the evidence that subjective probabilities are inverse S-shaped with respect to true probabilities, or more simply that individuals overweight small probabilities and underweight large probabilities.[24] Interestingly enough, our findings might then provide evidence that individuals perceive distributions like probabilities, therefore we could re-interpret the questions we ask in round 2 slightly differently. For example consider being asked what percentage of people might be shorter than yourself. This could be re-interpreted as asking for the probability that a random individual is

[24]There is a large literature on this, but a summary can be found in Camerer, Loewenstein and Rabin (2003).

shorter, and so would be subject to the inverse S-shape relationship. We would then expect to see lower probability events (for instance the chance that someone who is very short being drawn) being overestimated. This would imply a cumulative distribution function biased in a similar way to those we derive from our data and would provide grounds for an interesting characterization of subjective beliefs. We cannot entirely rule out this explanation, however this would not explain why perceived averages are correlated in 5 cases of 6 to individuals own values, as we can observe from figure 4.

5 Applications

The implications and applications that stem from our work are of three types. Firstly the self-centered bias identified in this paper has important implications for economic modelling generally and the interpretation of data derived from economic models. Secondly, there are implications for certain theories, particularly those which have beliefs at their core, such as informational herding or auction theory. Finally, the results from the specific cases we examine, such as height and weight, can also be used directly, for instance when designing government policy to tackle obesity.

To start with the first type of implication, Manski (2004) highlights how departures from rational expectations can leave economists in some difficulty when seeking to identify the correct model through revealed preference. The bias identified in this paper represents such a departure and the ubiquitous nature of our findings suggests that in many behavioral or choice contexts a self-centered bias may be playing a role in forming expectations. It is therefore difficult to know whether any decision is taken through an erroneous belief or through the reasonable use of private information or preferences. On that basis it is hard to draw conclusions concerning which model is correct from choices when the choices themselves may have come from more than one competing model and

the paradigm of revealed preference becomes suspect. To be more specific, consider the example of the ultimatum game used by Manski. In the game the sharing rule is determined by the proposer's expectations concerning the respondent's behavior. Manski shows that if one departs from the assumption that the objective probabilities of the respondent's behavior are known, data cannot identify a single model for play in the game. The self-centered bias identified in this paper plays a similar confounding role since it is plausible that the proposer forms his expectation on the basis of his perception of the distribution of preference for fairness which may be self-centered.

We can also consider lessons that come from our findings for economics and policy-makers. We would warn policy-makers and survey-designers that the assumption that beliefs are on average correct, even concerning such seemingly straightforward characteristics as height or weight, seems woefully inadequate in any context where a self-centered bias may emerge and our findings indicate that such a context may be far more wide-ranging than has hitherto been considered. For example, bias in the perception of the happiness of others might bias individuals' attitudes towards altruism and redistribution and in turn might bias responses in surveys and should lead us to be more careful in the use of such surveys to guide policy. To restate, for important policy decisions or even in the development of new economic theory it makes sense to think about whether biased beliefs will render a model inaccurate or a policy counter-productive, and if so, it makes further sense to think about how to measure the beliefs of the target population. The solution suggested by Manski (2004) is to make greater use of subjective probabilities in survey-based work and our findings lend empirical support to that recommendation.

Next we can examine the ramifications for models which rely on a clear understanding of how beliefs are formed. Biased perceptions of others' choices can lead to bias in your own choice when that choice is itself a function of the choice of others. This is clear for network goods but extends to any goods or services where quality might be

uncertain as discussed in the informational herding literature initiated by Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992). People overstate their own signals or information (believing it is over-representative) so overweight their information when updating. This could be applied to any decision problem but has particular resonance for the experimental herding literature since it can help to explain the well-known finding that individual's overweight their own private signals. This has been a standard feature of sequential herding experiments since Anderson and Holt (1994) and remains a feature of herding experiments in endogenous time (see Sgroi, 2003), herding with efficient prices (see Cipriani and Guarino, 2005, and Drehmann, Oeschler and Roider, 2005) and herding in endogenous time with efficient prices (see Park and Sgroi, 2012). To give another example which links our findings to the social learning literature, in a model of social learning, the perception that individuals are more likely to use their own information than is the case should lead individuals to apply an upward weight to the informational-content of their actions. This would make those in a sequence of decision-makers more likely to overweight the actions of those at the very start of the chain. This fits the story in Guarino and Jehiel (2012) who analyze behavior in a herding model where agents are unable to understand other agents' strategies in their finest detail, and is itself an application of Jehiel's (2005) Analogy-Based Expectation Equilibrium concept.

Economics typically has, at its core, the belief that economic agents are fully aware of the distribution that corresponds to an unbiased distribution surrounding the truth. This may even be a requirement for certain core theories to hold. Take for example the centre-piece of auction theory, the revenue equivalence theorem, which requires that bidders know the true distributions of valuations. The bias identified in this paper would suggest otherwise: bidders are likely to think that other's valuations are more highly correlated with their own than is justified by the truth. The ramifications of such a bias for any policy decision or economic model is considerable: with biases across such

a wide range of characteristics and choices, it is easy to see how policies could go badly wrong or models become misspecified.

