Jan 2018 (revised May 2018 and Sep 2019) No.359

Ignoring Good Advice

David Ronayne and Daniel Sgroi

WORKING PAPER SERIES

This paper has been revised September 2019 previous version titled
'When Good Advice is Ignored: The Role of Envy and Stubbornness' originally circulated January 2018 and previous update May 2018.

Centre for Competitive Advantage in the Global Economy

Department of Economics

Ignoring Good Advice*

David Ronayne† and Daniel Sgroi‡

First version: 17 January 2018

This version: 5 September 2019

Abstract

We ran an experiment where 1,503 subjects (advisees) completed tasks, and then had the choice to submit either their own score or that of other, high-scoring subjects (advisers). Good advice was ignored: about 25% of the time, advisees chose to submit their own score instead of higher-scoring advisers, reducing their payoff. When the adviser was superior in skill (rather than luck), good advice was ignored more often. When the adviser was relatively highly paid, subjects were less likely to make use of them. We offer an explanation of the data focused on two behavioral forces: envy and the sunk cost fallacy. The role of envy was complex: more envious advisees, as measured using a dispositional envy scale, opted to follow advisers more often in the skill-based task revealing a positive, motivational effect of envy. However, higher adviser remuneration reduced this effect, demonstrating a negative side of envy. Susceptibility to the sunk cost fallacy had a negative impact on the uptake of good advice. This is consistent with the idea that subjects feel resistant to changing their answers when they put in effort to formulate them. We also present findings from a survey of 3,096 UK voters who took part in the national referendum on EU membership, which are consistent with some of our experimental results. (*JEL: C91, C99, D91*)

Keywords: experiment, individual decision-making, relative comparisons, inter-personal comparisons, good advice, skill vs. luck, adviser remuneration, envy, sunk cost fallacy, stubbornness, Brexit survey.

*The authors thank Johannes Abeler, Gordon Brown, Alexander Cappelen, Vince Crawford, Andrew Dilnot, David Gill, Andrea Isoni, Ole Jann, Meg Meyer, Nattavudh Powdthavee, Neil Stewart, Sevgi Yuksel, Bill Zame, Florian Zimmermann, Daniel Zizzo, and participants of various seminars for helpful comments. The authors also thank Nuffield College for funding. The experiment described in this paper was pre-registered at the AEA RCT Registry on June 28, 2017: `https://doi.org/10.1257/rct.2022-3.0`. Ethical approval for the experimental work was obtained from the Department Research Ethics Committee of the University of Oxford on January 20, 2017 under reference ECONCIA17-18-001 and from the Centre for Experimental Social Sciences of the University of Oxford on February 14, 2017 under reference ETH-170118191-1.

†University of Oxford, david.ronayne@economics.ox.ac.uk

‡University of Warwick, daniel.sgroi@warwick.ac.uk

"America has always been a country of amateurs where the professional, that is to say, the man who claims authority as a member of an elite which knows the law in some field or other, is an object of distrust and resentment"

W. H. Auden, Faber Book of Modern American Verse, 1956.

"People in this country have had enough of experts"

Michael Gove MP, UK Justice Secretary, 2016.

1 Introduction

People are not experts in every domain. As such, individuals go through life adopting the advice of others in order to make better decisions in all manner of areas e.g. medical, legal, educational, social, political, financial, religious etc. However, it seems as though the advice of experts is often ignored, as some recent examples across the globe highlight. Both the debate in the UK surrounding "Brexit" (the national referendum on Britain's EU membership held in 2016) and the 2016 US presidential election indicated an increasing distrust of expert opinion among voters. One of the UK's most respected scientists, Stephen Hawking, described Brexit, which was opposed by many professional groups including doctors, scientists and economists,[1] as "the moment when the forgotten spoke, finding their voices to reject the advice and guidance of experts and the elite everywhere."[2] In the run up to the 2016 presidential election, groups of prominent experts in various fields, e.g., scientists and economists,[3] spoke up against the potential election of Donald Trump; advice insufficient to prevent him winning the election. More generally, the marketing firm Edelman has run large-scale international surveys for 18 years and recently reported distrust of a variety of different expert bodies across the world. Their survey of over 33,000 individuals across 28 countries documents that 60% of respondents found peers, defined as "people like you", to be as useful as academic or technical experts (see Edelman, 2017). However, signs that expert advice may often go ignored stretch back to well before the present day. For example, W. H. Auden, the noted British poet who moved to the USA in 1939 seemed to believe that the distrust of experts is something of a national characteristic of the USA (see the quote above).

There are many reasons why advice may be ignored. For example, there is nearly always uncertainty over whether advice from a given source is good. We define advice from one individual (an "adviser") to another (an "advisee") as "good" if accepting the advice increases the

[1] See respectively: The Guardian, June 15, 2016 `https://www.theguardian.com/politics/2016/jun/15/eu-referendum-british-medical-journal-urges-doctors-to-vote-remain`, the BBC, June 11, 2016 `http://www.bbc.co.uk/news/uk-36505736` and The Guardian, May 28,2016 `https://www.theguardian.com/politics/2016/may/28/economists-reject-brexit-boost-cameron`.

[2] The Express, December 3, 2016, `http://www.express.co.uk/news/politics/739320/Brexit-Stephen-Hawking-Remain-MPs-subvert-EU-referendum-result`.

[3] See respectively: Reuters, September 20, 2016 `https://www.reuters.com/article/us-usa-election-scientists/more-than-300-scientists-warn-over-trumps-climate-change-stance-idUSKCN11Q2PD` and Time, September 21, 2016 `http://time.com/4555032/economists-against-donald-trump/`

expected material payoff of the advisee, or "bad" if it reduces it. In reality, if advice is perceived as bad e.g., as a result of coming from a biased or corrupt source, it can of course be rational to ignore it (see e.g., Cain, Loewenstein, and Moore, 2005). In contrast, if *good* advice is ignored, it is a concern for society at large as well as for experts themselves, including economists. In this paper, we study whether, when and why individuals may irrationally ignore good advice.

To address these questions, we conducted a pre-registered online experiment.[4] In total, we recruited 1,578 participants through Amazon's Mechanical Turk, a relatively well-explored (see e.g., Paolacci and Chandler, 2014; Paolacci, Chandler, and Ipeirotis, 2010) and commonly-used participant pool in social science, including economics (e.g., DellaVigna and Pope, 2017; Kuziemko, Norton, Saez, and Stantcheva, 2015). Running an experiment allowed us to control for some of the complicating issues likely to arise with field data including uncertainty over the quality and implications of advice, perverse or corrupt incentives, dynamic issues such as history or reputation, and any confounding issues inherent to some particular context. At the same time, we aimed to construct an experiment possessing the main characteristics of advice-taking scenarios: A choice between the two probability distributions implied by following, and not following, the advice on offer where these two distributions are inextricably linked to (generated by) the adviser and the advisee, respectively. In our experiment, individuals completed tasks where their responses generated their probability of winning a bonus payment. They were then presented with the opportunity to change their probability of winning a bonus payment to be the same as another participant's (an adviser's) by using the other's responses instead of their own. Therefore, the advice offered to a participant was good (bad) when the adviser's probability of winning the bonus was higher (lower) than the participant's.[5] How accepting the advice would alter their expected payoff was explained to participants: raising it in the case of good advice; lowering it in the case of bad advice. Our first main finding is that in our experiment, while $< 3\%$ of bad advice was accepted, good advice was ignored about 25% of the time.

To better understand when good advice is ignored, we conducted novel treatments both within-subject and between-subject. To motivate our treatments, consider that in a free-market economy more skillful individuals can be expected to be remunerated more highly. Therefore, it is important to understand whether the skill or remuneration of individuals per se impacts the uptake of any advice they provide. Our within-subject treatment was designed to measure the impact of the relative skill of the adviser. Each participant, including those who were chosen as advisers, completed two tasks that differ in their relative levels of luck and skill. The more luck-based task required subjects to guess "heads or tails" in a sequence of coin tosses. The more skill-based task was ten questions from the Raven's visual IQ test. These tasks generated participants' respective probabilities of winning a bonus payment. We found that good advice was ignored more often when the adviser's superior position had been generated through higher skill. Our between-subject treatment was designed to measure the impact of the relative remuneration of the adviser. In our control condition, the participants who were chosen to be advisers were paid \$0.50 (the same bonus that was available to the advisee participants).

[4]See the AEA RCT Registry entry: `https://doi.org/10.1257/rct.2022-3.0`

[5]Note that neutral language was used in place of words such as "advice" or "adviser".

In our first between-subject treatment, advisers were instead paid \$100. On average, this high remuneration treatment *reduced* the propensity to accept good advice by 5.5 percentage points. In our second between-subject treatment, advisers were remunerated as in the control condition with the addition that the advisee could directly affect the adviser's payoff: the adviser was paid an additional \$0.25 for each participant that accepted their advice. In contrast to the first remuneration treatment, we found no evidence of an effect here. Together, the results are consistent with the hypothesis that the decision to ignore good advice is not affected by a desire to affect the other individual's pay, but rather by the scale of the remuneration.

Our design produces variation in the gain in expected payoff from accepting advice, which allowed us to see how the propensity to accept good advice varied with the value of advice. Here we found the more valuable the advice, the more likely participants were to accept it. This suggests that participants traded-off rationality with other factors. We investigated particular behavioral explanations which arise from our conceptualization of advice-taking. We suppose that the decision of whether to accept or ignore advice boils down to a choice between two lotteries with specific properties. First, the set of outcomes and payoffs are the same in each lottery; only the probabilities differ. Second and crucially, the lotteries are inextricably linked to individuals: one is generated by the advisee, one by the adviser. In this framework, when an advisee can choose a lottery (A) generated by an adviser which stochastically dominates the lottery they themselves generated (B), we define A as "good advice". Therefore, in the case where good advice is offered, an advisee compares their inferior probability distribution to the adviser's superior probability distribution. Therefore, we hypothesize that there may be a role for the fundamental human trait of envy to arise during an advisee's comparison of their lottery's probabilities and other attributes of themselves, against those of the adviser. For our study, we measured envy by requiring participants to complete the eight-question "dispositional envy scale" of Smith, Parrott, Diener, Hoyle, and Kim (1999). To our knowledge, we are the first to employ such a scale in economics. The second determinant we consider is an individual's susceptibility to the sunk-cost fallacy. In the real world (and our experiment) individuals spend resources e.g., time, money, emotion or effort, in forming their positions on issues. When presented with a superior viewpoint, a rational individual would rather adopt it than ignore it. However, where individuals suffer from the sunk-cost fallacy, there may be some resistance to moving away from their existing position. In order to measure this potential effect we built the first (to our knowledge) scale designed to measure susceptibility to the sunk cost fallacy, inspired by the work of Thaler (1999) and Arkes and Blumer (1985). Also, to separate the role of the sunk costs from a more general notion of stubbornness, we included a third (and final) psychological measure, taken from Wilkins (2015).[6]

We find that envy played a varied role in the decision of whether to accept good advice. When the adviser was superior in skill, we found that envy was a significant determinant of whether good advice was followed. Here, we found a positive role of envy on average: a higher dispositional level of envy was positively associated with the acceptance rate of good advice.

[6]These three behavioral determinants (envy, sunk-cost fallacy and stubbornness) were the only three determinants we intended to measure before conducting the experiment, as documented in our pre-trial registration (Ronayne and Sgroi, 2017).

This positive role of envy is supportive both of the notion that envy can be harnessed as fuel for self-betterment, and notions of inequality aversion e.g., Fehr and Schmidt (1999). However, this average effect masks a negative interaction effect with adviser remuneration. In the control condition where subjects could win the same remuneration as the adviser received, a higher level of dispositional envy was associated with a higher propensity to accept advice. In the high-remuneration treatment however, the effect of envy was significantly reduced, becoming insignificantly different from zero. Envy of high adviser remuneration prompted subjects to shun advice to such an extent that there was no longer evidence of a positive effect of envy. This interaction showcases both sides of envy and is inconsistent with the aforementioned notion of inequality aversion. Furthermore, the size of the effects of envy were meaningful. When advisers were paid the same as was available to the advisees i.e., where we found a positive role for envy, we predict that the least and most envious in our sample follow good advice 71% and 93% of the time respectively, a difference of 22 percentage points. We also found that our measure of susceptibility to the fallacy was robustly negatively associated with the propensity to take good advice. This measure also had a substantial effect: we predict the least susceptible in our sample to follow good advice 21 and 16 percentage points more often than the most susceptible, in the luck and skill tasks respectively. In contrast, we did not find the more general measure of stubbornness to have predictive power.

In addition to our experiment, we report results from a new panel of 3,096 voters in the UK's 2016 referendum on membership of the European Union. We asked members of the panel questions about the influence of expert opinion on their vote, their own opinions regarding expert remuneration, and their opinions on the link between influence and remuneration. Though we cannot hope to exercise the same level of control in a survey as in an experiment, we note that the survey's results are supportive of the results of our experiment's between-subject treatment. Specifically, we found a negative correlation between the reported influence of expert advice and agreement that expert remuneration is too high.

The paper proceeds as follows. Section 2 discusses related literatures. Section 3 provides a model to formalize our notion of advice and derive the econometric specification we use. Section 4 details the experimental design and highlights its key features. Section 5 provides the results. Section 6 presents our Brexit survey findings. Section 7 concludes.

2 Literature

Our work draws and expands on several important areas of literature both within and outside economics.

