CONFIDENTIAL: DO NOT DISTRIBUTE WITHOUT PERMISSION OF AUTHORS

Captains, Carbon, and Charity: A Field Experiment on the Impacts of Information, Targets, and Altruistic Incentives on Airline Pilot Behavior

Greer Gosnell, London School of Economics
John List, University of Chicago
Robert Metcalfe, University of Chicago
March 27, 2015

Abstract: Increasing fuel use in the airline industry has negative implications for both airlines' profitability and the environmental impact of the aviation industry, which is responsible for approximately 3.5% of total anthropogenic radiative forcing carbon emissions (projected to grow to 5% by 2050). While technology adoption and market-based mechanisms continue to appear on the aviation industry's agenda, little work has been done to understand the potential for emissions savings from changes in behavior of transport personnel. In this paper, we report a framed field experiment that is innovative with respect to both its methodology and its context. We provide personalized information, exogenous targets, and charitable incentives to airline captains to encourage the implementation of fuel-efficient behaviors. This study is the first to use field experimentation within the aviation industry and in fuel efficiency more generally. We find that exogenous targets are most effective in motivating captains to perform fuel-relevant behaviors—in some cases, increasing the incidence of these behaviors by more than 10%—while charitable incentives do not provide further motivation beyond these targets. These findings hold important theoretical implications for labor economics and provide practical solutions for the transport industry.

Keywords: aviation, fuel efficiency, altruism, information, targets, employee motivation, prosocial incentives, field experiment
JEL: D01, J3, Q5, R4

Acknowledgments: We thank the Templeton Foundation and the Science of Philanthropy Initiative at the University of Chicago for providing the generous funds to make this experiment possible. Further thanks to Emma Harvey, Claire Lambert, Paul Morris, Capt. David Kistruck, and the 'trusted pilots' within Virgin Atlantic for their essential roles in the implementation of this experiment. Finally, we thank OSyS for their cooperation in providing high-quality data in a timely manner. These parties are in no way responsible for the analyses and interpretations presented in this paper. Gosnell: g.gosnell@lse.ac.uk; List: jlist@uchicago.edu; Metcalfe: metcalfe@uchicago.edu.

I. INTRODUCTION

Excessive fuel use in the global aviation industry poses a severe threat to both airlines—for which fuel represents 33% of operating costs, on average (ATAG, 2014)—and to the global environment. The air transport sector currently accounts for 2-3% of global greenhouse gas emissions and is responsible for an estimated 3.5-5% of global radiative forcing due to the high altitudes at which emissions are generated and the formation of cirrus clouds from contrails (IPCC, 1999; Lee et al., 2009; Burkhardt & Kärcher, 2011). Technology adoption and market-based instruments continue to appear on the industry's agenda as primary means to reach its goals of carbon neutral growth by 2020 and halving emissions from 2005 levels by 2050 (see ICAO, 2013). However, almost no work has been done to understand the potential for cost and emissions savings from changes in the behavior of (air) transport personnel (Metcalfe & Dolan, 2012). While lack of such research is surprising given immense fuel costs, airlines are beginning to investigate a larger role for fuel savings from operations.

In the aviation industry, airline captains maintain a considerable amount of decision-making autonomy due to their high levels of experience and expertise (Smith, 2013).[1] In general, once trained, a captain will rarely come into contact with the airline's training and flight operations personnel, who are responsible for ensuring efficient aircraft operations. As in all classical principal-agent problems, this historical communication barrier between captains and fuel operations personnel has persisted due to highly asymmetric information with respect to both best practice and captains' 'effort' levels. Airlines have not traditionally observed captains' fuel-saving efforts and have been unable to monitor their individual behavior and performance as such. That is, captains maintain private information regarding the effort they expend. This information asymmetry leads to suboptimal implementation of fuel- and cost-saving behaviors and precludes fuel operations employees from incentivizing good performance. In addition to monitoring problems, captains may not believe operations personnel possess the practical knowledge to advise on best practice and, therefore, enforce optimal contracts.

[1] The "Captain"—as opposed to the "First Officer"—is the pilot on the aircraft who makes command decisions and is ultimately responsible for the flight's safety. As a rule, Captains are the most senior (and therefore most experienced) pilots in an airline. In the cockpit of a typical flight from New York to London, there would be one Captain, one First Officer, and a Flight Engineer (who does not actively fly the aircraft). Moreover, a vast majority of airline captains survive rigorous job market competition to secure their jobs, investing thousands of hours of training (privately or elsewhere) before obtaining the opportunity to be considered for a flying career with a major airline.

We capitalize upon recent technological developments that capture rich in-depth flight-level data from aircraft, allowing for robust scientific assessment of our uniquely designed interventions, which are aimed at improving captains' fuel efficiency. As a result, the principal can now monitor the fuel behaviors of the agents and incentivize profit-maximizing (i.e. fuel-minimizing) behaviors. However, while traditional labor economics would prescribe the use of various financial incentives and contract structures to align the interests of principals and agents, several factors prohibit the use of financial incentives in this context (e.g., contract costs, union barriers, and established industry conventions).

We report the first ever field experimental study on the behavior of transport personnel, focusing our attention on three main behavioral motivations for optimizing fuel use: complete information, targets motivated by reference-dependent preferences, and prosocial incentives. Provision of the right information about personal performance will allow the agent—here, the captain—to better optimize her behavior. Information provision may motivate behavioral change in two respects. First, the agent becomes aware that she is being monitored and that her behavior is clearly visible to the principal. Second, the information may be inconsistent with her own perceived performance, highlighting potential areas for improvement and stimulating an intrinsic desire to improve. The information provided is the percentage of flights for which each of three select fuel-efficient behaviors is properly implemented.