Finally, we can also draw from our findings to give direct policy ramifications in specific cases. For example, an overweight individual might perceive themselves to be close to average weight and this might leave them content with their diet and exercise regime. That same individual might choose to exercise more or eat more healthily if they realized their true position in the distribution. The same argument applies symmetrically for individuals at the other extreme of the scale. A direct policy ramification would be that one relatively low cost direct government policy intervention to tackle growing obesity is simply to publicise accurate data on average weight, height and BMI together with "ideal" figures.

Bias in the perception of the political beliefs of others could change political actions, including for instance the decision to vote strategically. If a voter thinks that an extreme party close to her views is more representative than is the case, that party may pick up more votes, rather than loosing votes because of a perception that they have little chance to secure victory. Ironically bias in perception in this case would actually raise the party's votes and so be partially self-confirming. More generally we might model voting behavior as a function of the proximity of the party in question to a voter's own beliefs and a function of the voter's belief that the party will win (in the vein of a Keynesian beauty contest) and our findings certainly suggest that potential voters, even for quite extreme parties, have unreasonably high expectations that their favored party is likely to win.

The perception that the brand of the good I am using is more common that it is in reality may have important consequences in terms of future consumption choices especially for network goods and status goods, but more generally if I think that market share is a good indicator of quality. This is good news for lower market share (niche) items since they are likely to benefit from incorrect beliefs that they are more popular

than is the case.

6 Conclusions

Our results highlight a subtle form of self-deception when individuals consider their own characteristics and choices: those at the extremes seem to perceive themselves as much closer to average than is the case. Our second consistent finding is that estimates concerning averages tend to be a function of individual's own values. The biases we uncover are both powerful and ubiquitous, they apply across a variety of different characteristics, choices and parameters, and apply regardless of the observability of these characteristics. Moreover, our experimental design allows us to use the observability of certain characteristics to differentiate between statistical and behavioral theories. Our subject pool are in constant daily contact with their peers and so should have an unbiased recall of the distribution of height and weight and where they stand within that distribution. However the biases we uncover do not disappear under observability. In this way our findings offer strong support to behavioral theories which work irrespective of observability. We also present numerous applications which should highlight the importance of our findings which range from the specific (policy applications linked to the characteristics we examine) through to the general (the implications for theoretical modelling in specific areas of economics and more generally).

References

[1] Baccara, M., and Yariv, L. (2012). "Homophily in peer groups", forthcoming in *American Ecocomic Journal: Microeconomics*.

[2] Banerjee, A. V. (1992). "A simple model of herd behavior", *Quarterly Journal of Economics* 107: 797-817.

[3] Becker, G. S. (1973). "A theory of marriage: part I", *Journal of Political Economy* 81: 813-46.

[4] Benabou, R., and Tirole, J. (2002). "Self-confidence and personal motivation." *Quarterly Journal of Economics* 117: 871-915.

[5] Benoit, J-P, and Dubra, J. (2011). "Apparent overconfidence." *Econometrica* 79: 15911625.

[6] Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). "A theory of fads, fashion, custom and cultural change as information cascades", *Journal of Political Economy* 100: 992-1026.

[7] Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). "A Theory of Fads, Fashion, Custom and Cultural Change as Information Cascades", *Journal of Political Economy* 100: 992-1026.

[8] Burks, S., Carpenter, J., Goette, L., and Rustichini, A. (2013). "Overconfidence is a social signaling bias." Forthcoming in *The Review of Economic Studies.*

[9] Camerer, C.F., Loewenstain, G., and Rabin, M. (2003). *Advances in Behavioral Economics.* Princeton, New Jersey: Princeton University Press.

[10] Carrillo, J.D., and Mariotti, T. (2000). "Strategic Ignorance as a Self-Disciplining Device", *Review of Economic Studies* 67: 529-544.

[11] Cipriani, M. and Guarino, A. (2005). "Herd behavior in a laboratory financial market" *American Economic Review* 95, 1427-1443.

[12] Clotfelter C.T. and Cook, P.J. (1993) "The gambler's fallacy in lottery play" *Management Science* 1993 39:12, 1521-1525.

[13] Coate, S., and Loury, G.C. (1993). "Will affirmative-action policies eliminate negative stereotypes?", *The American Economic Review* 83: 1220-1240.

[14] Compte, O., and Postlewaite, A. (2004) "Confidence-enhanced performance", *The American Economic Review* 94: 1536-1557.

[15] Cronbach, L.J. (1955). "Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin* 52: 177-193.

[16] Currarino, S., Pin, P., and Jackson, M. (2009) "An economic model of friendship: homophily, minorities and segregation", *Econometrica* 77: 1003-10045.

[17] Drehmann, M., Oeschler J. and Roider, A. (2005). "Herding and contrarian behavior in financial markets: An internet experiment" *American Economic Review* 95, 1403-1426.

[18] DellaVigna, S. (2009). "Psychology and Economics: Evidence from the Field", *Journal of Economic Literature*, June, Vol. 47, pp. 315-372.

[19] Engelmann, D., and Strobel, M. (2012). "Deconstruction and Reconstruction of an Anomaly", *Games and Economic Behavior*, 76(2), 678-689.

[20] Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman & Hall.

[21] Farmer, A. and Terrell, D. (1996). "Discrimination, Bayesian updating of employer beliefs and human capital accumulation", *Economic Inquiry* 34: 204219.