Whether or not people take advice has been studied with regards to many specific contexts and factors, using various methodologies. In finance, Önkal, Goodwin, Thomson, Gönül, and Pollock (2009) find that stock-price forecasts can be favored when they are made by humans rather than machines. Mullainathan, Noeth, and Schoar (2012) document that financial advisers do not de-bias their clients, which can often lead to higher profits for advisers. On the perceived credibility of experts on policy issues, Lachapelle, Montpetit, and Gauvin (2014) show that the way in which experts frame information matters, while Doberstein (2016, 2017) finds large differences depending on the type of institution publishing the work (academic, think tank or advocacy group). Hilger (2016) shows theoretically how providers of credence goods will make use of their informational advantage to overcharge consumers in equilibrium. Medical surveys have long documented a significant proportion of patients not following advice (see e.g., Davis, 1968; for a review of patient non-adherence see Kardas, Lewek, and Matyjaszczyk, 2013). The face-to-face experiments of the "judge-adviser system" have been used to identify various determinants of advice-taking related to sense and appearance such as adviser confidence (Swol and Sniezek, 2005). More generally with uncertainty over the quality of advice, many issues arise, including the discounting of advice relative to one's own opinion (Weizsäcker, 2010; Yaniv and Kleinberger, 2000), which has been shown to vary with task difficulty (Gino and Moore, 2007), among others e.g., overconfidence, risk attitudes etc. Ding and Schotter (2015) and Chaudhuri, Schotter, and Sopher (2009) explore the role advice can play in establishing truthful behaviour and enabling coordination respectively. The related question of whether people take *bad* advice has also been studied and it has been found that bad advice is accepted in certain contexts e.g., Powdthavee and Riyanto (2015). Charness, Oprea, and Yuksel (2018) look at biased information sources where recipients can select among them. Also related, Cook and Lewandowsky (2016) document a polarization of the beliefs of US participants in response to being presented with information concerning the consensus in the scientific community regarding anthropogenic global warming.

An underlying assumption in all these works is that if advice were known to be good, it would be taken. In contrast, our experiment provides the first abstracted setting with neutral, context-free language to study the questions of whether, when and why good, unbiased advice is ignored. To do so, we conceptualize the advisee's position and the adviser's advice as probability distributions, the general idea of which dates back to papers such as Morris (1974). We also control for the quality of advice and remove uncertainty over whether advice is good. More precisely, when advice is good (or bad) there is a simple stochastic dominance ordering of the advisee's position and the adviser's advice. This makes the rational action unambiguous and removes the scope for overconfidence and risk attitudes to play a role. We conduct novel treatments to determine the effects of the adviser being superior in skill vs. luck, and the remuneration of the adviser. We utilize our experimental setting to identify these effects, which would be difficult to isolate in the real world: whether an adviser reached their position through relative luck or skill, and in what combination, is difficult to estimate and likely endogenous to the value of their advice; the remuneration of an adviser is potentially observable, but is likely endogenous to the quality of the adviser, as well as to any corruption or bias. Finally, we provide at least a partial explanation of why good advice may not be followed in terms of underlying psychological measures, which we discuss next.

According to Parrott and Smith (1993), envy "occurs when a person lacks another's superior quality, achievement, or possession and either desires it or wishes that the other lacked it." This definition presents the essential dichotomy arising from envy itself: it can lead to an active attempt to diminish others or an active attempt to improve oneself. The philosopher Bertrand

Russell argued that envy was quite possibly the root cause of unhappiness in Western society, but he too highlighted the potential dual effects arising from envy: arguing that while it brings unhappiness, it can also be harnessed to positive ends. Indeed, he cited the rise of democracy as a tangible good that has come from envy (see Russell, 1930). Van de Ven, Zeelenberg, and Pieters (2009) take this a step further, pushing the idea that envy is really two separate concepts: a positive force for self-betterment, and a negative force of self-destruction. If envy really is one of the key driving forces of Western society and the primary cause of (un)happiness, then it might seem odd that it has received surprisingly little attention from economists.[7] However, this may relate to the ambiguity over the implications of the concept. For example, in our context, if individuals are envious of the ability or income of experts they might wish to take good advice in order to raise their own payoff and narrow the gap, consistent with notions of inequality aversion e.g., Fehr and Schmidt (1999). On the other hand, it might be that envy results in an emotional response to shun advice, even if it harms an individual's own payoff.

Within economics, envy has been studied in a variety of contexts though nothing directly related to our own. Following Brenner (1987) and Kuziemko, Norton, Saez, and Stantcheva (2015) we know that individuals care about their relative economic status, and Elster (1987) even makes the further refinement that some may be motivated by the desire to avoid generating envy in others. Falk (2017) links envy to status concerns which he argues provokes moral disengagement and possibly violence. The positive and negative sides of envy are examined by Gershman (2014) as potential factors in driving economic growth. In their analysis of a panel of 18,000 individuals, Mujcic and Oswald (2018) show that envy is a powerful predictor of falling levels of mental well-being, especially among young adults. Related, Winkelmann (2012) finds that the prevalence of luxury cars in a municipality has negative consequences for life satisfaction which he attributes to envy. Within welfare economics, a number of papers have examined the normative significance of envy, for instance Baumol (1986) as part of his broader examination of fairness, and Foley (1967) and Varian (1974) who consider the implications of an envy-free welfare equilibrium. Banerjee (1990) examines how the distortions caused by envy might be removed through progressive income tax. Brennan (1973) makes the point that envy can motivate support for redistribution from those who want to see the rich made poorer, and Mui (1995) examines the envy that followers may feel towards an innovator. Within experimental economics, envy is also linked to the documented "money-burning" phenomenon (e.g., Zizzo and Oswald, 2001) where players in a game are willing to damage their own utility in order to punish others, and Kirchsteiger (1994) attempts to use envy as a possible explanation for behavior in the ultimatum game. More generally, money-burning and envy are linked to the literature on fairness, for example, Fehr and Schmidt (1999) suggest that individuals may possess a "willingness to sacrifice potential gain to block another individual from receiving a superior reward", though care needs to be taken not to confuse envy with a general desire for fairness, as noted in Kirchsteiger (1994).[8] Leibbrandt and López-Pérez (2012) examine the motivation

[7]An interesting exception is Gershman (2014) which discusses the impact on growth that could be attributed to both sides of envy.

[8]Kirchsteiger (1994) argues that while fairness is often considered a plausible explanation for deviations from traditional theory in the lab (pertaining to behavior in the ultimatum game), envy provides

for punishments (including envy) inflicted by affected second and unaffected third parties in a set of games, and Blanco, Engelmann, and Normann (2011) investigate inequality aversion (from the model developed in Fehr and Schmidt, 1999) in four games, relating the propensity to block the rewards of others to envy. Although related to "money-burning" in a loose sense, the effect of rejecting expert advice is different: by ignoring good advice, individuals harm their own payoff without directly affecting the adviser's. More generally, while none of these papers within economics have a focus on the take-up of advice (good or otherwise), they share the common theme that envy is an important and powerful psychological concept that needs to be better understood.

There are also important practical issues on the measurement of envy: if we measure envy through behavior or outputs then issues of endogeneity are likely to hamper any attempt to understand the effects of the concept further. For this reason we deploy a dispositional envy scale (Smith, Parrott, Diener, Hoyle, and Kim, 1999) and are the first to do so in economics.[9]

There are also interesting links to work on status. First note that in our design we were clear to both the advisors in wave 1 of our data collection and the potential advisees in wave 2 that wave 1 remuneration was not known at the time decisions were made: advisors only knew they would receive at least \$0.50, so the higher-paid advisors did not know they were going to receive \$100. Therefore, advisees knew the quality of advice was independent from how much their adviser was paid. This is reinforced by the fact that they are told the accuracy of expert advice so issues like the effort or ability of the advisor should be irrelevant. Nevertheless, there is a possible route through which status could play a role. The payment of \$100 or \$0.50 to advisors was random. However, Ball, Eckel, Grossman, and Zame (2001) point out that even randomly assigning high status can have an effect. In their paper the effect is to boost the bargaining power in a buyer-seller market of anyone who is randomly allocated a gold star. If the random allocation of \$100 in our paper is seen as similar to a gold star then, following Ball, Eckel, Grossman, and Zame we might expect to see a greater take-up of advice from higher paid advisors simply because of status. We, however, saw the opposite: a lower take-up of advise from higher-paid advisors. This does not rule out a status effect, but suggests that the effects we document out-weigh any such effect in our context.

There is also a literature on stubbornness which has largely been developed within psychology and management with the main focus being on understanding and measuring the phenomenon, typically through surveys. Wilkins (2015) provides a summary and a scale which we modify and use as a general measure of stubbornness. We are the first to use such a scale within economics and to link it to the avoidance of good advice. More common within economics is the study of a related but different concept: the sunk cost fallacy. There are numerous papers that try to evaluate the extent of this fallacy, through which individuals fail to realize that resources (e.g., time, money, emotion, energy etc.) that have already been incurred (sunk) should be ignored when making a decision. We believe we are the first to offer a scale to measure susceptibility to

a good alternative explanation that is not subject to the criticism that "...people are not concerned about every deviation from a fair share. They are only concerned if this deviation is disadvantageous for themselves."

[9]Though it has been referenced within economics before, see for instance Falk (2017).

the sunk cost fallacy, for which we draw on the classic works of Thaler (1999) and Arkes and Blumer (1985).

Ignoring good advice could also be linked to the concept of motivated reasoning. As is made clear in a large and growing literature, individuals may choose to act in a seemingly irrational way in order to bolster their own self-image or self-confidence (for two recent surveys see Benabou and Tirole, 2016; Gino, Norton, and Weber, 2016). For instance, a participant in our experiment may wish to believe they are highly skilled and so should have performed well in the skill task. If accepting advice is an internal admission that they are less good at such tasks than they would like to believe, then motivated reasoning would dictate ignoring (even very good) advice. However, in order to minimize the role of (even quite flexible) beliefs about self from playing a role, we provided participants with a clear statement of how well they performed leaving little room for doubt. It might still be possible to ignore this, and in turn ignore advice as part of a sustained attempt to retain an inflated self-image, but this would be difficult in the face of direct and unambiguous evidence to the contrary. To make this clear, consider the model of memory recall in Benabou and Tirole (2006): individuals receive signals that indicate the true state of the world. These signals can be manipulated at a later date through (purposefully) faulty recall for some cost. In our experiment, individuals receive a correct statement of their performance and are asked to make a decision based on this information within a few seconds of receiving it. This minimizes the scope for individuals to manipulate the information they are given, or (in the language of Benabou and Tirole, 2006) makes the cost of such a manipulation extremely high.

Finally, it is worth noting the relationship between our own work and Fehr, Herz, and Wilkening (2013) who suggest that in some cases individuals might be willing to accept lower utility in order to retain a sense of authority. This result mirrors our own findings, albeit only for a substantial minority, in that some of our participants ignore good advice even though this entails reducing their expected payoff. In Fehr, Herz, and Wilkening (2013) the design makes it clear that authority is lost when individuals delegate decision-making. This is less apparent in our setting, though it is perhaps possible to interpret our design as one in which participants are being given the option of delegating decision-making to an expert, implied by accepting the expert's choice of action. Under this interpretation, our work not only provides support for Fehr, Herz, and Wilkening (2013), albeit in a different setting, but also offers behavioral explanations for this apparent desire to retain authority over decision-making.

3 A Model of Advice

This section serves three roles. Firstly, it provides a conceptualization of advice which leads to a more formal set of definitions of the terms used throughout the paper. Second, it maps the model's terms to the core elements of the experimental design. Finally, it derives the econometric specification used to analyze the experimental results.

3.1 Conceptualization

Suppose that there are many individuals denoted $i = 1, ..., n$ and that each individual engages in some costly activity which determines their probability, p_i, of receiving a positive payoff (with probability $1 - p_i$ they receive a payoff of zero). Now suppose that individual i has the opportunity to choose between p_i and p_j for some $j \neq i$. Given this, we define the terms "advisee", "adviser", "advice", "good advice", "bad advice" and "value of advice".

Definition (Advisee). *Individual i is an advisee if they have the option to replace p_i with p_j for some other individual $j \neq i$.*

Definition (Adviser). *Individual j is an adviser if there exists an individual $i \neq j$ who has the option to replace p_i with p_j.*

Definition (Advice). *If j is an adviser, p_j is advice.*

Definition (Good and bad advice). *For advisee i and adviser j, p_j is good advice when $p_j > p_i$ and bad advice when $p_j < p_i$.*

Definition (Value of advice). *For advisee i, the value of advice from adviser j is $p_j - p_i$.*

A rational individual, i, who is offered a choice between p_i and p_j chooses p_i if and only if $p_i \geq p_j$. Therefore, when advice is good, i.e., $p_j > p_i$, it should be accepted. Our conceptualization of advice therefore boils down to an advisee being offered a choice between two, two-outcome (winning a positive payoff or not) lotteries, where the probability in one lottery has been generated by the advisee, and the probability in the other, by the adviser. Where advice is good, the lotteries are strictly-dominance ordered: the lottery generated by the adviser first-order stochastic dominates the advisee's lottery.

Our abstracted definition of good advice subsumes multiple interpretations. For example, suppose advisee and adviser have mutually exclusive theories about some state of the world (laid out by the adviser in e.g., a meeting, newspaper or blog). If the adviser's theory is more likely to correctly identify the state, then the theory (or the choice it recommends) is good advice, and the advisee should adopt the adviser's theory (or recommendations from it). As a second example, suppose the adviser provides information additional to that held by the advisee. Because the additional information (at least weakly) leads to actions that raise the advisee's expected payoff, this information is good advice and should not be ignored by the advisee.