We additionally introduce feedback to the employees providing non-binding targets for attainment of the three behaviors that encapsulate reference-dependent preferences. We posit that a change in one's personal expectations from the status quo to an improved outcome can boost performance and, consequently, job satisfaction. In line with our theory, we therefore expect job satisfaction (i.e. utility) to increase for those who meet their targets. As in the Koszegi and Rabin ("KR", 2006) model of reference-dependent preferences, we assume individuals are loss averse in their reference dependency. In other words, performing below the target level will cause more disutility than exceeding the target level will benefit the individual. Unlike the KR model, we examine reference dependency in terms of individual performance relative to a given target attainment level, as opposed to monetary gains and losses. The targets are displayed just below the performance information to allow for direct comparison of actual and 'ideal' performance.

Our last behavioral motivation—prosocial incentives—is rooted in theories of (pure and impure) altruism. The warm glow model (Becker, 1974; Andreoni, 1989, 1990) is grounded in the notion that altruism exists in two varieties: pure and impure. The former requires that an individual derive utility from the benefits she directly receives from the provision of a public good. In other words, others' giving perfectly crowds out that of the individual, since she is only concerned with the amount of the public good provided and not the source of its funding. The latter posits that individuals gain private utility from the act of giving itself, so that an individual whose altruism is completely impure will provide the same amount 'x' toward the public good regardless of the provision of others toward the public good. Both pure and impure altruism provide positive utility to (altruistic) economic agents, and we assume that individuals are characterized by some combination of the two (and do not attempt to distinguish between them). We therefore incentivize some captains with charitable donations when they meet their performance targets (£10 per target met each month) to test the prediction that altruistic motivations will enhance performance.

In addition to presenting the first theory of the behavior of transport employees, we provide robust reduced-form estimates of three experimental treatments on the targeted behaviors.[2] Our experiment aims to increase fuel efficiency through the provision of salient information, exogenous targets, and altruistic (i.e. prosocial) incentives to aircraft pilots. Specifically, we run a large-scale framed field experiment that tests our theoretical predictions by providing monthly tailored feedback to 339 airline captains—the entire captain population of Virgin Atlantic Airlines—from February 2014 through September 2014. These captains flew primarily long-haul flights using various types of airbuses. We include a map of the routes flown below.

[2] There have been some investigations of different methodologies for assessing the welfare benefits of transport technology – see Anderson et al. (2011).



.

Our novel flight dataset identifies three behaviors airline captains regularly perform that contribute to overall fuel efficiency. The first is fuel load efficiency (i.e. Zero Fuel Weight adjustment), which assesses the accuracy with which fuel is loaded onto the plane at the captain's request given all relevant factors (e.g., weather, aircraft weight, etc.). Put simply, the airline wants to prevent captains from under-fuelling the aircraft (for safety purposes), and from over-fuelling the aircraft (for cost and safety purposes). Captains make this decision prior to taxi out. The second is in-flight efficiency (i.e. Efficient Flight), or how fuel-efficiently the captain operates the aircraft between takeoff and landing. In-flight fuel efficiency depends upon captains' optimization of speeds and altitudes, ideal flap settings, and compliance with air traffic control. The largest gains from efficiency occur during this phase of flight, though in-flight fuel efficiency is also largely dependent on initial fuel uptake. The third is taxiing efficiency (i.e. Reduced-Engine Taxi-In), or how efficiently the captain operates the plane once on the ground. Under most circumstances, the airline encourages captains to turn off at least one engine while taxiing to the gate.[3]

[3] These are three of many behaviors that the captain engages with during a given flight. There is a growing literature surrounding field labor economics, but most experiments have focused on simple tasks (List & Rasul, 2010; Levitt & Neckermann, 2015). Bandiera et al. (2007; 2009; 2010) demonstrate the effects of managerial compensation and social connections and in the workplace on worker productivity and selection in the fruit picking industry. Shearer (2004) find that piece rate wages improve worker productivity in tree planting relative to fixed rate wages. Field experiments on the impact of retail store-level tournaments on sales show mixed results (Delfgaauw et al., 2013; 2014; forthcoming). One exception to such task simplicity is Gibbs et al. (2014), who analyze the effects of a rewards program on innovation at a large Asian technology company in a field experimental setting. In an attempt to increase the value added provided to its customers, the company encouraged its employees to submit ideas on process and product improvements. They find

To Treatment Group 1, we provide basic performance feedback in terms of percentages of flights for which each of the three aforementioned behaviors is implemented. This group allows us to test and control for the impact of salient information regarding one's performance, ensuring that such information—as well as the consistent reminder of being monitored—is accurate and complete.[4] Treatment Group 2 receives the same monthly information as the first treatment group but the feedback additionally includes individualized targets for attainment of these three behaviors. The targets are not incentivized—they merely give the captain a personal reference point (i.e. expectation) for attainment.[5,6] According to our theoretical model, a captain will seek to meet the given targets due to the effect of her own perceived job performance on job satisfaction.

In accordance with the theories of (pure or impure) altruism (Andreoni 1989, 1990), our final treatment group (Treatment Group 3) is incentivized with charitable donations, or *prosocial incentives*. For each target met in a given month, a captain is credited with a donation of £10 (~$17) to a charity of her choice. In other words, each captain can donate up to £30 (~$51) per month for increasing her implementation of fuel-optimizing behaviors. To our knowledge, ours is the first randomized field study to use performance-based charitable incentives to increase employee effort.[7] Within our experiment, we did not change any organizational structures, training

---

that providing rewards for idea acceptance substantially increased the quality of ideas submitted. In an envelope-stuff experiment, Al-Ubaydli et al (forthcoming) find that quality is actually *higher* under piece-rate wages (contrary to predictions from economic theory), speculating a role for beliefs about employers' ability to monitor. In an artefactual field experiment with bicycle messengers, Burks et al. (2009) do find that performance pay reduces cooperation relative to a flat wage.

[4] This information is different to that in many other relevant field experiments, which provide feedback relative to a norm or peers (see Barankay, 2010; 2012). Our performance feedback is absolute.