[22] Geanakoplos, J. (1989). "Game theory without partitions, and applications to speculation and consensus", Cowles Foundation Discussion Paper No.914. Available from http://cowles.econ.yale.edu

[23] Golub, B., and Jackson, M. (2011). "Network structure and the speed of learning: measuring homophily based on its consequences", *Annals of Economics and Statistics*, Forthcoming.

[24] Grossman, Z. and Owens, D. "An unlucky feeling: Overconfidence and noisy feedback", *Journal of Economic Behavior & Organization*, Volume 84, Issue 2, November 2012, Pages 510-524

[25] Guarino, A., and Jehiel, P. (2012). "Social learning with coarse inference", *American Economic Journal: Microeconomics*, forthcoming.

[26] Jehiel, P. (2005). "Analogy-Based Expectation Equilibrium", *Journal of Economic Theory* 123: 81-104.

[27] Kagel, J.H., and Roth, A.E. (1995). *The Handbook of Experimental Economics.* Princeton, New Jersey: Princeton University Press.

[28] Manski, C. (2004). "Measuring expectations." *Econometrica* 72, 13291376.

[29] Marks, G. and Miller, N. (1987). "Ten years of research on the false-consensus effect: An empirical and theoretical review." *Psychological Bulletin* 102: 72-90

[30] McPherson, M., Smith-Lovin, L., and Cook, J. (2001). "Birds of a feather: homophily in social networks." *Annual Review of Sociology* 27, 415-44.

[31] Moore, D.A., and Healy, P.J. (2008) "The trouble with overconfidence." *Psychological Review* 115: 502-517.

[32] Morris, S. (1995). "The common prior assumption in economic theory." *Economics and Philosophy* 11: 227-253.

[33] Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118, 1209-1248.

[34] Park, A. and Sgroi, D. (2012). "Herding, contrarianism and delay in financial market trading" *European Economic Review* 56, 1020-1037.

[35] Oswald, A. (2008). "On the curvature of the reporting function from objective reality to subjective feelings." *Economics Letters* 100, 369-372.

[36] Robbins, H. (1952). "Some aspects of the sequential design of experiments." *Bulletin of the American Mathematical Society* 55, 52735.

[37] Ross, L., Greene, D., and House, P. (1977). "The 'false consensus effect': an egocentric bias in social Perception and Attribution Processes." *Journal of Experimental Social Psychology* 13, 279-301.

[38] Rothschild, M. (1974). "A Two-Armed Bandit Theory of Market Pricing." *Journal of Economic Theory* 9, 185202.

[39] Rabin, M. (2002)"Inference by Believers in the Law of Small Numbers" *Quarterly Journal ofEconomics* 117 (3): 775-816

[40] Sgroi, D. (2003). "The right choice at the right time: A herding experiment in endogenous time" *Experimental Economics* 6, 159180.

[41] Tversky, A., and Kahneman, D. (1973). "Availability: a heuristic for judging frequency and probability." *Cognitive Psychology* 5, 207-232.

[42] Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and biases". *Science*, 185, 1124, 1128-1130.

Tables and Figures

Table 1: Main Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Weight Female	55.71	7.007	43	73	69
Weight Male	72.427	12.481	45	109	82
Height Female	162.418	7.237	138	181	67
Height Male	179.613	9.858	151	208	80
Political Stance	3.805	1.166	1	7	154
Happiness	4.766	1.021	1	7	154
Male	0.545	0.5	0	1	154

Figure 1: **Weight and Height:** The left panels represent the real CDFs (continuous lines) and the beliefs over the distributions (dots) and their respective Lowess function (dashed lines). The right panels represents the 45 degree lines (solid lines) and the local polynomial interpolation of the perceived and the real distributions (dashed lines), the shadow represents the 95% confidence interval.