The abstraction of modelling advice as a probability distribution goes back to works such as Morris (1974): regardless of the context or specific domain of recommendation, accepting advice is akin to accepting a probability distribution over outcomes. In our experiment, these distributions are particularly simple. We give every advisee participant, i, binary decisions over whether to accept advice (p_i or p_j for $i \neq j$) which determines their probability of winning a positive payoff or not (winning a bonus payment or not). Naturally, this simplicity excludes several aspects of reality. Firstly, depending on the domain of the advice, it is often possible to partially accept advice or to accept it in some dimensions rather than all. However, the idea that advice can be simplified down to a binary decision (yes or no, accept or ignore, stick or switch) is a widely used convention; as Calvert (1985) puts it (p. 534): "This feature represents the basic nature of advice, a distillation of complex reality into a simple recommendation." Second, an advisee would usually not be presented with probabilities directly, but rather with a set of recommended actions from which they are expected to infer the probabilities in order to decide whether to accept the advice. However, the focus of this experiment is not to study how people translate recommendations into probabilities. Rather, we want to study how people treat advice when they know it is good, so unlike in reality, we let advisees know both their own, and the adviser's probability of being correct. The fact that we run an experiment allows us to do this in a direct and controlled manner: advisees are told p_i and p_j, so whenever $p_j > p_i$, the advisee knows that the advice is good.

3.2 Mapping

We now map the model onto our experimental design. Our experiment consists of two waves. In the preliminary wave, we ask a small number of participants to complete tasks. Based on their performance, for each task we select some participants to act as advisers. In the main wave, all participants are advisees. There are two tasks. In the "luck task", participants guess whether a coin landed "heads" or "tails" for each of the ten times it was tossed. In the "skill task", participants select an answer to ten non-verbal reasoning IQ questions from eight alternatives.

If advisee i obtains $4/10$ in the coins task, then i's probability of winning the bonus using their own answers is $p_i = 0.4$. However, advisees also have the option to use the answers of an adviser. Suppose the adviser j obtained $7/10$, then $p_j = 0.7$. Next, the advisee, i, is faced with a choice between the two probabilities of winning a bonus: $p_i = 0.4$ or $p_j = 0.7$. Therefore, absent any "non-rational" forces, when participant i is faced with the decision between p_i and p_j, i chooses p_i if and only if $p_i \geq p_j$. Where this is how i acts, we say that i is rational, or, that i has taken a rational action.

3.3 Econometric specification

Our interest is in the possibility that individuals may depart from rationality by ignoring good advice in systematic ways. In other words, where $p_i - p_j > 0$ i.e., the value of advice is positive, and advisee i ignores advice. To study this, we allow for a general effect of $p_i - p_j$ as well as for additional factors which may determine the probability that individuals follow good advice. We denote these other factors as variables $x_k, k = 1, \ldots, K$, and allow for an error term, u. In our analysis, these K variables of interest include our between-subject treatments, psychometric measures such as envy, susceptibility to the sunk-cost fallacy, stubbornness, interactions of these variables, and participant demographics.

We now derive a probit specification for estimation from a latent variable model in a standard way. Firstly, fix a common adviser with advice $\bar{p}$, and restrict attention to those individuals for whom this advice is good i.e., i such that $\bar{p} > p_i$.[10] For any such advisee, i, we suppose that i's utility function is such that the difference in i's expected utility from accepting and ignoring

[10]In the experiment, we set $\bar{p} = 0.7$ in the luck task and $\bar{p} = 0.9$ in the skill task.

advice takes the following form:

$$E\left[U_i\left(\text{accept}\right)\right]-E\left[U_i\left(\text{ignore}\right)\right]=\sum_v \delta_v \mathbb{I}_v^{\bar{p}-p_i}+\sum_{k=1}^K \beta_k x_{i,k}+u_i,$$

where $\mathbb{I}_v^{\bar{p}-p_i}$ is an indicator variable, equal to one if the value of advice for i, $\bar{p}-p_i$, is v, and equal to zero otherwise. Assuming i chooses $y_i \in \{\text{accept}, \text{ignore}\}$ to maximize expected utility implies:

$$y_i=\text{accept} \iff \sum_v \delta_v \mathbb{I}_v^{\bar{p}-p_i}+\sum_{k=1}^K \beta_k x_{i,k}+u_i \geq 0.$$

We assume that $u_i \sim \text{iid}N(0,1)$[11] and are therefore left with the following probit specification for estimation:

$$pr(y_i=\text{accept})=\Phi\left(\sum_v \delta_v \mathbb{I}_v^{\bar{p}-p_i}+\sum_{k=1}^K \beta_k x_{i,k}\right).$$

Overall, our empirical strategy is to first establish whether good advice is indeed ignored and then, if so, to assess the importance of our x_k variables on the decision to accept or ignore good advice, using this probit specification where necessary.

4 Experimental Design

The participants in our study were from the Amazon Mechanical Turk (MTurk) online pool of subjects. MTurk's participant population has been shown to have the advantages of being more demographically diverse, and producing data of a comparable quality to more traditional participant methods (Chandler, Mueller, and Paolacci, 2014; Paolacci and Chandler, 2014). This has been shown through many studies replicating classic experiments in various domains including cognitive psychology (e.g., Goodman, Cryder, and Cheema, 2013; Paolacci, Chandler, and Ipeirotis, 2010) and economics (e.g., Horton, Rand, and Zeckhauser, 2011). The software used to perform the experiment was Qualtrics.

The experiment was registered in advance in the AEA RCT Registry (Ronayne and Sgroi, 2017). There, we detailed our intention to study the following: whether (good) advice is followed; the effect of luck vs. skill-based tasks; the effect of two remuneration treatments; subject envy, susceptibility to the sunk cost fallacy, and stubbornness. All of these measures are presented and discussed in this paper, and, there are no measures we included in the pre-registration that are not presented or discussed in the paper (see the section Primary Outcomes under Experimental Details of Ronayne and Sgroi, 2017). Here we provide a summary of the overall experimental design before going into detail on each part.

Our design involved two waves of data collection. In the first wave, 75 subjects undertook two incentivized tasks: the first was to guess "heads" or "tails" in a series of coin flips, the

[11]The choice of $\sigma=1$ is without loss of generality: $\sigma \neq 1$ could also be assumed but as σ cannot be identified, convention is to set $\sigma=1$.

second was a short Ravens visual IQ test. Some of the subjects who performed well in this first wave acted as the advisers who featured in the second wave. In the second wave, 1,503 subjects undertook the same two tasks. The addition over wave 1 is that subjects in wave 2 had the option to switch their own answers for that of one of the advisers selected from wave 1. For each task, they were told both their score and the score of the adviser and then offered the choice to submit their own answers or the answers of the adviser. For ten randomly chosen subjects, a single problem was chosen from their submitted answers and a bonus payment made if the correct answer was provided. These choices were followed by demographics and three sets of questions designed to measure envy, stubbornness and the sunk cost fallacy. The subsections below provide more detail.

The final subsection below discusses some key features of our design, but here we point out those most salient. One central feature was to make it clear to participants when advice was good or bad. In order to achieve this, we fixed the quality of advice and made this quality known to participants. In the experiment, advisees had the choice between two known probabilities of success (winning a bonus): one generated by their answers which applied if they ignored the advice and one generated by the adviser's answers which applied if they accepted the advice.[12] A second feature is that, in line with the model presented in Section 3, our experiment is simplified and stark compared to reality, with much of the richness and complexity of real-world decision-making stripped away. This feature is a key part of the design. In the context of real-world decision-making there are many reasons why advice might be ignored and in Section 2 we summarized literature which examines many of these other important factors. Our design allows us to control for many of these, including concerns about the potential gains of the advice since participants are told both the true quality of the advice, and the probability that they are correct absent advice. In the real world there is likely doubt surrounding either or both of these factors so that corresponding field data would suffer from identification issues. In contrast, our experimental design identifies these variables.

4.1 Wave 1: Choosing Advisers

The 75 subjects in this wave were paid $2.00 for completing the experiment which took an average of 7 minutes 34 seconds to complete, corresponding to an hourly wage of $15.86. Participants were first asked to guess the outcome of a series of ten coin flips, one at a time. They were told that doing "especially well" in this task would result in a bonus payment of at least $0.50. They undertook two practice questions (flips) to acclimatize to the software and were given feedback on their performance in the practice questions prior to starting the main questions. They were then asked to undertake a set of ten Ravens visual IQ questions. Again, they were told that doing especially well in this task would result in a bonus payment of at least $0.50 and were first tasked with a practice question and received feedback on their performance before undertaking the full test. Before receiving feedback on their scores, they were asked to complete a set of questions on the difficulty of the tasks followed by demographic questions.

[12]In the instructions we use the wording "Keep my answers" or "Use the other worker's answers". For details, see the transcript included at the end of this submission.

This wave was specifically designed to provide advisers for the main, second wave. The performance of subjects in wave 1 was recorded and advisers were chosen to be those with scores that placed them towards the top of the distribution for each task: the adviser for the luck task (coin-tosses) was a participant with a score of 7/10 and for the skill task (Raven's visual IQ), a participant with 9/10. Crucially, their scores were chosen so that they could be beaten by only a small minority of wave 2 subjects. This meant that we would have some data on how subjects with higher scores than the advisers behaved.[13] Another important feature was the wording of the bonus payment instructions: we carefully informed wave 1 subjects that they would receive "at least \$0.50" which gave us freedom to allocate different bonus payments and left performance incentives identical across subjects who ended up receiving different bonus payments.

4.2 Wave 2: Advice

We recruited a further 1,503 MTurk participants to take part in wave 2.[14] Participants in this wave were paid \$2.00 for completing the experiment which took an average of 12 minutes 11 seconds to complete, corresponding to an hourly wage of \$9.85. Wave 2 began identically to wave 1. Once again, subjects guessed the outcome of a series of ten coin flips, one at a time. They were told that a good performance in this task would result in a bonus payment of at least \$0.50, and undertook two practice questions to acclimatize to the software (and were once again given feedback on their performance in the practice questions prior to facing the main task). Again, they were next asked to undertake a set of ten Ravens visual IQ questions and told that if they did especially well in this task it would result in a bonus payment of at least \$0.50. Once again they were first tasked with a practice question and received feedback on their performance before undertaking the main test. The next part of the experiment marked the first difference between wave 1 and 2: subjects in wave 2 were now informed about the existence of the advisers from wave 1 as follows:

> "[...] we here describe the experiences of two other workers who, some time ago, completed the exact same tasks you have just tried for the same \$2.00 HIT reward. [...] These two workers both saw the same instructions you did. This means they were both told that they could receive at least \$0.50 for doing especially well in a task. In addition, they each did especially well in one of the tasks: one scored 7/10 in the coin task, and got a bonus of X; the other scored 9/10 in the logic puzzles task, and got a bonus of X."

The X in the text they were given was varied by treatment. We set $X = \$0.50$ for all participants except for those allocated to the "high-remuneration" treatment where we set $X = \$100.00$. Note from the wording that it was made clear to subjects in wave 2 that the earlier workers who did well also undertook the same tasks and with the same instructions.

[13] We ran pilots using 501 subjects in total to give us prior information on the likely distribution of scores for the tasks in waves 1 and 2, and to test our other measures. More detail on the pilot studies are available on request.

[14] Participants from wave 1 were excluded from participating in wave 2.

Next was the crucial "accept" or "ignore" decision. There was a careful explanation of how decisions would translate into possible bonus payments. Subjects were informed that ten participants would be chosen at random and that if they were chosen, one of their answers (from either task) would be selected, and if correct, they would win a bonus. Before proceeding, they were asked to either keep their answers or use the other worker's answers where the "other worker" was the adviser just described on the previous page.[15] Participants were asked this for each of the two tasks i.e., they could independently choose to keep their own answers or use the adviser's in each task. On the same page, we showed the subjects their own score and reminded them of the score of the adviser, and the bonus payment the adviser received. We also explained the probabilities of winning the bonus payments conditional on being selected, both in the case where they chose to keep their own answers and in the case where they chose to use the other participant's answers. Additionally, for those in the second, "per-follower" remuneration treatment, participants were informed that for every wave 2 subject who accepted the answers of the wave 1 participant's answers, the wave 1 subject's bonus would rise by \$0.25.

The wave 2 subjects across all three treatments then answered questions concerning demographics before moving on to the final page of the experiment.

4.3 Wave 2: Envy, Stubbornness and the Sunk Cost Fallacy

Wave 2 concluded with three sort questionnaires designed to generate three behavioral measures: dispositional envy, susceptibility to the sunk-cost fallacy, and stubbornness.

The first set of questions was the dispositional envy scale (DES) of Smith, Parrott, Diener, Hoyle, and Kim (1999), which yielded a measure with a reasonable degree of variation.[16] The subjects were asked to evaluate the extent to which they agreed with eight statements following a simple 5-point Likert scale (strongly disagree, moderately disagree, neither agree nor disagree, moderately agree, strongly agree). Some example statements are: "I feel envy every day" or "The success of my neighbors does not make me resent them". The scale was generated for each participant by adding the scores from each question where responses indicating a higher level of dispositional envy received a higher score. With eight questions generating a score of between 1 and 5, the total score for each individual was between 8 and 40 with a higher score representing a higher level of dispositional envy.

We restrict ourselves to the study of situations when potential advisees have a prior opinion which forms their initial probability of success in the decision-making task. In our experiment this prior opinion is formed through time and effort spent in the tasks which suggests there may be a role for the "sunk cost fallacy". Indeed, earlier work has documented the widespread role of the sunk cost fallacy which leads individuals to undertake utility-reducing actions because they have paid in advance with resources such as time, money, emotion or effort. This trait may also affect the uptake of advice if individuals have already sunk time and effort to form beliefs which lead them in one direction even if those beliefs are likely to generate lower payoffs than simply following expert advice. In this paper we build the first stubbornness scale designed to

[15] Note that we refrained from using words like "adviser" or "expert" in the text.