[5] This treatment also allows us to control for the impact of such exogenous reference points when assessing the impact of charitable incentives, a common oversight in prior research

[6] The closest research in this area is the one-shot laboratory experiment with students by Goerg & Kube (2012) where the individual could choose his or her target and the financial incentive was randomized. They found no impact of the incentive but a positive and significant impact of the target on productivity. In a natural field experiment conducted with workers in a Chinese consumer electronics factory, Hossain and List (2012) find that workers respond differently if their bonuses are framed as losses as opposed to gains. The bonus works to increase productivity in both contexts, but the magnitude of the effect is greater for the loss frame. Similar results are found in Fryer et al. (2012) and Englmaier et al. (2012). Our treatment is different to these two papers because we introduce a reference point (i.e. the target) for the individual without changing the framing of a financial or non-financial incentive.

[7] See Imas (2014) and Charness et al. (2014) for lab experiments on the effect of charitable incentives on effort, and Anik et al. (2013) for a field study of unconditional charitable bonuses. Field experimental research into unconditional gifts is a burgeoning area of research – see Gneezy and List (2006) Bellemare and Shearer (2009) Hennig-Schmidt et al. (2010) Englmaier and Leider (2012), Kube et al. (2012), and Cohn et al. (2013).

schemes, or contracts with the airline captains, although we recognize that these could be important to productivity and efficiency.[8]

The experimental design allows us to first and foremost determine the efficacy of charitable incentives in motivating employees in a real-world high-stakes labor setting. To our knowledge, we are the first to perform a large-scale field experiment on firm employees in such a setting. Moreover, we are able to assess whether there are additional benefits of such incentives beyond sole provision of information and personal targets, i.e. free of cost to the principal.[9] In line with our theory, if the cost of effort for any given behavior exceeds the altruistic utility one derives from 'earning' charitable donations, we would expect that she will not improve her behavior, and vice versa. Likewise, if the cost of effort exceeds the satisfaction derived from performing one's job to the best of her ability, our theory predicts that she will not improve her performance.

Over an eight-month field experiment, we found nine significant and important results. Firstly, all three treatments were successful in increasing the prevalence of fuel-efficient behaviors. Secondly, exogenous targets are most effective in motivating captains to perform fuel-relevant behaviors—under some specifications, increasing the average incidence of these behaviors by more than 10 percentage points (0.26 standard deviations). Thirdly, while also effective in provoking behavior change when compared to a control group, charitable incentives do not provide further motivation beyond exogenous targets without incentives. In some cases these incentives perform significantly worse than the exogenous targets on their own, while adding an additional monetary cost to the principal. Regardless, even our most conservative estimates indicate that both treatments that provide exogenous targets to individuals significantly improve two of the three selected behaviors.

Fourthly, we do not find large decay effects in our treatment after each treatment is delivered. Fifthly, we find that our treatments work better on those who are not trainers – those captains who train other captains and first officers are less likely to positively respond to our three treatments. Sixthly, we find that our Target and Charity interventions work better on younger

[8] See Nagin et al. (2002), Hamilton et al. (2003), Karlan and Valdivia (2011), Bandiera et al. (2013), Bloom et al. (2013a;b), Karlan et al. (2014), Lyons (2014).

[9] The closest research to this free lunch approach is depicted in the field experiments of Gino and Grant (2010), Kosfeld and Neckermann (2011), Bradler et al. (2013), Chandler and Kapelner (2013), Gubler et al. (2013), Ashraf et al. (2014a; 2014b), Kosfeld et al. (2014).

captains than older captains, while the Information intervention worked better on older captains. Interestingly, we do not have any heterogeneous treatment effects with respect to captain seniority, although this is positively correlated with age.

Seventhly, in accordance with our theory, we find that job performance increases job satisfaction. This finding is consistent with the assertion that utility is reference dependent. Among the subjects who received personalized targets, those who met more targets over the course of the experiment are more satisfied with their jobs. Further investigation of this phenomenon reveals that performance on reduced-engine taxi in drives these results, increasing job satisfaction by 0.13 points (on an eleven-point scale) per monthly target met. In other words, a captain who met all of his RET targets (out of a possible 24) would rate his overall job satisfaction 3.12 points higher than a captain who did not meet any of his RET targets. Additionally, when we regress job satisfaction on treatment group indicators, all coefficients are positive, indicating that none of the interventions reduced job satisfaction for this population. In fact, captains in the charitable incentives group are statistically significantly more satisfied with their jobs than captains in the control group. This finding bodes well with the theoretical assumption that performance increases satisfaction, since our treatments had only neutral or positive effects on performance.

Eighthly, heterogeneity in captains' underlying preferences helps to explain our treatment effects. While risk preferences do not strongly predict the magnitude of the treatment effects of the three groups, we find a 2.5 and 3.3 percentage point increase (per unit increase in the risk aversion parameter r[10]) in ZFW implementation for the information and prosocial incentive groups, respectively. We find that ambiguity aversion is a strong and significant predictor of the treatment effect of the targets group. For instance, an increase of one on an eleven-point self-reported uncertainty aversion scale reduces implementation of EF and RET by approximately 1.4 percentage points and 2.5 percentage points (respectively) for the targets treatment group. Additionally, we find that prosocial incentives worked best for those captains with higher social preferences before the experiment started. Importantly, when examining differences in self-reported giving between 2013 and 2014, we find no evidence that charitable incentives crowd out private charitable giving behavior.

[10] We use the constant relative risk aversion range in Binswanger (1989) and Eckel & Grossman (2008) in calculating *r*. See section III.D. for more information.

Finally, we estimate that our treatments saved 392,750kg of fuel altogether, resulting in a savings of $390,000 for the airline over the eight-month experimental period. The experiment appears to have changed the behavior of all captains – control and treatment – due to the framed field experimental nature of the experiment. We therefore argue that the effects we find are lower bound estimates of what we would expect to find in reality (i.e. in a natural field setting).