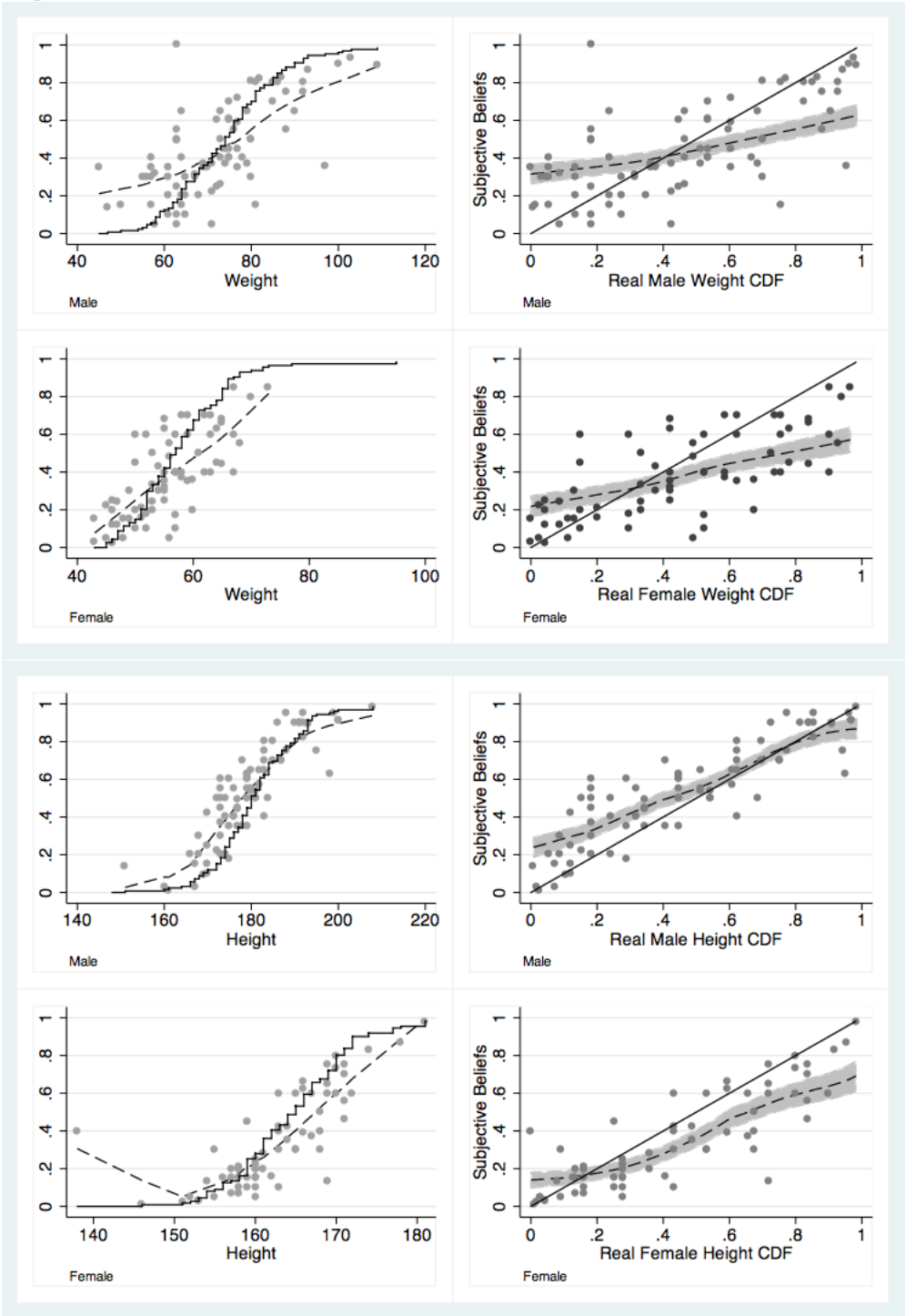


Figure 2: **Happiness and Political Stance:** The left panels represent the real CDFs (continuous lines) and the beliefs over the distributions (dots) and their respective Lowess function (dashed lines). The right panels represents the 45 degree lines (solid lines) and the local polynomial interpolation of the perceived and the real distributions (dashed lines), the shadow represent the 95% confidence interval.

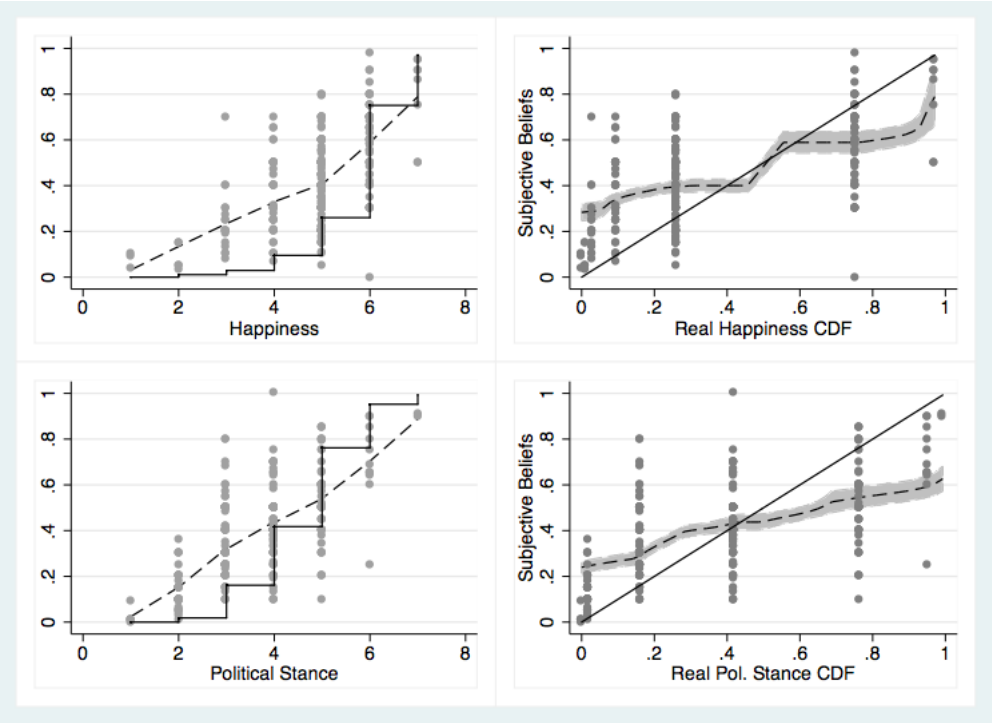


Figure 3: **Mobile Phones and Restaurant Choices:** The left panels represent the histograms of the simple real distributions (with the different characteristics ordered by frequencies), the beliefs over the distribution (dots) and their respective Lowess function (dashed lines). The right panels represents the 45 degree lines (solid lines) and the local polynomial interpolations of the perceived and the real simple distributions (dashed lines) with the shadow representing the 95% confidence interval.