[16] A histogram of the data generated by the DES is given in Figure A1 in the Appendix.

measure susceptibility to the sunk cost fallacy, inspired by the work of Thaler (1999) and Arkes and Blumer (1985). Our scale is based on answers to five scenarios, for example:

> "Imagine that you have spent $20 on a ticket to a concert. The day of the concert comes and unfortunately it is snowing heavily, and you feel tired after a tough day. You know you would not have decided to go to the concert if you hadn't already bought the ticket, but you also know that you cannot get a refund. On balance you decide to go to the concert."

Subjects were asked to note their agreement on a 5-point Likert scale. Summing across the five questions created a scale with a range from 5 (minimum susceptibility) to 25 (maximum susceptibility).[17] Arkes and Blumer (1985) note that individuals commit the sunk cost fallacy when they continue a behavior or endeavor as a result of previously invested resources (e.g., time, money, emotion or effort) across a variety of dimensions: we try to capture some of these dimensions in our set of five scenarios, with one of our scenarios based on Thaler (1999). To our knowledge we are the first to form a scale to measure susceptibility to the sunk-cost fallacy within economics, though the general method follows the same principle as the DES or other similar scales in psychology e.g., and perhaps most famous of all, the Big Five Inventory used to measure various personality traits (see John and Srivastava, 1999).

Finally, the subjects faced a set of statements designed to generate a stubbornness score to provide an alternative to our measure of susceptibility to the sunk cost fallacy. Again, participants were asked to indicate agreement using the same 5-point Likert scale ranging from strongly agree to strongly disagree. Five questions were asked yielding an overall score between 5 and 25. An example statement is "I do something I want to do even if no one else wants to do it." Again, to the best of our knowledge we are the first to use such a scale in economics, but tests of this type are common in the management literature e.g., Wilkins (2015), which provided the five questions used in our scale. For each scale, questions were in some cases "inverted" so that agreement indicated a lack of envy, susceptibility to the sunk-cost fallacy or stubbornness and in other cases agreement indicated the opposite. This was done to prompt subjects to read the questions carefully and to discourage them from answering all questions identically.

4.4 Features of the Design

Here we examine some special features of the design which may not be apparent at first glance, but which highlight the importance of the level of control we have through running an experiment.

First, let us turn to the use of two different tasks. In reality, people's perception concerning whether others have achieved status or success through luck or skill, is often confused (see Frank, 2016). In order to identify luck and skill in our experiment, it was important for us to select a task that could not be interpreted as skill-based. The task of guessing the results of coin-tosses meets that criterion. This allows us to not only consider the role of task-type, but also to make

[17] As is standard, the sum is used for analysis because responses to each of the five scenarios are expected to be highly collinear (they are all designed to get at the same underlying phenomenon).

within-subject comparisons without fear of learning between tasks. The tasks themselves were chosen as they reflect differing levels of luck and skill. While the coin task is entirely luck-based, the Ravens visual IQ test is largely a skill-based task (albeit with luck playing some role because participants may select the correct answer by chance). Each individual in our experiment decides whether to accept advice separately in both the luck task and the skill task allowing us to test whether perceptions around luck and skill really matter.

A second important feature is the ability to isolate the roles of adviser remuneration, envy and stubbornness. Surveys such as the Edelman Trust Barometer and our Brexit survey (reported in Section 6) are indicative but do not establish a causal link even if any plausible factors are identified. The level of control we need is only likely to come about in an experimental setting in which we can control the environment and incentives. For example, in practice, higher quality experts are likely to be more highly remunerated so expert pay can be viewed as a signal of quality. This can act as a confound when attempting to study the relationship between the propensity to follow advice and adviser remuneration in a real-world setting, whereas our design fixes expert quality independently of adviser remuneration. In each condition, different remuneration levels (bonuses) were awarded, but all participants had the same information throughout the experiment so remuneration had no effect on performance. This was made clear to wave 2 participants, meaning they knew adviser performance was not driven by any extra effort that may have arisen from the higher incentives. In turn, this means that any acceptance of their advice was not in order to reward the advisers for higher effort, enabling us to isolate the effect of remuneration per se.

Our design also includes the endogenous production of advice, which brings several advantages. If advice was disseminated by the experimenter we would have to contend with possible reciprocity from participants (the so-called "demand effect", see e.g., Zizzo, 2010). If participants were selected as advisers by the experimenter there might also be issues of fairness and an additional channel for envy. Recall also the finding from the Edelman survey that people are about as likely to listen to advice from people like themselves as from academic or technical experts. By having advisers endogenously emerge from a prior wave of the experiment and drawn from the same MTurk pool of workers, we provide a measure of control against this bias.

The marginal financial incentives for participants to take good advice varies by the value of the advice, which will differ across participants, i.e., some perform well enough themselves to gain little or nothing from advice, while others perform poorly and would benefit more from advice. The incentives provided were sufficient in the sense that the majority (75%) accepted good advice, and the value of advice was found to be a significant determinant of acceptance (detailed results follow in Section 5). Notwithstanding this point, the relatively low marginal incentives reflect many of the applications we have in mind, e.g., situations where a voter is highly unlikely to be pivotal, but might feel some other motivation to get a decision correct.

5 Results

Wave 2 consisted of a total of 1,503 participants.[18] Each of these participants made choices regarding both the luck and the skill task. In total, this makes 3,006 decisions of whether to follow advice or not. Table 1 displays a breakdown of these decisions. In a total of 2,757 of the decisions, advice was good and the rational action was for the participant to accept the advice (accepting would have strictly increased expected payoff). In 73 cases, the participant achieved a strictly higher score than the adviser so the rational decision was to ignore the advice. In the remaining 176 decisions, the participant's score tied with the adviser's making either decision rational.

Table 1: Rational decisions

	Observations		
Rational decision	Luck	Skill	Total
Accept	1,322	1,435	2,757
Ignore	52	21	73
Indifferent	129	47	176
Total	1,503	1,503	3,006

Our first result is that good advice is frequently ignored. Specifically, we show that the rate at which good advice is accepted is significantly lower than the frequency implied instead by random error. To do so, we compare the proportion of participants making the rational decision to ignore the advice against the proportion making the rational decision to accept the advice. The results are presented in Table 2. Of the decisions where it was rational to ignore the advice almost all took the rational action; 96.2% and 100.0% in the luck and skill tasks, respectively. However, of the decisions where it was rational to accept the advice, 71.7% and 78.9% took the rational action in the respective tasks; a total rate of 75.4%. The difference in the proportion of rational decisions between the two groups is significantly different from zero within both the luck and skill tasks ($P < 0.001$ and $P = 0.018$, respectively): good advice is frequently ignored.

To offer some context, we note that in experiments of choices over lotteries, the rate at which participants violate stochastic dominance varies depending on the complexity of the lotteries offered. However, when they are in their simplest forms, violation rates have been documented to be low. For example, Birnbaum (1999) studied violations of stochastic dominance in online samples, where in perhaps the simplest binary choice offered, 6% of participants chose the dominated option.[19,20] We expected this level of error to constitute an upper bound on the

[18]Demographic information on these participants is shown in Table A1 in the Appendix. In wave 1, data from 75 participants was collected from which the advisers were chosen. Results from this wave are available upon request.

[19]This choice, offered to participants was between [$4 with 0.5, $96 with 0.3, $100 with 0.2] and [$4 with 0.5, $12 with 0.3, $100 with 0.2] (choice 3 of his Table 3) where 6% chose the latter.

[20]In Charness, Oprea, and Yuksel (2018), subjects chose between stochastic information sources that are framed as advisers. Although not as straightforward as choices between simple lotteries, their simplest control featured two Blackwell-ranked sources where subjects chose the less-informative source 11% of the time (see their Figure 2).

error rate in our experiment for three reasons: our participants face an even simpler, binary choice, where each lottery has two outcomes rather than three; the outcomes we offered (dollar amounts) were constant across the two lotteries (and one of the outcomes was zero); and we offered a verbal explanation of the choices. Indeed, we found that bad advice was accepted in only $2/73 < 3\%$ of all decisions, where the stochastically-dominated option was chosen. Against such a marker, the rate of 25% at which participants ignored good advice stands out starkly.

Table 2: Good advice is ignored

	Advice			Pr(rational)
	accepted	ignored	n	
Luck				
Rational to accept	948	374	1,322	0.717
Rational to ignore	2	50	52	0.962
Difference in proportions				0.244
P-value				<0.001
Skill				
Rational to accept	1,130	305	1,435	0.789
Rational to ignore	0	21	21	1.000
Difference in proportions				0.211
P-value				0.018

Our experimental design allows for variation in the value of advice (as defined in Section 3). Table 3 below shows how the proportion of participants taking good advice varied with value. If all participants took good advice, or did so with some unconditional error, these slope coefficients would be zero. However, the correlation between the proportion taking good advice and value is positive for both tasks as indicated by the non-zero regression coefficients ($P < 0.001$ and $P = 0.010$ for luck and skill, respectively). The relationship between the value of advice and the decision to take good advice suggests that our participants traded off this value against other, non-rational forces.

We have established that participants did not always take the rational decision to accept good advice. We now investigate the effect of our treatments. Firstly, we use our within-subject

Table 3: Trading-off rationality

$y = \mathbb{I}\,(\text{took good advice})$	Luck	Skill
Value	0.078 (0.010)	0.013 (0.005)
Constant	0.512 (0.032)	0.722 (0.028)
Observations	1,322	1,435

OLS regressions. Robust standard errors are shown in parentheses.

treatment to test whether good advice is ignored more or less frequently when the adviser achieved their status through skill rather than luck. Each participant completed both a luck and a skill task as well as deciding whether to take advice in each. In order to look at the effect of adviser skill on the propensity to take good advice, we restrict attention to individuals for whom advice was good in both tasks (Table 2 shows in nearly all cases where advice was bad, advice was ignored). There were 1,260 such participants, corresponding to 2,520 decisions. Table 4 displays the decisions made by these participants. The data are split into three groups depending on participants' relative value of advice across tasks. Specifically, panels A, B and C show data from participants for whom good advice was, respectively: more valuable in the luck task, equally valuable, and more valuable in the skill task. If all decisions were rational, all participants would have accepted in both tasks, so that the only non-zero entry in the 2×2 grids of Table 4 would be the top-left cell, and, as we would expect from Table 2 this is where the majority of the observations lie. Positive counts in the bottom-right cell of each panel shows that some participants ignored advice in both tasks. But to look at the effect of task, we look at the counter-diagonal cells in each panel. These cells show the data of participants who accepted good advice in one task and ignored it in the other. Panels A and C show that these participants were more likely to accept in the task for which the advice was more valuable, as one may expect (conditional on only accepting good advice for one task). In panel B however, the advice in each task was equally valuable to participants. Hence, the task itself is isolated as the meaningful difference between the two decisions. We find that all such participants who ignored advice in one task, ignored it in the skill task.

Table 4: The effect of value and adviser skill

Participants' decisions				Pr(ignore; skill) - Pr(ignore; luck)	P-value from McNemar's test
A: for whom advice was good in both tasks, but more valuable in luck ($n = 168$)					
		skill			
		accepted	ignored		
luck	accepted	125	27		
	ignored	1	15	0.155	<0.001
B: for whom advice was good, and equally valuable, in both tasks ($n = 143$)					
		skill			
		accepted	ignored		
luck	accepted	115	12		
	ignored	0	16	0.084	<0.001*
C: for whom advice was good in both tasks, but more valuable in skill ($n = 949$)					
		skill			
		accepted	ignored		
luck	accepted	603	12		
	ignored	158	176	-0.154	<0.001

*The exact Binomial test is used when the counter-diagonal frequencies sum to less than 25.

Next, we use our between-subject treatment to investigate the effect of adviser remuneration on the propensity to take good advice. In the control condition, advisers received \$0.50 as a reward for being made advisers. In our first remuneration treatment, advisers instead received a high lump-sum amount of \$100.00 as a reward. In our second remuneration treatment, advisers received \$0.50 plus an additional \$0.25 "per-follower" i.e., for every advisee that accepted their advice. In all conditions, the remuneration received by advisers was told to advisees. The results for the two treatments are provided by the coefficients on the treatment dummies in Table 5, which provides the average marginal effects (AMEs) from probit regressions as per the specification derived in Section 3. We find that those assigned to the treatment where the adviser received a higher level of remuneration exhibited a lower propensity to follow good advice ($P = 0.048, 0.045, 0.034, 0.036$ in specifications (3), (4), (7) and (8), respectively). This treatment effect was approximately 5.5-6.0 percentage points. In contrast, we did not find a significant difference in the propensity to take good advice across the control and the "per-follower" treatment.

We now provide evidence that behavior in our experiment is driven by underlying psychological variables. In turn, we discuss the roles of envy, stubbornness and the sunk cost fallacy. It was not clear ex-ante whether or when envy would have a positive or negative effect on the propensity to take good advice. In general, and as discussed earlier, it has been said that there are two sides to envy (Russell, 1930; Van de Ven, Zeelenberg, and Pieters, 2009). It could be that through relative comparisons to the adviser, envy could encourage self-betterment and directly increase the proportion of individuals taking good advice. Alternatively, it may be that those more disposed to envy incur a negative emotional response to exposure to someone superior to themselves in some dimension, which here could result in good advice being ignored. Through the use the dispositional envy scale of Smith, Parrott, Diener, Hoyle, and Kim (1999) and our design, we investigate when the effect of envy is to drive a positive or negative response in the propensity to take good advice.