The above findings hold important theoretical implications for both labor economics and the economics of charitable giving, and additionally provide a strong first step toward identifying practical and cost-effective fuel solutions for the air transport industry. The remainder of this paper is structured as follows. Section 2 presents the first theoretical model of the (fuel-use) behavior of transport personnel, with specific reference to airline captains. Section 3 describes the field experimental design and details its procedures. Section 4 displays the results, and section 5 concludes.

II. THEORETICAL MODEL

We model an airline captain's fuel use decisions per flight. We include her response in fuel usage to receiving our experimental treatments, which are i) information about her recent fuel usage, ii) information about target fuel usage, and iii) a donation to her chosen charity conditional on achieving the target usage. We base our model on the charitable giving model in DellaVigna, List, & Malmendier (2012; "DLM", hereafter), who provide a simple yet comprehensive theoretical framework on individual-level behavior related to charitable giving that comprises both self-interested and altruistic utility. We extend the model to incorporate a reference-dependent component to capture the effects of exogenous targets, and we omit the social pressure component since it is irrelevant to our study.

II.A. Industry Motivation

Fuel is a major cost in airlines' production functions, accounting for roughly 30 percent[11] of total operating costs (IATA, 2014). For this reason airlines are interested in cost-effective and efficient ways to lower the total amount of fuel used per flight. However, given their renowned expertise and prestige in the industry, airline captains are granted significant autonomy in their

[11] For the airline represented in this study, fuel accounts for 44% of operating costs.

decision-making across several fuel-relevant behaviors. Here, we exploit an airline captain's preferences for high-level job performance and prosocial outcomes to reduce airlines' fuel usage and costs.

Airline captains, like typical labor market participants in agency theory, want to maximize wealth and leisure while minimizing effort. However, when surveyed, captains in our study expressed a concern over fuel usage and fuel cost, both for environmental reasons and company profitability[12]. Since captains have a vast amount of flying experience[13], we assume that they are using the equilibrium amount of effort, and therefore also fuel, at time $t = 0$ (i.e. before the experiment begins).[14]

II.B. Model Setup

In our model, we assume that captains are at an equilibrium fuel usage with respect to their wealth, experience, effort, and concerns for safety, the environment, and company profitability. Since we are only manipulating some of these components in our experiment—namely wealth and effort— and because we assume that safety and environmental concerns are unaffected by our interventions, we do not incorporate any other variables into the captains' utility function.[15] This exclusion of parameters does not affect our predictions since, by nature of the study, these parameters are exogenous to treatment.

[12] Captains likely express concern for company profitability due in part to the structure of the labor market in which they operate. To become an airline captain requires many years of training and experience *within* an airline; if a captain loses her job with one airline and seeks employment in another, she loses her prior seniority and must work for many years to reinstate it. Thus, for the sake of their own job security, captains care about minimizing fuel costs.

[13] Explain the experience of the population we are looking at (i.e. minimum years of experience). In order to become captain one must consistently demonstrate her competence/expertise and airlines will only promote to captain once they feel comfortable with the consistency of their behavior (*consider this more thoroughly and rephrase*), so in this respect it does not matter that experience varies across captains.

[14] This assumption implies that a captain's concern for the environment does not interact with our experimental treatments to affect her effort or fuel usage and is, therefore, left out of her utility function.

[15] While it may seem instinctive that encouraging captains to carry more or less fuel and turn off engines while taxiing would influence the safety of the flight, it must be noted that all airlines operate within robust international safety standards. In the case of our UK-based partner airline, these standards comply with those set by the UN International Civil Aviation Organization, the European Aviation Safety Agency, and the UK Civil Aviation Authority. Moreover, as in our case, many airlines maintain stricter safety standards than those imposed upon them by these external bodies. All communications with the captains as part of this study have stressed the overriding importance of safety as the airline's utmost priority and assured captains that these interventions fall within the airline's stringent safety standards. Finally, all targets are capped at 90% to alleviate pressure and to allow captains some flexibility in their decision-making on any given flight.

We define:

$$i := \begin{cases} 1, & if\ receive\ information\ on\ fuel\ use \\ 0, & otherwise \end{cases}$$

$$r := \begin{cases} 1, & if\ i = 1\ and\ receive\ a\ fuel\ usage\ target \\ 0, & otherwise \end{cases}$$

$$d(e) := \begin{cases} p, & if\ r = 1\ and\ a\ charitable\ donation\ is\ made\ when\ fuel\ usage\ target\ is\ met \\ 0, & otherwise \end{cases}$$

The probability of meeting the target conditional on effort, $p \in [0, 1]$, is characterized by complete information in the map from effort to fuel usage on both the lower and upper bounds. That is, captains can be certain that they do not meet a target if they put in little effort, and they can be certain that they have achieved the target if they put in sufficient effort. However, whether or not a charitable donation is granted depends on the captain's actions over a month-long period, so p takes a value on the interior of the support for the majority of the month. For example, if a captain chooses to exert a high amount of effort in the beginning of the month, she is not guaranteed to meet the target. Conversely, if she chooses to exert a low amount of effort in the beginning of the month, she may still be able to meet the target.

We consider a decision theory problem where captains decide how much effort to exert while on the job. Let $e_{ijt}^{q} \in \{e_{ijt}^{l}, e_{ijt}^{h}\}$ be captain i's effort choice of quality q for task j in time period t. The tasks consist of the aforementioned flight metrics, i.e. Zero Fuel Weight adjustment, Efficient Flight, and Reduced-Engine Taxi-In. Captains observe their effort and a signal of fuel usage, $f_{it} + (1\text{-}i)*\mathcal{E}_{it} = \overline{f_{it}}$. The signal is noisy unless the captain receives information. Based on captains' extensive flying experience—and the basic information provided to all captains regardless of treatment (e.g., pre-flight zero fuel weight reports)—we can assume $E(\mathcal{E}_{it}) = 0$.

Each captain also has an ideal fuel usage f_I, which is based on her own experience and preferences. In period $t = 1$, if a captain receives an exogenous target (r=1), $f_I = f_T$, where f_T is a specific number set by the airline.