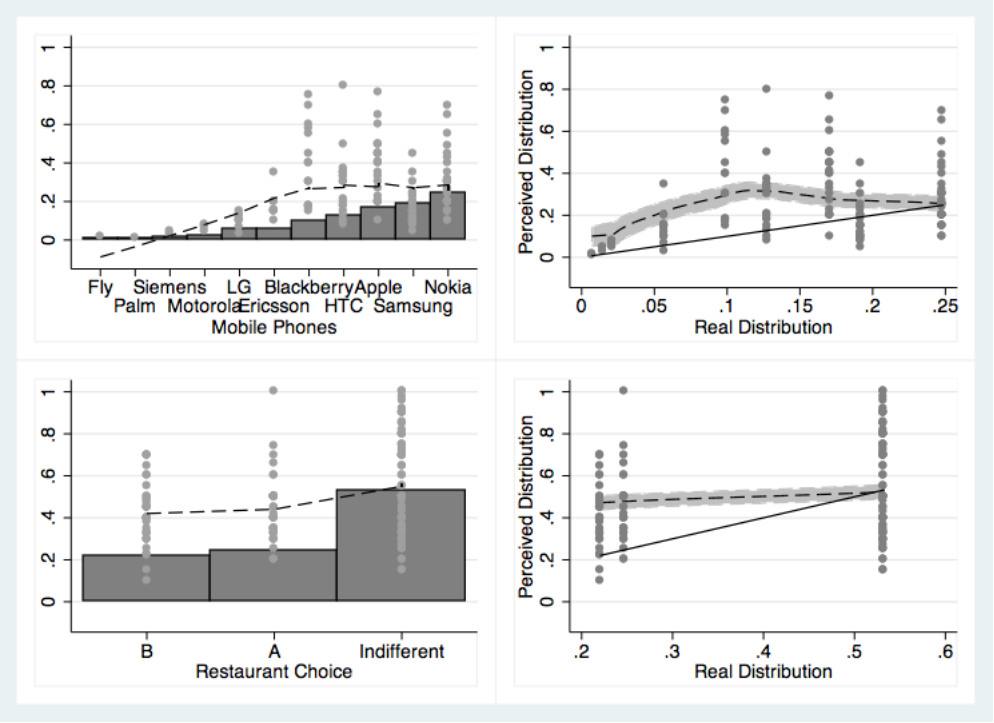


Table 2: **Determinants of Perceived Cumulative Distributions**. Dependent Variables: Perceived Share of Individuals below each individual Happiness, Political Stance, Weight and Height. Robust Standard Errors in Brackets.

	1 Happiness CDF b/se	2 Pol. Stance CDF b/se	3 Weight CDF b/se	4 Height CDF b/se
Happiness	0.4569***			
	(0.0415)			
Political Stance		0.4809***		
		(0.0372)		
Female*Weight			0.5502***	
			(0.0585)	
Male*Weight			0.5538***	
			(0.0682)	
Female*Height				0.7469***
				(0.0627)
Male*Height				0.7907***
				(0.0479)
Male	0.0053	0.0055	0.0612	0.1423***
	(0.0215)	(0.0204)	(0.0481)	(0.0394)
Constant	0.2671***	0.2100***	0.1327***	0.0245
	(0.0195)	(0.0183)	(0.0301)	(0.0277)
r2	0.336	0.399	0.484	0.783
N	273	267	151	147

Table 3: **Determinants of Perceived Distributions**. Dependent Variables: Perceived Share of Individuals with the same mobile-phone brand and the same information processing behavior. Robust Standard Errors in Brackets.

	1 Mobile Distribution b/se	2 Deference Distribution b/se
Mobile	0.4288**	
	(0.1825)	
Deference		0.4330***
		(0.1060)
Male	–0.0557*	0.0141
	(0.0289)	(0.0324)
Constant	0.2251***	0.3217***
	(0.0401)	(0.0394)
r2	0.059	0.097
N	141	154

Figure 4: **Beliefs about averages of characteristics plotted against the individuals' own characteristics:** the solid lines represent the the linear interpolating function and the grey area represent the 95% confidence interval. The solid lines parallel to the axes represent the average values in our sample, in the y axis we also reported the confidence interval of the estimated averages