Stubbornness reflects a general unwillingness to budge from one's current position or opinion. We consider both a general measure of stubbornness (Wilkins, 2015) and a measure of susceptibility to the sunk-cost fallacy for which we constructed a novel scale. Susceptibility to the sunk-cost fallacy is especially relevant to advice-taking contexts because individuals have often invested resources (time, effort, money etc.) in forming their position. In our experiment, this was reflected by the time and effort participants put into the tasks. By pitting these two measures against each other in regressions, we are able to retrieve the relative importance of the resources sunk in forming a position over and above a general reluctance to leave one's position.[21]

We now examine the effect of our psychometric measures on the probability of accepting good advice. Table 5 reports that the AME of envy on the probability of accepting good advice is positive and significant in the domain of skill ($P = 0.030, 0.029, 0.024, 0.022$ in specifications (5)-(8) respectively) but not in the domain of luck ($P = 0.679, 0.641, 0.379, 0.540$ in (1)-(4) respectively). Interpreting the estimated AMEs in the domain of skill, a one-standard-deviation

[21] Histograms of our participants' scores on all three psychometric scales are provided in Figures A1-A3 in the Appendix.

increase in dispositional envy increases the probability of taking good advice by roughly 2.5 percentage points on average.

However, this positive effect of envy is averaged across the different remuneration treatments. We now show that the effect of envy on the propensity to take good advice changes significantly depending on whether or not the adviser was highly remunerated. In other words, we find a significant interaction effect between envy and whether a subject was in the high-remuneration treatment condition. Figure 1 shows the AMEs of a one standard deviation increase in envy on the propensity to take good advice in the hypothetical scenario where all participants are allocated to the control versus where all are allocated to the high-remuneration treatment. The estimates reveal that the positive overall effect of envy found in Table 5 can be meaningfully decomposed by whether the adviser was highly remunerated or not. When all are assumed to be in the control, where the adviser was remunerated with the same low amount that the advisee could achieve, envy had a large positive effect: 5.6 percentage points for a one standard deviation increase in envy ($P = 0.002$). However, when all participants are assumed to be in the treatment where the adviser had been highly remunerated, the effect of envy was significantly different ($P = 0.004$ from a contrast of AMEs test) with an estimate of -1.6 percentage points for a one standard deviation increase in envy.[22]

To describe the effect of envy more precisely, Figure 2 provides average adjusted predictions of the probability of accepting good advice in the two scenarios, over the range of participants' dispositional level of envy. Where participants are assumed to be in the control condition (left panel), there is a monotonic increasing relationship between envy and the propensity to take good advice (we knew the relationship was increasing on average from the AME of 5.6 in Figure 1). The effect of envy can be seen to be large here. For the least envious participants, we predict on average that they will follow good advice about 71% of the time. In contrast, we predict on average that the most envious participants will follow good advice about 93% of the time, an increase of more than 20 percentage points. On the other hand, where participants are assumed to be in the high-remuneration treatment (right panel), the positive relationship is lost. For less-envious participants, the differences in the predicted probabilities (across panels) are not significant. However, as we consider more envious participants, the predictions across conditions diverge, culminating in the most envious participants following advice 24% less often when the adviser was highly-remunerated.

The AME of our measure of susceptibility to the sunk-cost fallacy is significant and negative across all the specifications of Table 5. The AMEs reported suggest that a one-standard-deviation increase in susceptibility to the sunk-cost fallacy decreases the probability of taking good advice by between 2.5-3.8 percentage points on average. Figure 3 provides the average adjusted predictions of the probability of accepting good advice in the domains of luck and skill, over the range of participants' susceptibility to the sunk cost fallacy. The difference between the least and most susceptible participants' in the propensity to follow good advice was 21 percentage points (79% for the least, 58% for the most) and in the skill task, 16 percentage points

[22]Congruent with earlier results, the corresponding AME of envy for remuneration treatment 2 was not significantly different from the AME where all participants are assumed to be in the control condition.

Table 5: Determinants of following good advice

Average Marginal Effects	Luck				Skill			
$y = \mathbb{I}$(took good advice)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Envy	0.005	0.006	0.011	0.008	0.025	0.025	0.025	0.026
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)
Sunk-Cost Fallacy	-0.038	-0.035	-0.036	-0.034	-0.027	-0.025	-0.030	-0.026
	(0.012)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.011)
Stubbornness	-0.005	-0.007	-0.012	-0.007	-0.013	-0.013	-0.018	-0.014
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)
Remuneration Treatment 1			-0.060	-0.060			-0.056	-0.055
			(0.030)	(0.030)			(0.026)	(0.026)
Remuneration Treatment 2			0.015	-0.013			-0.003	-0.001
			(0.029)	(0.029)			(0.025)	(0.025)
Remuneration × envy			X	X			X	X
Psych. interactions		X	X	X		X	X	X
Value dummies	X	X	X	X	X	X	X	X
Participant demographics	X	X		X	X	X		X
Observations	1,317	1,317	1,317	1,317	1,427	1,427	1,427	1,427

Average marginal effects are shown with standard errors in parentheses following probit regressions. The estimates from the probit regression are included in Table A2 in the Appendix. All specifications pass various mis-specification tests of functional form and heteroskedasticity. The binary dependent variable is 1 (0) if the decision made was to accept (ignore) the good advice. “Value dummies” refers to the inclusion of a dummy for every level of value in the underlying probit regression (bar one, to serve as the reference category). “Psych. interactions” refers to the inclusion in the underlying probit regression of all possible interaction terms between the three psychometric variables (envy, susceptibility to the sunk-cost fallacy and stubbornness). Each participant decided whether to take the advice in each task, hence there were 1,503 decisions for each task. Of those in the skill (luck) task, 68 (181) were excluded because they out-performed the adviser i.e., advice was bad. Of the remaining participants, 8 (4) were excluded because they did not disclose their sex. Finally, there was only one respondent with a value of 7 in the luck task, who was dropped for the analysis.

(84% for the least, 68% for the most).

In contrast, we did not find any predictive power of the general measure of stubbornness. This provides support for the notion that people may be unwilling to take good advice not because they are stubborn per se but because they are sensitive to the fact they used resources in order to form their position on a matter.

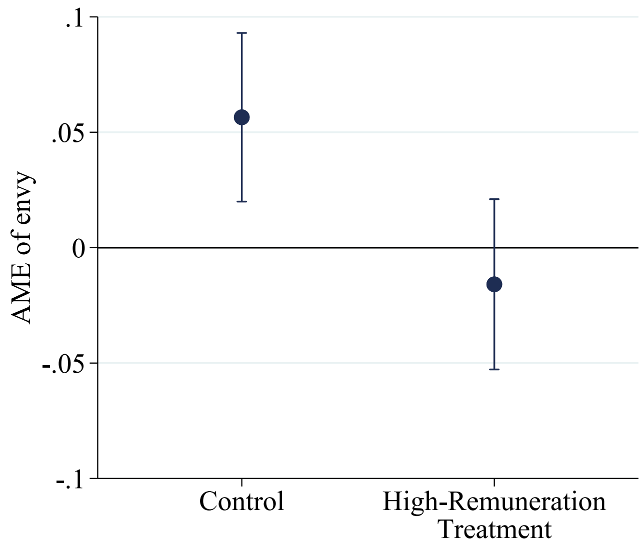
6 Survey Evidence

In this section, we offer survey evidence which is supportive of the results of our between-subject treatment regarding adviser remuneration. Firstly, Edelman (2017) reports the results from a survey covering over 33,000 people across 28 countries and suggests a link between resources and expert credibility. As well as a general distrust in experts, they find that 85% of people surveyed feel the "system" is biased towards the rich. However, this broad measure does not allow us to know whether that sentiment has affected the decision to accept advice, and if it did, whether the advice was rationally ignored because it was perceived to be bad, or ignored for other reasons. To make further progress in a real-world context, we put some questions to a separate panel of 3,096 voters in the UK's referendum on membership of the European Union.[23]

Respondents were asked how much they agreed with several statements on a 0-100 scale. The primary statement was "The advice of experts influenced my decision about how to vote". There, the average response was 38, consistent with the general notion that expert advice is frequently ignored. The next questions were designed to examine whether expert remuneration contributed to this low level of influence. Table 6 presents the statements respondents faced, along with output from a linear regression of the responses. The independent variable of interest is the response to the statement "Experts earn too much money". In our experiment, we were able to isolate the impact of adviser remuneration per se. However, in this field setting, we cannot control for what high expert pay could be perceived to signal. On the positive side, high pay could signal high quality. On the negative side, it could signal bias or corruption and therefore that the advice may be bad. The second and third variables are measures which attempt to capture these negative perceptions of why high expert pay would mean that their advice would be bad and hence rationally ignored. Controlling for these factors, we find a negative correlation between the reported influence of expert advice and agreement that expert pay is too high. In other words, the estimates suggest that there may be a negative effect of expert remuneration on the decision to follow advice even when some of the reasons why advice may be rationally ignored are taken into account. Such a finding is congruent with our experimental results that good advice is frequently ignored and that high adviser remuneration can reduce the propensity to accept good advice.

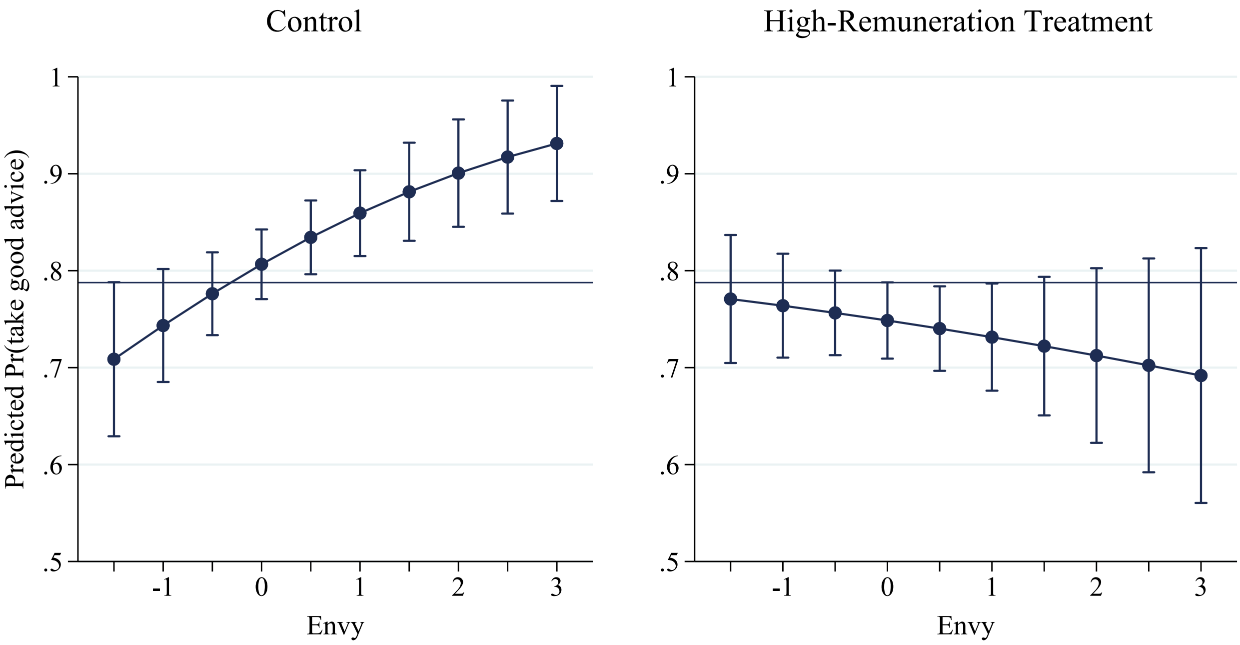
[23]Data was collected in November 2017. The questions were included in larger survey run by the Centre for Social Investigation at Nuffield College, University of Oxford.

Figure 1: The two sides of envy: average marginal effects



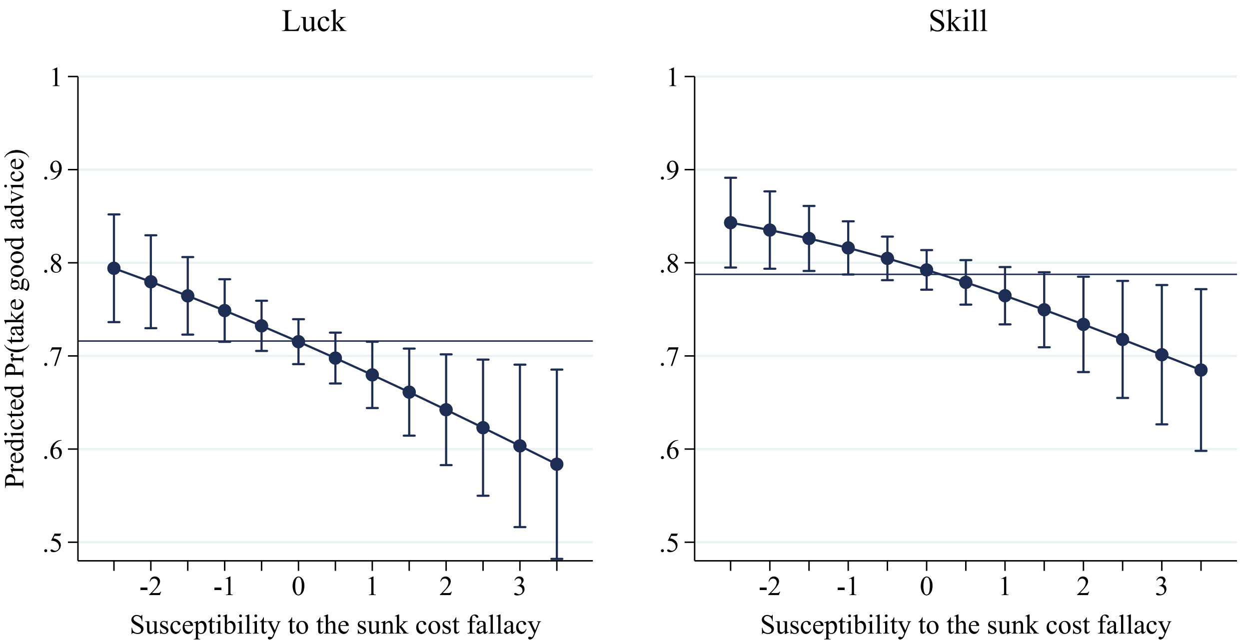
The average marginal effects shown were computed using the estimates from specification (8) of Table 5. 95% confidence intervals are shown. A contrast of AMEs test provides evidence ($P = 0.004$) that the AMEs shown are significantly different.