The captain has utility:

$$U(w,s,e,f,g) = u(w,e,g) + av(d(e)*g, g_o, G_{-i}) + y(s,e,f) - c_j(e) \qquad (1)$$

In this model, *u* is utility a captain receives from wealth, *v* is utility from giving to charity, *y* is utility from job performance, and *c* is utility from exerting effort. Each of these components is discussed in detail below.

Utility from wealth. Similar to DLM, for wealth *W*, charitable giving from the airline *g* for meeting the target (if applicable), and other charitable giving g_o, *u* is defined as follows:

$$u(w,e,g) := u(W - g_o(d(e)*g) + \tilde{a}*d(e)*g) \qquad (2)$$

where,

$$\tilde{a} = \begin{cases} 0 \; if \; a < 0 \\ a \; if \; 0 \leq a \leq 1 \\ 1 \; if \; a > 1 \end{cases}$$

Private consumption is an individual's wealth minus the amount given to charity from that person's wealth (i.e. not from this study). However, to ensure that *u* is continuously differentiable, we need to account for the effect of charitable donations resulting from our treatments on utility from private consumption. To capture this effect, we multiply the individual's expected donation, *d(e)*g,* by a function of *a*—a parameter capturing preferences for giving—which we call $\tilde{a}$. As in DLM, the parameter *a* is non-negative in the case of pure or impure altruism and negative in the case of spite[16], and $\tilde{a}$ is simply *a* truncated at 0 and 1.

The reasons for creating such boundaries on the term capturing preferences for giving are twofold. First, an individual with spiteful preferences ($a < 0$) does not get less utility from private consumption when she donates to charity than when she does not donate to charity; therefore, $\tilde{a}$ is censored from below at 0. Second, an individual with pure or impure altruistic preferences will get

[16] As defined in Andreoni (1989, 1990), pure and impure altruism capture two possible motivations for giving. The first stems from a preference solely for provision of the public good, so that an individual's donations are entirely crowded out by donations from other sources. Impure altruism, on the other hand, refers to the phenomenon whereby individuals receive direct utility from the act of giving itself, i.e. through "warm glow". Spite, as defined in DLM, exists when an individual gets disutility from donating to the charity.

additional utility from her private consumption by giving to charity through our treatment because it corresponds to an outward shift in the budget constraint in the dimension of giving to the chosen charity. However, $\tilde{a}$ is censored from above at 1 because an individual will experience weakly more utility from increases in *W* than from giving to the chosen charity (i.e. $\frac{\partial u}{\partial W} \geq \frac{\partial u}{\partial g}$). This relation holds since increases in *W* shift the budget constraint outward in all dimensions—including the charitable giving dimension—so these must be weakly preferred to shifts in only one dimension. This stipulation is important to assume differentiability in *u* in a standard expected utility framework, as in DLM.

Please note that the amount an individual gives to other charities will be related to the amount that she gives to charity in the context of this study. Captains will smooth their consumption for giving. If a captain normally gives $100 to charity each year and this year she gives $10 through the context of the study, we would expect her total giving to be between $100 and $110, or $g_o + g \in [100,110]$. The realization of the sum depends on the value of *a* and whether *a* stems from pure altruism, impure altruism, or spite. We should expect that an individual who has a negative *a* value does not donate to charity outside of the context of this study since donating to charity decreases that individual's utility.

Utility from charitable giving. The *v* term is also adapted from DLM and follows the same properties for each type of individual (pure or impure altruistic and spiteful). The main difference between the *v* term in this study and that in DLM is that in this study, not everyone has the opportunity to donate to charity (i.e. $d(e) > 0$ for only one treatment group). We also assume that *v* is separable in its parameters, as follows:

$$v(d(e) * g, g_o, G_{-i}) = a v_1(d(e) * g, G_{-i}) + a v_2(\theta g_o, G_{-i})$$

where θ is the cost of giving through channels other than the study. In this specification, v_1 represents utility from giving in the study context and v_2 represents utility from giving from one's personal wealth. Note that $v_1(0, G_{-i}) = 0$ since if $d(e) = 0$, then a captain is not able to donate to the charity through the context of the study, so v_1 should not affect the utility function (similar to the spite case). Based on the arguments made above with respect to consumption smoothing, $v|_{d=0} \leq v|_{d=p}$, $0 = v_1|_{d=0} \leq v_1|_{d=p}$, $v_2|_{d=0} \geq v_2|_{d=p}$. That is, a captain's utility from giving is at least as high for those captains for whom $d(e)=p$ as it is for those captains for whom

d(e)=0, which follows from our assumption that giving in the study context can only decrease giving from one's own wealth or not affect it at all. Finally, since $\frac{\partial p}{\partial e} > 0$, we have $\frac{\partial v}{\partial e} \geq 0$.

In the case of pure altruism, an individual should get the same utility from giving to charity from her personal wealth as from giving to charity through the context of the study, since the benefit to the charity is identical. In this sense, *v* can be thought to represent the charity's production function. In the case of impure altruism, an individual should also get the same utility from donating to charity through her personal wealth as she does from donating through the context of the study because the amount donated on her behalf is the same. Lastly, in the case of spite, $g_0 = 0$ since giving to charity decreases utility and so those individuals will not give to charity independently of the study. Note, $v(0) = 0$ because if a person does not give to charity in person then her utility from giving to charity in person is 0.