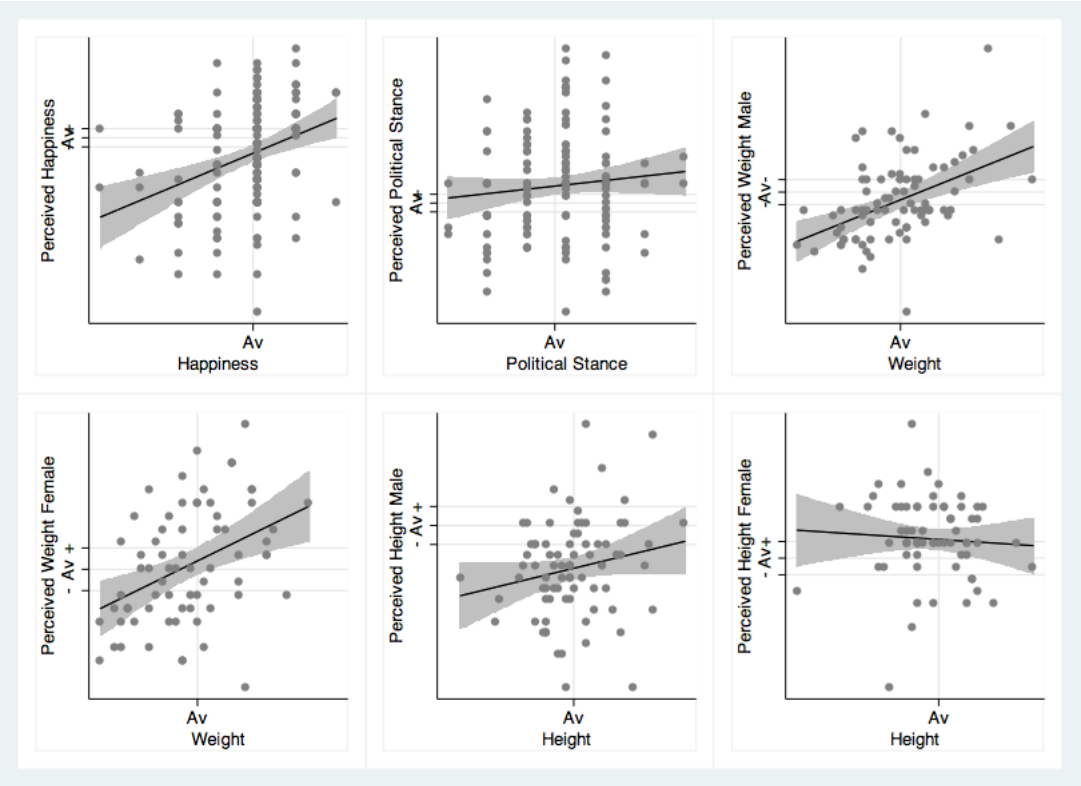
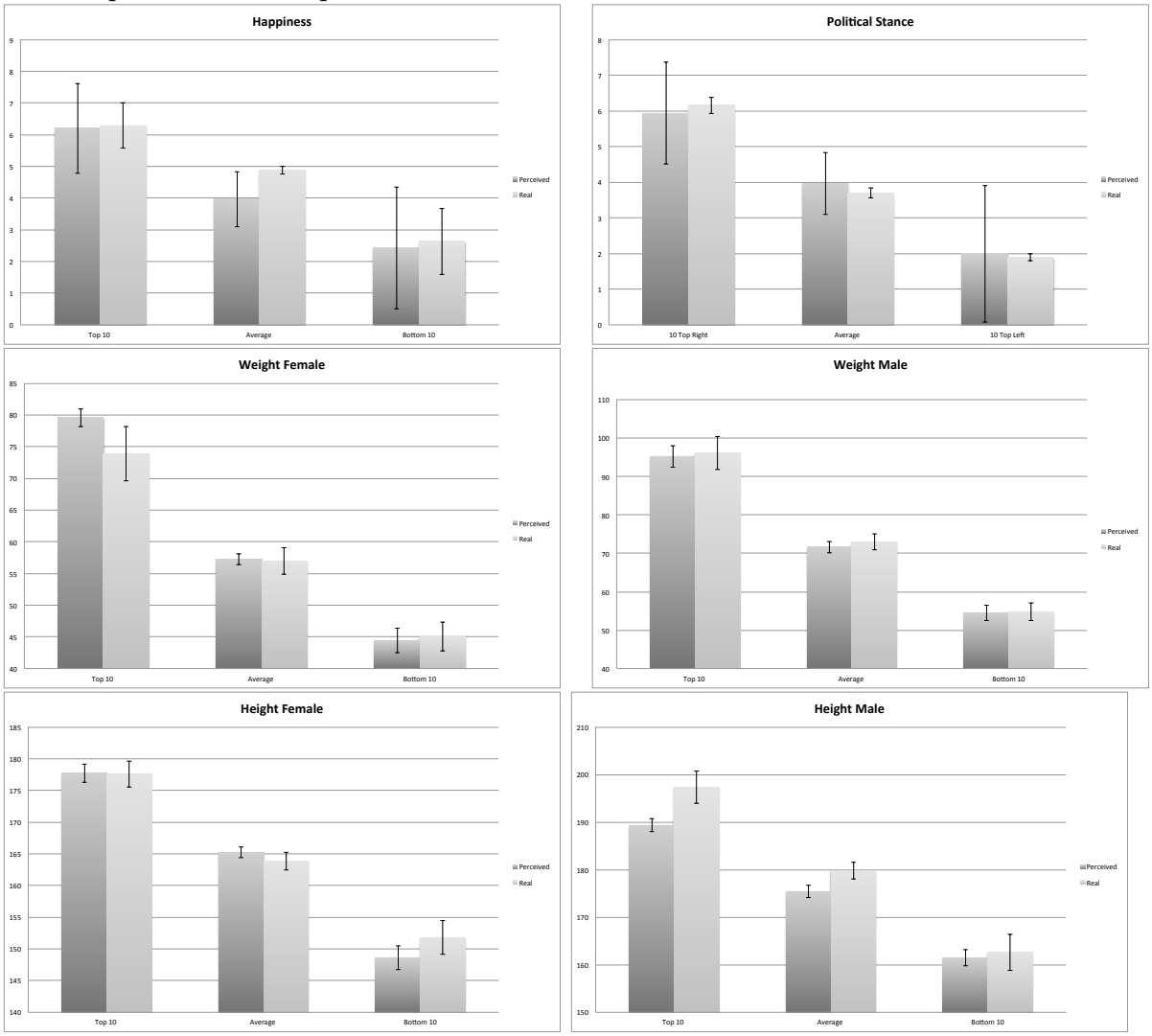