Figure 2: The two sides of envy: average adjusted predictions



The average adjusted predictions shown were computed using the estimates from specification (8) of Table 5 under the counter-factual assumptions that all participants have the dispositional envy level as shown on the x-axis, and were in the control (left panel); high-remuneration treatment (right panel). The horizontal reference line is the average predicted value of the probability of accepting good advice where the prediction for each participant is generated using their actual (observed) data. The lowest (highest) observed standardized level of envy in our sample was -1.60 (3.27). To provide a smoother illustration we allow envy to take 0.5 standard-deviation increments between -1.5 and 3.0. 95% confidence intervals are shown.

Figure 3: Susceptibility to the sunk cost fallacy leads to a lower propensity to take good advice: average adjusted predictions



The average adjusted predictions shown in the left and right panels were computed using the estimates from specifications (4) and (8) of Table 5 respectively under the counter-factual assumptions that all participants have the susceptibility level as shown on the x-axis. The horizontal reference lines are the average predicted value of the probability of accepting good advice where each participant's prediction is generated using their actual (observed) data in the domain of luck and skill respectively. The lowest (highest) observed standardized level of susceptibility in our sample was -2.57 (3.57). To provide a smoother illustration we allow envy to take 0.5 standard-deviation increments between -2.5 and 3.5. 95% confidence intervals are shown.

Table 6: Results from a Brexit survey

y = The advice of experts influenced my decision about how to vote.	
Experts earn too much money.	-0.107 (0.028)
I feel that expert advice was generally objective and unbiased.	0.436 (0.023)
Too often, experts have their own agenda.	-0.177 (0.034)
Observations	3,096

OLS regression. A constant was included but not reported. Robust standard errors are shown in parentheses. All four statements (including the dependent variable) were measured on a 0-100 scale, where 0 = "Completely disagree" and 100 = "Completely agree".

7 Concluding Remarks

We have offered an analysis of whether, when and why people ignore good advice. We provided an abstracted experimental setting in which it was clear that the rational action was to accept good advice. However, we found good advice was frequently ignored; in our experiment about 25% of the time. When good advice was more valuable, participants were more likely to accept it which suggests that participants traded off rationality with other forces.

What might those forces be? The literature documents many practical and context-specific factors that can play a role in determining the propensity to follow advice. We focus on *good* advice in an abstract setting and to our knowledge we are the only paper that does so. We investigated when good advice is ignored via novel treatments concerning adviser remuneration and task type. On the issue of why good advice might be ignored, we reveal that envy and the sunk cost fallacy play important roles in our data, while a general measure of stubbornness does not.

Before summarizing our findings in more detail we discuss other plausible behavioural forces which we did not examine. In our abstracted setting, the instructions of our experiment gave participants the opportunity to select either their own answers or the answers of an adviser. This may lead to the interpretation that participants were offered a chance to "cheat" rather than to take advice. We consider this interpretation from several angles, but find it unconvincing. First, we might think about an individual feeling guilty about breaking the "rules of the game" by using answers other than their own. This is not a direct concern for our experiment as the design explicitly allows people to consider adopting advice and makes it clear that doing so is not breaking any experimental rules. A second interpretation of "cheating" might be that a participant feels that using someone else's answers breaks some social norm. For instance, perhaps they feel they are "cheating the experimenter" by not sticking with their own answers. Interpreted this way, our experiment would report that 75% of subjects elected to "cheat", by adopting good advice. In contrast, estimates from a meta study shows that US participants leave on average over 80% of the gains from cheating on the table (see Figure A2 from Abeler, Nosenzo, and Raymond, 2018).[24] A third interpretation might refer to cheating the adviser in some way linked to the notion of social justice or fairness, through which a participant might feel guilty about the adviser not receiving a fair recompense for their good advice and so might opt to ignore the advice in consequence. However, if this was the case then we would expect to see participants follow advice more readily under our second between-subject treatment in which advisers received a bonus for every participant who followed their advice. In fact, we found behaviour under the second treatment to be statistically indistinguishable from the control in which advisers' compensation did not depend on the number of participants who followed their advice. More generally, we note that under alternate interpretations and other possible behavioural motivations for ignoring good advice, a key point is whether these are likely to confound or contradict the forces we studied, such as adviser remuneration, task type, envy or

[24] In the meta study of Abeler, Nosenzo, and Raymond (2018), participants roll a die and are then asked to report the number in order to receive payment. "Cheating" is therefore defined as lying in order to extract more money from the experimenter than is warranted by the true die roll.

the sunk cost fallacy. For instance, if concerns over fairness or cheating are the true explanation for our findings they would need to account for why advice is taken less often when advisers are paid more: but a more highly paid adviser would presumably reduce feelings of guilt and so enable advice to be taken more readily, which is the reverse of our finding.

Returning to the forces that we do document, we found that the fundamental human trait of envy played a major and varied role in determining whether good advice was followed. Our within-subject treatment revealed that good advice was followed less often when the adviser was more skillful than advisees, rather than luckier. Moreover, when the adviser was superior in skill, envy played a "positive" role: those with a higher dispositional level of envy were on average more likely to take good advice. In a between-subject treatment, we showed that good advice was followed less often when the adviser was highly remunerated, and there, a "negative" side of envy emerged: the positive association between envy and the propensity to take good advice was significantly lower, becoming insignificantly different from zero. We also showed that susceptibility to the sunk-cost fallacy was robustly negatively associated to whether good advice was accepted, suggesting an unwillingness to leave one's position because it was costly to form. We measured this susceptibility through a novel scale we constructed, based on works such as Thaler (1999) and Arkes and Blumer (1985). In contrast, a more general scale of stubbornness was not found to have predictive power.

Envy has long been viewed as a complex human trait. It is a latent characteristic sitting dormant, waiting to be activated. The Oxford English Dictionary defines envy as "the feeling of mortification and ill-will occasioned by the contemplation of superior advantages possessed by another". Important questions for social science are: what superior advantages possessed by another provoke envy and which decisions does envy then affect? To this end, our work suggests that comparisons of one's skill relative to another's, and one's remuneration relative to another's, can lead envy to be activated and to affect the decision of whether to accept good advice from that other individual. These forces may be particularly relevant in free-market economies where advisers tend to relatively highly skilled and remunerated. In our experiment, we also demonstrated both positive and negative sides of envy. When the adviser was superior in skill, more envious types were more likely to accept good advice, increasing their expected payoffs and reducing inequality between themselves and the adviser, which seems supportive of notions such as inequality aversion (e.g., Fehr and Schmidt, 1999). However, we showed that other payoff-irrelevant factors, e.g., adviser remuneration, can substantially reduce such an effect. Our treatments were instrumental in showing envy's multifaceted role, but naturally leave many interesting questions open. For example, our study focused on the behavioral determinants of following good advice conditional on being an advisee (the intensive margin) rather than the propensity to search for advice (the extensive margin).

In summary, our results show that individuals ignore advice even when the rational action is to take it, seemingly trading off rationality with other forces. The findings also suggest that the relative skill of advisers, adviser attributes such as remuneration, and underlying fundamental psychological traits of advisees, such as envy and the sunk cost fallacy, could be key to understanding when and why good advice is ignored.

References

Abeler, J., D. Nosenzo, and C. Raymond (2018): "Preferences for Truth-Telling," *forthcoming, Econometrica.*

Arkes, H. R., and C. Blumer (1985): "The psychology of sunk costs," *Organizational Behavior and Human Decision Processes*, 35, 124–140.

Ball, S., C. Eckel, P. J. Grossman, and W. Zame (2001): "Status in markets," *Quarterly Journal of Economics*, 116(1), 161–188.

Banerjee, D. (1990): "Envy," in *Economic Theory and Policy: Essays in Honor of Dipak Banerjee*, ed. by D. Ray, B. Dutta, S. Gangopadhyay, and D. Mookherjee. Oxford University Press.

Baumol, W. J. (1986): *Superfairness: Applications and Theory.* MIT Press.

Benabou, R., and J. Tirole (2006): "Belief in a just world and redistributive politics," *Quarterly Journal of Economics*, 121(2), 699–746.

——— (2016): "Mindful Economics: the production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 30(3), 141–164.

Birnbaum, M. H. (1999): "Testing critical properties of decision making on the Internet," *Psychological Science*, 10, 390–407.

Blanco, M., D. Engelmann, and H. T. Normann (2011): "A within-subject analysis of other-regarding preferences," *Games and Economic Behavior*, 72(2), 321–338.

Brennan, G. (1973): "Pareto desirable redistribution: the case of malice and envy," *Journal of Public Economics*, 2(2), 173–183.

Brenner, R. (1987): *Rivalry: In Business, Science, among Nations.* Cambridge University Press.

Cain, D. M., G. Loewenstein, and D. A. Moore (2005): "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest," *Journal of Legal Studies*, 35, 1–25.

Calvert, R. L. (1985): "The value of biased information: A rational choice model of political advice," *Journal of Politics*, 47(2), 530–555.

Chandler, J., P. Mueller, and G. Paolacci (2014): "Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers," *Behavior Research Methods*, 46(1), 112–130.

Charness, G., R. Oprea, and S. Yuksel (2018): "How do people choose between biased information sources? Evidence from a laboratory experiment," Working Paper.

Chaudhuri, A., A. Schotter, and B. Sopher (2009): "Talking ourselves to efficiency: coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice," *The Economic Journal*, 119(534), 91–122.

Cook, J., and S. Lewandowsky (2016): "Rational irrationality: Modeling climate change belief polarization using Bayesian networks," *Topics in Cognitive Science*, 8(1), 160–179.

Davis, M. S. (1968): "Variations in patients' compliance with doctors' advice: an empirical analysis of patterns of communication," *American Journal of Public Health and the Nations Health*, 58(2), 274–288.

DellaVigna, S., and D. Pope (2017): "What motivates effort? Evidence and expert forecasts," *Review of Economic Studies*, 85(2), 1029–1069.

Ding, T., and A. Schotter (2015): "Intergenerational advice and matching: An experimental study," Working paper.

Doberstein, C. (2016): "Whom do bureaucrats believe? A randomized controlled experiment testing perceptions of credibility of policy research," *Policy Studies Journal*, 45(2), 384–405.

——— (2017): "The Credibility Chasm in Policy Research from Academics, Think Tanks, and Advocacy Organizations," *Canadian Public Policy*, pp. 1–13.

Edelman, R. (2017): *2017: Trust in Crisis.* Edelman Trust Barometer: `http://www.edelman.com/trust2017`. Last accessed March 13, 2018.

Elster, J. (1987): "Envy in Economic Life," in *Strategy and Choice*, ed. by R. Zeckhauser. MIT Press.

Falk, A. (2017): "Status inequality, moral disengagement and violence," Working Paper.

Fehr, E., H. Herz, and T. Wilkening (2013): "The lure of authority: motivation and incentive effects of power," *American Economic Review*, 103(4), 1325–1359.

Fehr, E., and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114(3), 817–868.

Foley, D. K. (1967): "Resource allocation and the public sector.," *Yale Economic Essays*, 7, 45–98.

Frank, R. H. (2016): *Success and Luck: Good Fortune and the Myth of Meritocracy.* Princeton University Press.

Gershman, B. (2014): "The two sides of envy," *Journal of Economic Growth*, 19(4), 407–438.

Gino, F., and D. A. Moore (2007): "Effects of task difficulty on use of advice," *Journal of Behavioral Decision Making*, 20(1), 21–35.

Gino, F., M. I. Norton, and R. A. Weber (2016): "Motivated Bayesians: feeling moral while acting egoistically," *Journal of Economic Perspectives*, 30(3), 189–212.

Goodman, J. K., C. E. Cryder, and A. Cheema (2013): "Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples," *Journal of Behavioral Decision Making*, 26(3), 213–224.

Hilger, N. G. (2016): "Why Don't People Trust Experts?," *Journal of Law and Economics*, 59(2), 293–311.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011): "The online laboratory: Conducting experiments in a real labor market," *Experimental Economics*, 14(3), 399–425.

John, O. P., and S. Srivastava (1999): "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of Personality: Theory and Research*, 2(1999), 102–138.

Kardas, P., P. Lewek, and M. Matyjaszczyk (2013): "Determinants of patient adherence: a review of systematic reviews," *Frontiers in Pharmacology*, 4.

Kirchsteiger, G. (1994): "The role of envy in ultimatum games," *Journal of Economic Behavior & Organization*, 25(3), 373–389.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015): "How elastic are preferences for redistribution? Evidence from randomized survey experiments," *American Economic Review*, 105(4), 1478–1508.

Lachapelle, E., É. Montpetit, and J.-P. Gauvin (2014): "Public perceptions of expert credibility on policy issues: The role of expert framing and political worldviews," *Policy Studies Journal*, 42(4), 674–697.

Leibbrandt, A., and R. López-Pérez (2012): "An exploration of third and second party punishment in ten simple games," *Journal of Economic Behavior & Organization*, 84(3), 753–766.

Morris, P. A. (1974): "Decision analysis expert use," *Management Science*, 20(9), 1233–1241.

Mui, V.-L. (1995): "The economics of envy," *Journal of Economic Behavior & Organization*, 26(3), 311–336.

Mujcic, R., and A. J. Oswald (2018): "Is envy harmful to a society's psychological health and wellbeing? A longitudinal study of 18,000 adults," *Social Science & Medicine*, 198, 103–111.