Utility from job performance. Since captains care about fuel efficiency, and since imposing exogenous targets on performance affects a captain's perception of how well she is doing her job, we include a parameter *y* capturing job performance.[17] We assume *y* is separable in safety (*s*) and fuel (*f*) because changes in fuel as a result of the study do not affect safety levels, as argued in our assumption above. A captain whose performance exceeds her target will achieve higher utility under this parameter than a captain who does not achieve her target. Similarly, a captain will experience less (more) utility the further below (above) the target is her performance. We therefore incorporate job performance into the model as follows:

$$y\,(s,e,f) = y_1(s) + y_2(e,f) = \; y_1(s) + \; y_2\left(-\bar{f}\middle|-f_I\right)$$

where

$$y_2\left(-\bar{f}\middle|-f_I\right) = \; y_{2m}\left(-\bar{f}\right) + \; y_{2n}\left(-\bar{f}\middle|-f_I\right)$$

and

$$y_{2n}\left(-\bar{f}\middle|-f_I\right) = r * \mu(y_{2m}\left(-\bar{f}\right) - \; y_{2m}(-f_I))$$

[17] Evidence indicates that influencing job performance positively influences job satisfaction (or utility), whether through increased self-esteem or perceived managerial support for autonomous decision-making (Christen, Iyer, & Soberman, 2006; Pugno & Depedri, 2009).

Here, y_2 is defined as in Koszegi & Rabin (2006; KR hereafter). We denote the components of y_2 "m" and "n" to mirror the notation in KR. As in KR, m represents the "consumption utility" and n represents the "gain-loss utility." These terms are separable across dimensions. Finally, μ is the "universal gain-loss function" and has the associate properties outlined in KR.

Note that pilots get utility from using less fuel $\frac{\partial y2}{\partial f} < 0$ and, conditional on receiving a reference point, get utility (disutility) from performing above (below) the target, which increases with distance from the target according to μ. We assume μ is linear and $\mu(x) = \eta x$ if $x > 0$ and $\mu(x) = \eta\lambda x$ if $x \leq 0$ for $\eta > 0$, $\lambda > 1$, in accordance with theories of loss aversion. Moreover, following naturally from our definition of μ, we assume $y(x) = x$. If a captain does not receive a reference point, her utility does not comprise gain-loss utility, so for these individuals $y_2 = y_{2m}$. That is, if r=0, captains do not receive information regarding ideal performance with respect to fuel efficiency, so their job performance parameter depends solely on fuel consumption.

Additionally, based on industry standards and emphasis on safety as well—as the design of the treatments—we assume that captains' job performance utility from flying safely is constant across treatments, therefore:

$$\frac{\partial y}{\partial s} = S$$

(Dis)utility from effort. Finally, c_j represents the cost of effort. The cost function is subscripted by j to convey that various tasks have different costs associated with them. The cost structure is a function of the difficulty of the task itself (e.g., it may be easier to turn off one engine after landing than to have an efficient flight for several hours) and resistance due to previous habit formation (e.g., captains who for many years have not properly performed the Zero Weight Fuel calculation may find it difficult or bothersome to begin doing so). Additionally, the costs for each task are separable since the tasks are done independently. Therefore,

$$c_j = \sum_j c(e_j^q), where\ c\left(e_j^q\right) \neq c\left(e_{-j}^q\right)\ for\ j \neq -j.$$

For a captain to decrease her fuel use, she must also increase her effort, i.e. $\frac{\partial f}{\partial e} < 0$. Note that $c_j(e)$ is subtracted in the utility equation, so $\frac{\partial u}{\partial c} < 0$, $\frac{\partial c}{\partial e} > 0$. Based on interviews with captains, the

cost of effort increases at an increasing rate. Defining the cost of effort as a quadratic function of effort implies that the cost of effort increases with the amount of effort exerted (i.e. $\frac{\partial^2 c}{\partial e^2} > 0$).

Proposition 1: Information will cause captains to increase effort and thereby decrease fuel usage.

Proposition 2: Captains in the target and charity conditions will choose to increase their effort the most in tasks that are the easiest to meet the target.

Proof: Since the firm sets the targets and donations exogenously, the utility for meeting a target is constant across tasks. The donation to charity is the same across tasks as exogenously determined and since the targets are also exogenously determined, the captains believe that the firm values them all equally by revealed preference. If the firm did not value them equally, then it would not offer the same reward. However, the cost function is not constant across tasks for reasons described earlier, which implies that the captains will choose to increase their effort on tasks for which targets are easiest to meet.[18]

III. EXPERIMENTAL DESIGN

The framed field experimental design[19] aims to capture the effect of information provision both on its own and interacted with exogenous targets and prosocial incentives. The information and incentives targeted three behaviors relevant to fuel use: Zero Fuel Weight adjustment (ZFW), Efficient Flight (EF), and Reduced-Engine Taxi-in (RET). The targeting of these three behaviors allows us to capture decision-making before takeoff, during the flight, and after landing (respectively).

[18] Our theory and interventions are rooted in Holmstrom's (1979) Informativeness Principle, which states that *any* accessible information about an agent's effort should be used in the design and enforcement of optimal contracts. Our interventions are not aimed at the efficient allocation of effort across these tasks—as proposed in Holmstrom & Milgrom (1991) and Baker (1992)—since we assume our three behaviors are not substitutable (since they occur during different phases of flight). We acknowledge the possibility that additional fuel-efficient behaviors exist that we do not measure that may be fully or partially neglected due to our treatments.

[19] Ideally, we would have conducted a natural field experiment in which the subjects did not know they were under scrutiny (Harrison & List, 2004), though privacy policies internal to the airline precluded such a design. By conducting a framed field experiment, we open up the possibilities that i) treated subjects change their behavior simply because they know they are being monitored (Hawthorne effects), or ii) control subjects alter their behavior in spite of the lack of treatment (John Henry effects).

Airline captains have access to basic information regarding their fuel use on any given flight, but they do not currently receive detailed information relating their decision-making to their fuel efficiency. The information provided to treatment groups in this study is intended to create a clear and direct link between behavioral inputs and fuel outputs, eliminating informational deficiencies regarding the consequences of captains' decision-making. Recent advances in data collection in the aviation industry allow us to obtain precise information regarding several variables relevant to both on-the-ground and in-flight fuel efficiency on a per-flight basis.

Therefore, we partnered with Virgin Atlantic Airlines to provide accurate monthly feedback to three treatment groups over the course of eight months (a control group did not receive any feedback) across 339 captains. Printed feedback forms with information from the previous month's flights were sent to the home address of each captain in the middle of each month of the study, so that captains received their first feedback form in March 2014 and their final feedback form in October 2014.