Figure 5: **Top, Bottom and Averages Beliefs:** For each characteristics we can observe the perceived averages of the different sections of the distributions and the real values in our sample. The bars represent the 95% confidence intervals.



NOT INTENDED FOR PUBLICATION

Appendix: Instructions

This Appendix presents the full on-screen instructions and questions for the entire experiment in the precise order observed by participants. The "session ID" box was automatically completed based on their initial login credentials. The questions marked with a "*" indicate required questions. Once they had answered each required question they could then proceed by hitting the "submit answers" button at the bottom of each web-page. After round 7 they were taken to a final page where they were asked to wait patiently until their fellow participants had finished before receiving payment.

ROUND 1

ID ENTRY
Session ID:*
All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

ROUND 1 QUESTIONS
Please answer all of the following questions to the best of your ability and please scroll down through all of the questions until you reach the "SUBMIT ANSWERS" button.

What is your gender?
Your answer:*

What is your height?
The menu includes heights measured metric and imperial measures, so you can choose the units with which you are most comfortable.
Your answer:*

What is your weight at the moment?
The menu includes weights measured metric and imperial measures, so you can choose the units with which you are most comfortable.
Your answer:*

How would you rate your happiness at the moment?
Please use a 7-point scale where 1 is completely sad, 2 is very sad, 3 is fairly sad, 4 is neither happy nor sad, 5 is fairly happy, 6 is very happy and 7 is completely happy.
Your answer:*

How would you rate your political beliefs at the moment?
Please use a 7-point scale where 1 is far left, 2 is left, 3 is centre left, 4 is centre, 5 is centre right, 6 is right and 7 is far right.
Your answer:*

What is your current brand of mobile phone? If you have more than one enter the one that you use most often.
Your answer:*

Imagine that you have to decide between two restaurants in which to have dinner alone. They are called restaurant A and B. You have some private information that A is better, but you know that an equally well-informed colleague has information suggesting that B is better. Would you choose to eat at A, B or are you indifferent?
Your answer:*

END OF ROUND 1
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will only be linked to your session ID and not to your personal identity or your university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)

ROUND 2

ID ENTRY
Session ID:*
All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

VERY IMPORTANT - BONUS PAYMENT EXPLANATION - PLEASE READ!
You will receive a certain £2.50 for showing-up but by answering the next few questions you have the chance to win a bonus. One of the questions on this page and the ones to follow will be chosen at random and secretely allocated to be the "prize question". If your answer is within 10% of the correct answer to the prize question you will win a bonus payment of £5.

ROUND 2 QUESTIONS
Please answer all of the following questions to the best of your ability as one of these may be the "prize question" and please scroll down through all of the questions until you reach the "SUBMIT ANSWERS" button.

What percentage of students at Warwick do you think are less happy than you are?
Your answer:*

What percentage of students at Warwick do you think are less right-wing than you?
Your answer:*

What percentage of students at Warwick of your gender do you think are shorter than you?
Your answer:*

What percentage of students at Warwick of your gender do you think are lighter than you?

Your answer:*

Think again about the mobile phone brand you listed in the last round. What percentage of students at Warwick do you think also use the same brand of mobile phone as their main mobile phone?
Your answer:*

Think again about the restaurant question in the last round. To remind you, you had to decide between two restaurants in which to have dinner alone. They were called restaurant A and B. You had some private information that A is better, but you knew that an equally well-informed colleague had information suggesting that B was better.
What percentage of your fellow Warwick students do you think would have chosen the same answer as you? Remember that the options were "indifferent", "A" or "B".
Your answer:*

END OF ROUND 2
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will only be linked to your session ID and not to your personal identity or your university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)

ROUND 3

ID ENTRY
Session ID:*
All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

ROUND 3 QUESTIONS

Please answer all of the following questions to the best of your ability as one of these may be the "prize question" and please scroll down through all of the questions until you reach the "SUBMIT ANSWERS" button.

What do you think is the average height for someone who is in the 10% tallest Warwick students of your gender?
Your answer:*

What do you think is the average weight for someone who is in the 10% heaviest Warwick students of your gender?
Your answer:*

What do you think is the average happiness for someone who is in the 10% happiest students at Warwick?
Please use a 7-point scale where 1 is completely sad, 2 is very sad, 3 is fairly sad, 4 is neither happy nor sad, 5 is fairly happy, 6 is very happy and 7 is completely happy.
Your answer:*

What do you think is the average political belief for someone who is in the 10% most right-wing students at Warwick?
Please use a 7-point scale where 1 is far left, 2 is left, 3 is centre left, 4 is centre, 5 is centre right, 6 is right and 7 is far right.
Your answer:*

END OF ROUND 3
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will only be linked to your session ID and not to your personal identity or your university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)

ROUND 4

ID ENTRY
Session ID:*

All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

ROUND 4 QUESTIONS
Please answer all of the following questions to the best of your ability as one of these may be the "prize question" and please scroll down through all of the questions until you reach the "SUBMIT ANSWERS" button.