Mullainathan, S., M. Noeth, and A. Schoar (2012): "The market for financial advice: An audit study," NBER Working Paper No. 17929.

Önkal, D., P. Goodwin, M. Thomson, S. Gönül, and A. Pollock (2009): "The relative influence of advice from human experts and statistical methods on forecast adjustments," *Journal of Behavioral Decision Making*, 22(4), 390–409.

Paolacci, G., and J. Chandler (2014): "Inside the Turk: Understanding Mechanical Turk as a Participant Pool," *Current Directions in Psychological Science*, 23(3), 184–188.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010): "Running experiments on Amazon Mechanical Turk," *Judgment and Decision Making*, 5(5), 411–419.

Parrott, W. G., and R. H. Smith (1993): "Distinguishing the experiences of envy and jealousy," *Journal of Personality and Social Psychology*, 64(6), 906—920.

Powdthavee, N., and Y. E. Riyanto (2015): "Would you pay for transparently useless advice? A test of boundaries of beliefs in the folly of predictions," *Review of Economics and Statistics*, 97(2), 257–272.

Ronayne, D., and D. Sgroi (2017): “When Good Advice is Ignored: The Role of Adviser Remuneration,” *AEA RCT Registry*, June 28, `https://doi.org/10.1257/rct.2022-3.0`.

Russell, B. (1930): *The Conquest of Happiness.* H. Liverwright.

Smith, R. H., W. G. Parrott, E. F. Diener, R. H. Hoyle, and S. H. Kim (1999): “Dispositional envy,” *Personality and Social Psychology Bulletin*, 25(8), 1007–1020.

Swol, L. M., and J. A. Sniezek (2005): “Factors affecting the acceptance of expert advice,” *British Journal of Social Psychology*, 44(3), 443–461.

Thaler, R. H. (1999): “Mental accounting matters,” *Journal of Behavioral Decision Making*, 12, 183–206.

Van de Ven, N., M. Zeelenberg, and R. Pieters (2009): “Leveling up and down: The experiences of benign and malicious envy.,” *Emotion*, 9(3), 419—-29.

Varian, H. R. (1974): “Equity, envy, and efficiency,” *Journal of Economic Theory*, 9(1), 63–91.

Weizsäcker, G. (2010): “Do we follow others when we should? A simple test of rational expectations,” *American Economic Review*, 100(5), 2340–2360.

Wilkins, M. M. (2015): “Signs that you’re being too stubborn.,” *Harvard Business Review*, May 21, `https://hbr.org/2015/05/signs-that-youre-being-too-stubborn`. Last accessed August 31, 2019.

Winkelmann, R. (2012): “Conspicuous consumption and satisfaction,” *Journal of Economic Psychology*, 33(1), 183–191.

Yaniv, I., and E. Kleinberger (2000): “Advice taking in decision making: Egocentric discounting and reputation formation,” *Organizational Behavior and Human Decision Processes*, 83(2), 260–281.

Zizzo, D. J. (2010): “Experimenter Demand Effects in Economic Experiments,” *Experimental Economics*, 13, 75–98.

Zizzo, D. J., and A. J. Oswald (2001): “Are People Willing to Pay to Reduce Others’ Incomes?,” *Annales d’Économie et de Statistique*, 63/64, 39–65.

Appendix

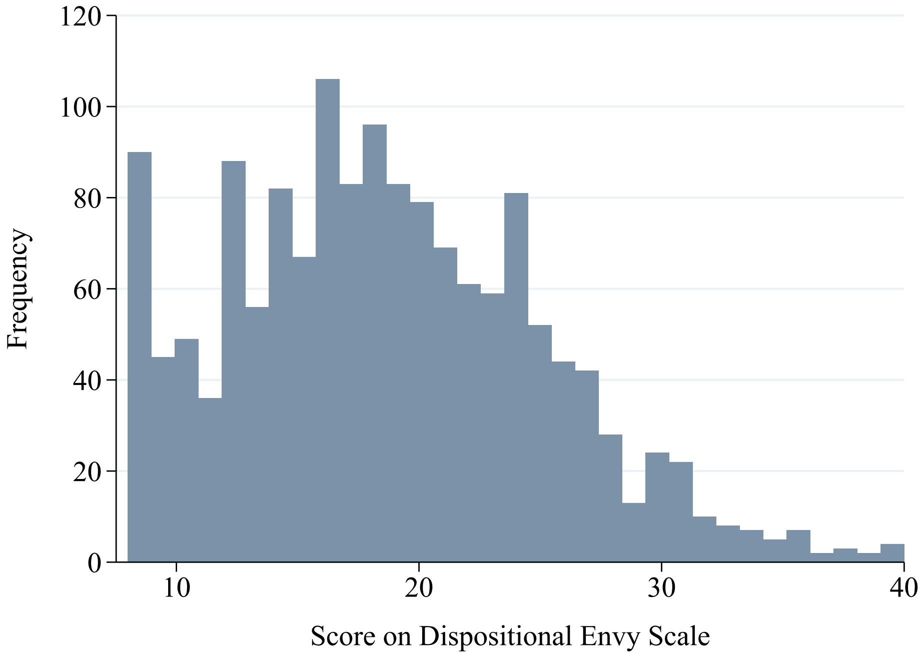
The Appendix provides demographic information on the participants of the main wave (Table A1); the distribution of scores obtained from participants' responses to three psychometric scales we used (Figures A1-A3); coefficient estimates of the probit coefficients underlying the regressions of Table 5 (Table A2); and a copy of the experiment's transcript (modified for readability and intended to be online only; not for publication).

Table A1: Participant Demographics from Main Wave

Characteristic	
Gender	
Male	701 (47)
Female	794 (53)
Other	4 (0)
Prefer not to say	4 (0)
Age, mean years [sd]	35.7 [11.7]
18-25	284 (19)
26-30	326 (22)
31-40	484 (32)
41-50	210 (14)
51+	199 (13)
Race	
White	1,105 (74)
Black or African American	125 (8)
Hispanic or Latino	96 (6)
American Indian or Alaska Native	11 (1)
Asian American	130 (9)
Native Hawaiian or Pacific Islander	6 (0)
Other	30 (2)
Income[a]	
0 – 9.999	77 (5)
10 – 19.999	158 (11)
20 – 29.999	197 (13)
30 – 39.999	215 (14)
40 – 49.999	175 (12)
50 – 59.999	167 (11)
60 – 69.999	127 (8)
70 – 79.999	113 (8)
80 – 89.999	54 (4)
90 – 99.999	53 (4)
100 – 124.999	85 (6)
125 – 149.999	38 (3)
150+	44 (3)
Education	
No schooling	1 (0)
Nursery, kindergarten, and elementary (grades 1-8)	1 (0)
High school (grades 9-12, no degree)	23 (2)
High school graduate (or equivalent)	160 (11)
Some college (1-4 years, no degree)	546 (36)
Bachelor's degree (BA, BS, AB, etc)	591 (39)
Master's degree (MA, MS, MENG, MSW, etc)	139 (9)
Professional school degree (MD, DDC, JD, etc)	26 (2)
Doctorate degree (PhD, EdD, etc)	16 (1)
Political Affiliation[b] [sd]	38.7 [29.6]
N	1,503

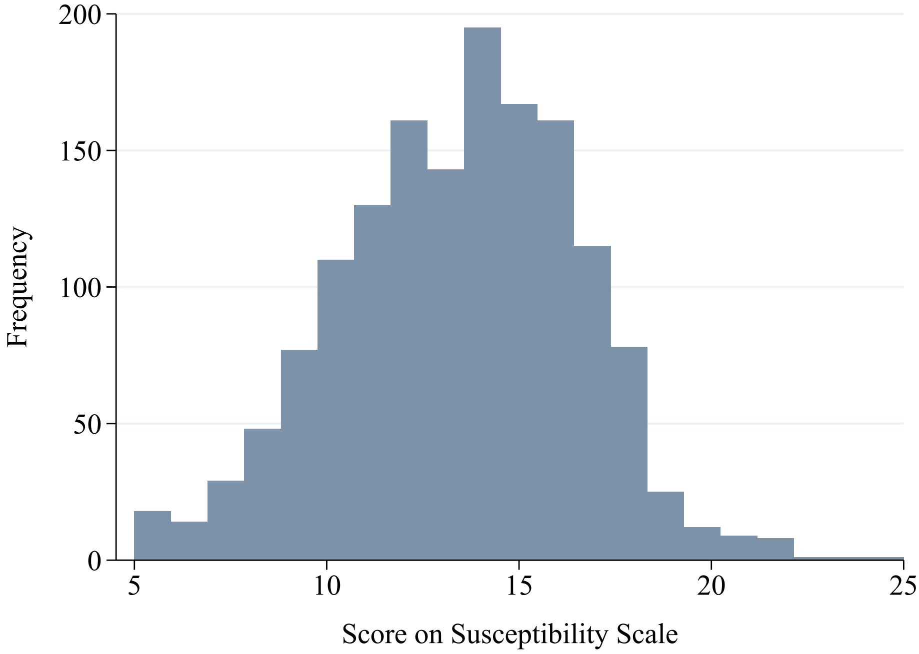
Frequencies; (% within characteristic); [standard deviation]
[a] Household annual pre-tax income in '000 USD
[b] 0 = "Entirely Liberal"; 100 = "Entirely Conservative"

Figure A1: Envy



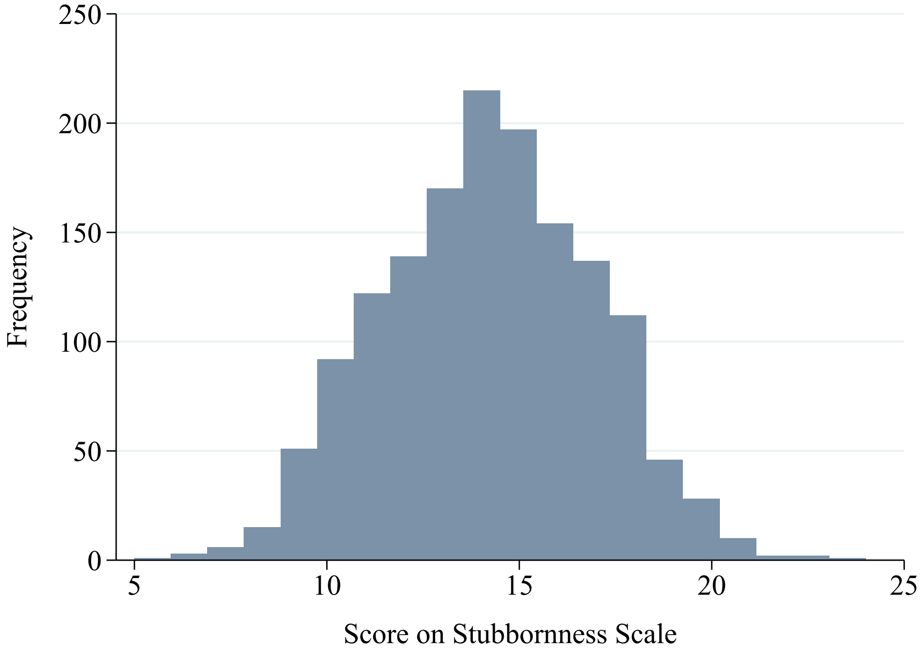
Participants' scores from the dispositional envy scale of Smith, Parrott, Diener, Hoyle, and Kim (1999) that comprises of 8 statements to which the subject must select how much they agree with them on a 5-point Likert scale from "Strongly disagree" to "Strongly agree". Before summing, the responses to each question are coded from 1-5 such that the higher the score, the higher the level of dispositional envy, $N = 1,503$.

Figure A2: Susceptibility to the Sunk Cost Fallacy



Participants' scores from our susceptibility scale that comprises of 5 statements to which the subject must select how much they agree with them on a 5-point Likert scale from "Strongly disagree" to "Strongly agree". Before summing, the responses to each question are coded from 1-5 such that the higher the score, the higher the level of susceptibility to the sunk cost fallacy, $N = 1,503$.

Figure A3: Stubbornness



Participants' scores from the responses to the stubbornness criteria listed inWilkins (2015) that comprises of 5 statements to which the subject must select how much they agree with them on a 5-point Likert scale from "Strongly disagree" to "Strongly agree". Before summing, the responses to each question are coded from 1-5 such that the higher the score, the higher the level of stubbornness, $N = 1,503$.