III.A. Background

Provision of information and targets with respect to these behaviors ensures that captains have the opportunity to conserve fuel—and, for some, donate to charity—in each stage of the flight experience. Moreover, these behaviors are highly deterministic of discretionary (i.e. non-mandated) fuel use, and detailed data allow for precise monitoring of captains' decision-making at each of these three stages of flight. [20]

[20] Globally, analysis performed by the ICAO's Expert Operational Goals Group (IEOGG) indicates a worldwide operational fuel efficiency of 87.25% in 2010, which corresponds to a "benefit pool" of 12.75% in potential efficiency gains (International Civil Aviation Organization 120-121). Furthermore, due to projected increases in demand for air travel and consequent rises in congestion levels, operational inefficiency on a business-as-usual trajectory will decrease to 81.5% by 2040. In light of this projection, the IEOGG's goal of reducing fuel usage per flight by an average of 9% from operational improvements alone by 2040 is quite ambitious, and there is not yet a defined plan for achieving these goals. The operational measures discussed in the report include both measures we capture in our experiment—e.g., reduced engine taxiing and speed and altitude optimization—as well as some that are beyond the control of airline captains and therefore outside the realm of our experiment, such as improved weather forecasting and air traffic management. However, the report does not account for inefficiencies due to over-fuelling and does not specify exact fuel savings from itemized operational changes, which is left for future research.

Zero Fuel Weight (ZFW) Adjustment. Before each flight, captains utilize flight plan information in conjunction with their own professional judgment to determine initial fuel uptake, which is approximately 90% of the anticipated fuel necessary for the flight. This amount is fuelled into the aircraft prior to the loading of passengers and cargo. Once all passengers and luggage are boarded, captains receive updated information regarding the final weight of the aircraft and may upwardly adjust the fuel on the aircraft accordingly. The information they receive from flight operations includes a ZFW measure, which indicates the weight of the aircraft with the revenue load (i.e. passengers and cargo) and without fuel, as well as the Takeoff Weight (TOW), which includes the revenue load and the fuel.

Captains may then perform a ZFW calculation in which they first calculate the amount by which they should increase fuel uptake based on the TOW—a formula that is standard across the airline—and subsequently compute a second iteration to account for the fuel necessary to carry the fuel that they have elected to add to the plane. If the amount of fuel already on the plane is sufficient according to these calculations, the captain may choose not to add any additional fuel. Additionally, should the captain choose to carry more or less fuel than that prescribed by the double iteration calculation, she should account for the deviation in the "ACARS box" (a computer system on the flight deck), which is then submitted back to the airline for data collection purposes.

Our first metric is a binary variable indicating whether this double iteration has been performed and fuel load adjusted accordingly. The captain successfully performs the behavior if her final fuel load—adjusted for her ACARS adjustments—is within 200kg of the "correct" amount of fuel. This allowance prevents penalizing captains for rounding and slight over- or under-fuelling on the part of the fueller. Sustainable Aviation (2012; hereafter "SA"[21]) estimates that more accurate fuel-loading can save approximately 0.5% of total fuel burn; with average fuel consumption per flight of 69,500kg, this fuel savings corresponds to about 350kg fuel per flight.[22] Accurate ZFW adjustment is the only behavior we monitor that our partner airline claims should be performed on every flight regardless of circumstances – hence it should ideally be 100% for all captains.

[21] Sustainable Aviation is an alliance of UK airlines, airports, aerospace manufacturers, and air navigation service providers with the long-term objective of reducing air and noise pollution.

[22] Using data from its 2014 sectors, Virgin Atlantic estimated an average fuel savings of 250kg per sector for correct ZFW calculations.

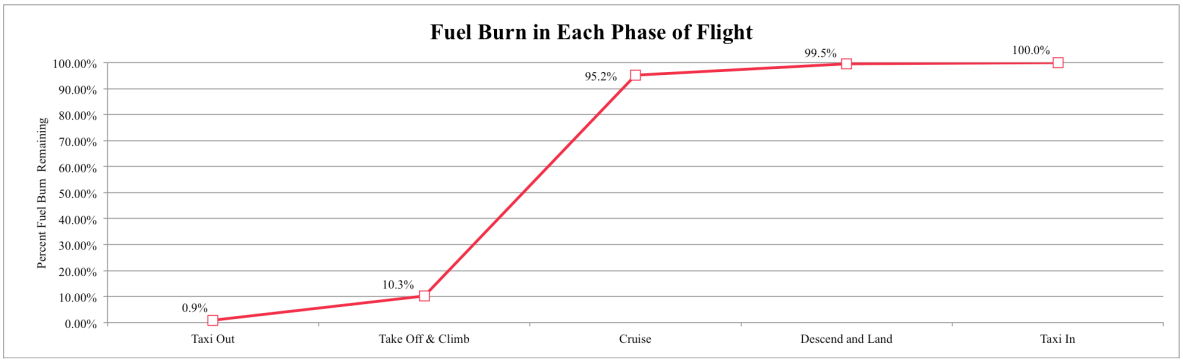
Efficient Flight (EF). With respect to optimization of in-flight behavior, our comprehensive EF metric captures whether captains use less fuel during flight than is expected according to the flight plan. We use this metric to understand whether captains have made efficiency-enhancing decisions between takeoff and landing. This measure incorporates several in-flight behaviors that augment fuel efficiency, such as requesting and executing optimal altitudes and shortcuts from air traffic control, maintaining ideal speeds, and ensuring optimal aerodynamic arrangements with respect to takeoff and landing gear and flap settings. While this measure is the sole targeted measure regarded as a behavioral output, it is the only in-flight measure that allows captains the flexibility to achieve the target while using professional judgment to ensure that safety remains the top priority. Under some circumstances, we would expect the captains to sacrifice fuel efficiency for the sake of maximizing safety, so we would not expect even a "model" captain to perform this metric 100% of the time.