What do you think is the average height for someone who is in the 10% shortest Warwick students of your gender?
Your answer:*

What do you think is the average weight for someone who is in the 10% lightest Warwick students of your gender?
Your answer:*

What do you think is the average happiness for someone who is in the 10% most sad at Warwick?
Please use a 7-point scale where 1 is completely sad, 2 is very sad, 3 is fairly sad, 4 is neither happy nor sad, 5 is fairly happy, 6 is very happy and 7 is completely happy.
Your answer:*

What do you think is the average political belief for someone who is in the 10% most left-wing students at Warwick? Please use a 7-point scale where 1 is far left, 2 is left, 3 is centre left, 4 is centre, 5 is centre right, 6 is right and 7 is far right.
Your answer:*

END OF ROUND 4
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will only be linked to your session ID and not to your personal identity or your

university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)

ROUND 5

ID ENTRY
Session ID:*
All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

ROUND 5 QUESTIONS
Please answer all of the following questions to the best of your ability as one of these may be the prize question and please scroll down through all of the questions until you reach the SUBMIT ANSWERS button.

What do you think is the average height for a Warwick student of your gender?
The menu includes heights measured metric and imperial measures, so you can choose the units with which you are most comfortable.
Your answer:*

What do you think is the average weight for a Warwick student of your gender?
The menu includes weights measured metric and imperial measures, so you can choose the units with which you are most comfortable.
Your answer:*

What do you think is the average happiness at the moment for a Warwick student?
Please use a 7-point scale where 1 is completely sad, 2 is very sad, 3 is fairly sad, 4 is neither happy nor sad, 5 is fairly happy, 6 is very happy and 7 is completely happy.
Your answer:*

What do you think is the average political belief for a Warwick student? Please use a 7-point scale where 1 is far left, 2 is left, 3 is centre left, 4 is centre, 5 is centre right, 6 is right and 7 is far right.
Your answer:*

Please estimate the percentage of your fellow Warwick students who use each of the following mobile phone brands as their primary mobile phone.

Note that the list is in alphabetical order (except for the "other" category) and you should insert an entry for each brand (including "other").

Note concerning payment: if this question is the one that is randomly chosen to be the payment question then a single brand will be chosen randomly and if you are within 15% of the correct answer you will receive the £5 bonus payment.

(Participants were then presented with a list of mobile phone brands in the following order: Apple Iphone, Blackbery, HTC, LG, Motorola, Nokia, Samsung, Siemens, Sony-Ericson, Other. For each one they were asked to select from a list of 5% intervals, beginning 0 to 5% and ending with 95 to 100%)

Think again about the restaurant question you were asked earlier in the session. To remind you, you had to decide between two restaurants in which to have dinner alone. They were called restaurant A and B. You had some private information that A is better, but you knew that an equally well-informed colleague had information suggesting that B was better.

What percentage of your fellow Warwick students do you think would have chosen to eat at restaurant A if they were asked the same question? Remember that the other options were indifferent and B.
Your answer:*

END OF ROUND 5
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will only be linked to your session ID and not to your personal identity or your

university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)

ROUND 6

ID ENTRY
Session ID:*
All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

ROUND 6 QUESTION
Please answer the following question to the best of your ability as it may be the "prize question" and please scroll down until you reach the SUBMIT ANSWERS button.

Consider the following gamble. You have a 20% chance of winning £100, a 40% chance of winning £10 and a 40% chance of winning £0. If you played this gamble many times what would you expect to be your average winnings per gamble? (in pounds)
Your answer:*

END OF ROUND 6
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will only be linked to your session ID and not to your personal identity or your university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)

ROUND 7

ID ENTRY
Session ID:*
All of your answers will be entirely anonymous. You have been allocated a Session ID and password, for example your ID might be "48576" or "60306". This ID and password are unique to you and will help us link the data you enter across several forms. However both the session ID and password are entirely anonymous and cannot be used to determine your personal identity or university ID.

ROUND 7 QUESTIONS
This is the final round. Please answer the questions to the best of your ability and please scroll down through all of the questions until you reach the "SUBMIT ANSWERS" button. Note that none of the following questions are possible "prize questions".

What is your age? (in years)
Your answer:*

What is your nationality?
Your answer:*

What is your degree subject?
Your answer:*

Did you study maths up to and including your final year at school? (e.g. to A-level or IB SL or HL?)
Your answer*

Please comment on the methods you used (if any) during the rounds when you had the opportunity to win bonus payments.
Your answer:

END OF ROUND 7
Privacy statement
REMINDER: All of your answers today will be entirely anonymous. No data which personally identifies you is collected on the form. Your answers will

only be linked to your session ID and not to your personal identity or your university ID. When you are ready to submit your answers please hit the "SUBMIT ANSWERS" button below.
(SUBMIT ANSWERS BUTTON)