Table A2: Probit Coefficients Behind AMEs of Table 5

Coefficients	Luck				Skill			
$y = \mathbb{I}$ (took good advice)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Envy	0.017	0.019	0.055	0.047	0.089	0.078	0.199	0.269
	(0.041)	(0.042)	(0.069)	(0.071)	(0.041)	(0.042)	(0.072)	(0.073)
Sunk-Cost Fallacy (SCF)	-0.119	-0.108	-0.108	-0.105	-0.096	-0.091	-0.108	-0.096
	(0.039)	(0.041)	(0.041)	(0.042)	(0.038)	(0.040)	(0.040)	(0.041)
Stubbornness	-0.016	-0.021	-0.036	-0.022	-0.045	-0.044	-0.060	-0.048
	(0.042)	(0.042)	(0.041)	(0.042)	(0.041)	(0.041)	(0.042)	(0.041)
Envy × SCF		0.040	0.037	0.037		0.085	0.077	0.075
		(0.038)	(0.038)	(0.038)		(0.037)	(0.037)	(0.037)
Envy × Stubbornness		-0.029	-0.027	-0.029		-0.001	0.002	0.003
		(0.036)	(0.036)	(0.036)		(0.035)	(0.035)	(0.034)
SCF × Stubbornness		-0.003	-0.001	0.001		-0.105	-0.097	-0.103
		(0.040)	(0.039)	(0.040)		(0.037)	(0.036)	(0.037)
Envy × SCF × Stubbornness		-0.010	-0.009	-0.008		-0.024	0.024	-0.025
		(0.026)	(0.025)	(0.026)		(0.025)	(0.025)	(0.025)
Remuneration T1			-0.181	-0.184			-0.203	-0.202
			(0.092)	(0.092)			(0.093)	(0.093)
Remuneration T1 × Envy			-0.073	-0.072			-0.268	-0.269
			(0.090)	(0.091)			(0.092)	(0.092)
Remuneration T2			0.049	-0.044			-0.018	-0.012
			(0.093)	(0.094)			(0.095)	(0.096)
Remuneration T2 × Envy			0.018	0.010			-0.066	-0.072
			(0.096)	(0.096)			(0.098)	(0.097)
Value dummies	X	X	X	X	X	X	X	X
Participant demographics	X	X		X	X	X		X
Observations	1,317	1,317	1,317	1,317	1,427	1,427	1,427	1,427

Coefficient estimates from probit regressions are shown with robust standard errors in parentheses. All specifications pass various mis-specification tests of functional form and heteroskedasticity. The binary dependent variable is 1 (0) if the decision made was to accept (ignore) the good advice. "Value dummies" refers to the inclusion of a dummy for every level of value (bar one, to serve as the reference category). Each participant decided whether to take the advice in each task, hence there were 1,503 decisions for each task. Of those in the skill (luck) task, 68 (181) were excluded because they out-performed the adviser i.e., advice was bad. Of the remaining participants, 8 (4) were excluded because they did not disclose their sex. Finally, there was only one respondent with a value of 7 in the luck task, who was dropped for the analysis.

Transcript (modified for readability: to appear online only)

Participation Agreement

You have been invited to take part in a research study run by researchers at the University of Oxford. Please read the following statements carefully.

<u>Our Commitments and Privacy Policy</u>

- We never deceive participants. For example, if we inform you that another participant is making a choice on which you can then react, this is indeed the case.
- We keep our promises made to participants. For example, if we promise a certain payment, participants will indeed receive it.
- In the event that we are responsible for a mistake that is to the disadvantage of participants, we will inform and compensate the respective participants.
- We design, conduct and report our research in accordance with recognised scientific standards and ethical principles.

We adhere to the terms of our privacy policy as stated below.

- The data in the participants' database will only be used for the purpose of the study.
- There is no link between the personal data in the participants' database and the data collected during a study.
- The generated anonymous data will be used for analysis. The end product will be publicly available.
- Your participation in this study is purely voluntary, and you may withdraw your participation or your data at any time without any penalty to you.
- Please be aware that Amazon user information, connected to Mturk worker IDs, can be visible to the public, depending on the privacy settings of your Amazon.com account. See also https://www.mturk.com/mturk/privacynotice for further information on Amazon.com's privacy policies.

There are no known risks associated with your participation in this research beyond those of everyday life. If there is anything about the study or your participation that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact the Requester via MTurk. For questions about your rights as a research participant, you may contact The Center for Experimental Social Science, Nuffield College, University of Oxford, 3 George Street Mews, Oxford OX1 2AA, at cess@nuffield.ox.ac.uk.

○ I agree

Coins: Task & Bonus Instructions

You will face ten timed questions where you are simply asked to guess whether a fair two-sided coin landed with "Heads" or "Tails" facing up.

You have ten seconds to answer each question. If you are happy with your answer and wish to move on to the next question before the ten seconds are up, simply hit the blue ">>" button at the bottom of the screen. If the time runs out and you have selected an answer, that answer will be submitted and you will move on automatically to the next question. If the time runs out and you have not selected any answer, that question will be marked as incorrect and you will move on automatically to the next question. Once you have moved on from a question, you cannot go back to it.

If you do especially well in this task you will be awarded a bonus payment of at least $0.50. Bonuses will be paid after the required number of workers have completed the HIT. Those who do not win a bonus will not be notified.

Before you take the quiz of ten questions, you first have two practice questions. These will allow you to get a feel for the format and time limit. They do not count for the bonus payments. The first practice question will begin immediately on the next page.

○ I understand these instructions

Coin Toss (Practice x2)

One coin is tossed. Guess which side landed face up:

○ Heads

○ Tails

Coin Tosses Practice Feedback

You scored <<their score>> out of 2 in the practice.

First coin-toss
Your guess: <<their guess>>

The coin showed: Tails

Second coin-toss
Your guess: <<their guess>>
The coin showed: Heads

The real questions will **begin immediately** on the next page. Make sure you are ready.

Coin Tosses (x10)

One coin is tossed. Guess which side landed face up:

- ○ Heads
- ○ Tails

Logic Puzzles: Task & Bonus Instructions

You will face ten timed multiple choice questions about logic.

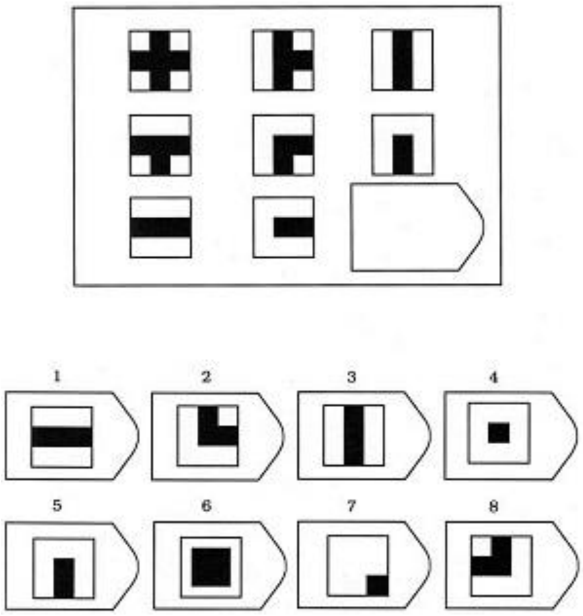
Each question shows a sequence of nine patterns with one missing. Your task is to select the missing pattern from the drop-down list. There is only one correct answer for each question. You have 30 seconds to answer each question. If you are happy with your answer and wish to move on to the next question before the 30 seconds are up, simply hit the blue ">>" button at the bottom of the screen. If the time runs out and you have selected an answer, that answer will be submitted and you will move on automatically to the next question. If the time runs out and you have not selected any answer, that question will be marked as incorrect and you will move on automatically to the next question. Once you have moved on from a question, you cannot go back to it.

If you do especially well in this task you will be awarded a bonus payment of at least $0.50. Bonuses will be paid after the required number of workers have completed the HIT. Those who do not win a bonus will not be notified.

Before you take the task of ten questions, you first have a practice question. This will allow you to get a feel for the format and time limit. It does not count for the bonus payments. The practice question will begin immediately on the next page.

○ I understand these instructions

Logic Puzzle Practice



(Dropdown menu of 1-8.)

Logic Puzzles Practice Feedback

You got the practice question correct!
[You got the practice question incorrect. Hopefully you will have better luck with the next questions.]

The real task will **begin immediately** on the next page. Make sure you are ready.

Logic Puzzles (x10)

<<Ten puzzles similar in nature to the practice were presented, each with 8 possible answers>>

Description of Other Worker

Before we reveal your scores...

... we here describe the experiences of two other workers who, some time ago, completed the exact same tasks you have just tried for the same $2.00 HIT reward. Please pay attention as we will be asking you some **comprehension questions** about them on the next screen.

These two workers both saw the same instructions you did. This means they were both told that they could receive <u>at least</u> $0.50 for doing especially well in a task. In addition, they each did especially well in one of the tasks: one scored 7/10 in the coin task, and got [If not in remuneration treatment 1: “a bonus of $0.50”] [If in remuneration treatment 1: “a large bonus of $100”]; the other scored 9/10 in the logic puzzles task, and [If not in remuneration treatment 1: “got a bonus of $0.50”] [If in remuneration treatment 1: “also got a large bonus of $100”].

- ◯ I am ready for comprehension questions on the text above

Comprehension Questions

What did the two workers who did the tasks some time ago know about the potential bonus payment for doing "especially well" in a task **before** they started it?

- ◯ It would be at least $0.50
- ◯ It would be exactly $0.50

What bonus payment did both the workers actually get for doing especially well?

- ◯ $0.50
- ◯ $100.00

A Chance for a Bonus

We are not going to ask you to repeat the tasks. On this page we explain how the answers you gave in each task translate into your chances of winning a bonus.

We will choose ten workers at random. If you are chosen, we will pick one question at random, and if you got it right, you will win a bonus. Let's look at your scores:

You got <<their coins score>>/10 in the coin task
You got <<their logic score>>/10 in the logic puzzles task

That means if you are chosen and we pick one of the coin questions, there is a <<their coins score>> in 10 chance of you winning the bonus. Similarly, if you are chosen and we pick one of the logic puzzle questions, there is a <<their logic score>> in 10 chance of you winning the bonus.

But before we go ahead, we would like to give you an opportunity to perhaps boost your odds of getting the bonus. Remember those two other workers we described earlier who did especially well? If you like, instead of us using your answers when we check if you have won the bonus we will look at their answers: that means your chances of getting the bonus would be 7 in 10 if we pick from the coin task or 9 in 10 from the logic puzzles task.

Regarding payment: The other worker got a bonus of $0.50. In your case, the bonus you might win is also $0.50. [If in remuneration treatment 1: “The other worker got a large bonus of $100.00. In your case however, the bonus you might win is $0.50.”] [If in remuneration treatment 2: “The other worker got a bonus of $0.50. In your case, the bonus you might win is also $0.50. Additionally, if you decide to use another worker's answers, they will get a further bonus of $0.25.”]

So, would you like us to use the answers you already gave, or the answers the other worker gave when we check whether you have won a bonus?

For the coin task:

- ○ Keep my answers
- ○ Use the other worker's answers

For the logic puzzles task:

- ○ Keep my answers
- ○ Use the other worker's answers

Comprehension Questions

What did the two workers who did the tasks some time ago know about the potential bonus payment for doing "especially well" in a task before they started it?

- ○ It would be at least $0.50
- ○ It would be exactly $100.00

What bonus payment did both the workers actually get for doing especially well?

- ○ $0.50
- ○ $100.00

Final Questions (page 1 of 2): Demography

What is your sex?

- ○ Male
- ○ Female
- ○ Other
- ○ Prefer not to say

What is your age?

What is your race?

- ○ White
- ○ Black or African American
- ○ Hispanic or Latino
- ○ American Indian or Alaska Native
- ○ Asian American
- ○ Native Hawaiian or Pacific Islander
- ○ Other

What is your household's annual income? (US dollars, before tax)

- 0-9,999
- 10,000 - 19,999
- 20,000 - 29,999
- 30,000 - 39,999
- 40,000 - 49,999
- 50,000 - 59,999
- 60,000 - 69,999
- 70,000 - 79,999
- 80,000 - 89,999
- 90,000 - 99,999
- 100,000 - 124,999
- 125,000 - 149,999
- 150,000 +

What is the highest grade of school you have completed, or the highest degree you have received?

- No schooling (or less than 1 year)
- Nursery, kindergarten, and elementary (grades 1-8)
- High school (grades 9-12, no degree)
- High school graduate (or equivalent)
- Some college (1-4 years, no degree)
- Bachelor’s degree (BA, BS, AB, etc)
- Master’s degree (MA, MS, MENG, MSW, etc)
- Professional school degree (MD, DDC, JD, etc)
- Doctorate degree (PhD, EdD, etc)

Generally speaking, which point on this scale best describes your political affiliation?
(A slider was presented with range [0,100] with “Entirely Liberal” over 0 and “Entirely Conservative” over 100.)

What is your Mturk ID? (please copy and paste it to avoid typos)

Final Questions (page 2 of 2): Personality

Please respond to the statements below using the scales provided: (each scale was a 5-point Likert scale with “Strongly disagree”, “Moderately disagree”, “Neither agree nor disagree”, “Moderately agree” and “Strongly agree”.)

(Dispositional Envy Scale)
I feel envy every day.
The bitter truth is that I generally feel inferior to others.
It doesn't frustrate me to see some people succeed easily.
Feelings of envy rarely torment me.
No matter what I do, envy always plagues me.
I am rarely troubled by feelings of inadequacy.
It somehow doesn’t seem fair that some people seem to have all the talent.
The success of my neighbors doesn't make me resent them.

(Susceptibility to the Sunk-Cost Fallacy Scale)
You have invested a good deal of your time into a project and it is failing. You have the option to start on something different that you now know is more likely to be successful but you know you cannot get the time back that you spent on the project so you decide to keep going with it.

You have an investment strategy that you have developed over several months. It is not working and you are losing money. There is no way for you to recover the lost effort put in to developing the strategy but you decide that it is better to start afresh anyway.

Imagine that you have spent $20 on a ticket to a concert. The day of the concert comes and unfortunately it is snowing heavily, and you feel tired after a tough day. You know you would not have decided to go to the concert if you hadn’t already bought the ticket, but you also know that you cannot get a refund. On balance you decide not to go to the concert.

You are staying in a hotel room, and you have just paid $6.95 to see a movie on pay TV. You find that you are bored 5 minutes into the movie and that the movie seems pretty bad. You decide that since you cannot get a refund you might as well continue watching the movie.

Your relationship with your partner is not going well. You have reasoned it out and you have realized that if you knew how it would go when you started the relationship you would not have gone through with it. You have the opportunity to break up but since you have been together for many months you decide to keep going.

(Stubbornness Scale)
I do something I want to do even if no one else wants to do it.
I never keep at an idea (or plan) when I know I am wrong.
When others present an idea, I tend to point out all the reasons it won’t work.
I agree to or commit half-heartedly to others’ requests, when I know all along that I’m going to do something entirely different.
I visibly feel anger, frustration, or impatience when others try to persuade me of something I don’t agree with.