Reduced-Engine Taxi-in (RET). Finally, once the plane has touched down, captains may choose to turn off one (or two, in a four-engine plane) of their engines while they taxi after landing. A captain meets the criteria for this metric if she turns off (at least) one engine during taxi in. SA estimates that this practice—in conjunction with reduction in auxiliary power use (APU) substitution—prevents the emissions of 100,000 tonnes of CO_2 annually. Fuel savings from RET depend on scheduling and delays as savings are estimated on a per-minute basis.[23] Similar to EF, there are circumstances under which we would not expect or prescribe the use of RET. Constraints to implementation of RET include inexperience, geographical constraints (e.g., the placement or layout of the runway), limited skill in taxiing, and the complexity of the taxi route (e.g., number of stops, turns, or cul-de-sacs encountered).

The graph below demonstrates fuel use in various stages of a typical long-haul flight within Virgin Atlantic. Zero fuel weight adjustment takes place prior to taxi out and is the only behavior we aim to improve that affects fuel burn through the entire duration of the flight, since extra fuel is burned to carry the weight of excessive fuel loaded. Efficient flight implementation affects fuel

[23] Fuel savings also depend on aircraft type and only begin to accrue after engines have cooled, which takes 2-5 minutes. Savings per minute for aircraft operated within our study are as follows: 12.5kg (B744, A330), 8.75kg (A346), and 6.25kg (A343).

burn between takeoff and landing, when most fuel is consumed. Finally, reduced-engine taxi in is responsible for the final 0.5% of fuel burn on a typical flight.



III.B. Interventions

Treatment Group 1: Information. Each feedback form outlines the captain's performance relative to the three behaviors for the prior month. The feedback detailed the behaviors under scrutiny and the percentage of flights flown for which the captain efficiently performed each of the three behaviors. For instance, if a captain flew four flights in the prior month, successfully performing ZFW and RET on two of the flights and EF on three of the flights, his feedback form would indicate a 50% attainment level for ZFW and RET and a 75% attainment level for EF.

Treatment Group 2: Targets. Captains in Treatment Group 2 received the same information as those in Treatment Group 1 but subjects in the former were additionally encouraged to achieve personal targets of 25% above their personal baseline attainment levels for each metric (capped at 90%). The targets were communicated to captains who participated in Treatment Groups 2 and 3 prior to the beginning of the experiment on a sample feedback form, which they received in the mail. An additional box is included in the feedback form to provide a summary of performance (i.e. how many targets were met in total). If at least two of the three targets were met, captains were recognized with an injunctive statement ("Well Done!") and encouraged to continue to fly efficiently the following month. If fewer than two targets were met, captains were encouraged to fly more efficiently to reach their targets. Captains were not rewarded or recognized in any public or material fashion for their achievements.

Treatment Group 3: Charitable Incentives. In addition to the information and targets provided to captains in Treatment Group 2, those in Treatment Group 3 were informed that achieving the same targets resulted in donations to charity. Specifically, for each target achieved in a given month, £10 was donated to a charity of the captain's choice on her behalf. Therefore, each captain in Treatment Group 3 had the opportunity to donate £30 (~$51) per month for a total of £240 ($408) to her chosen charity over the course of the eight-month trial. Captains were reminded each month of the potential remaining donations that could result from realizing their targets. Table 1 outlines the treatments (see Appendix for examples of each of the three feedback forms).

III.C. Experimental Setup

Randomization. To randomize subjects into the four study groups,[24] the pre-experimental data are first blocked on five dummy variables which captured whether subjects were above or below average for i) number of engines on planes flown, ii) flights flown per month, and iii) attainment for the three target behaviors. The former two variables were those that proved significant in determining our outcome behaviors in preliminary regressions, while the three target behaviors are our main dependent variables. Once blocked, subjects in each block were randomly allocated to one of the four study groups through a matched quadruplet design. To ensure that individual-specific observable characteristics are balanced across groups, we performed balance tests for gender, seniority, age, trainer status, and trusted pilot status. In addition to checking for balance across the variables on which the data were blocked, we checked for balance on flight plan ramp fuel (i.e. as a proxy for average distance of flights) and whether captains flew disproportionately on weekends (see Table 3).

Communication with captains. Two weeks prior to the beginning of the study, all captains were informed that Virgin Atlantic would be undertaking a study on fuel efficiency. The initial letter outlined the behaviors being measured and the possible study groups to which the captains may be assigned. Captains in treatment groups were to receive letters the following week to inform

[24] To determine the optimal experimental design, it is essential to consider the probability α of falsely rejecting the null hypothesis (type I error) and the probability $1-\beta$ of failing to reject a null hypothesis that is not true (type II error). Since there do not exist any obvious predecessors to our study, we do not have a reliable notion of the expected effect size and therefore cannot assess the ideal tradeoff between significance and power levels (see Zhang & Ortmann, Mimeo). For our sample size calculations, we use three months (September 2013 through November 2013) of pre-experimental data on captains' flight behaviors. We employ the conventional significance and power levels (α=0.05, β=0.20) to calculate our minimum detectable effect (MDE) sizes given the constraint on our sample size (N=339).

them of what to expect in the coming months. In the final week of January, letters were sent to all treatment group subjects informing them of the intervention to which they had been assigned and providing a sample feedback report, which included the individual's targets if she was assigned to Treatment Groups 2 or 3.

From February 1, 2014 to October 1, 2014, we gathered all flight-level data on a monthly basis for each captain in Virgin Atlantic and mailed a feedback report to the home addresses of all captains in Treatment Groups 1-3. Captains were encouraged to engage with the material and ask questions to an email address set up specifically for study inquiries. Once the experiment was complete, we sent treated captains a letter informing them of their overall monthly results with respect to their targets (if in Treatment Group 2 or 3) and their total charitable donations (if in Treatment Group 3). All (treatment and control) captains were informed that a follow-up survey would be sent to their company email addresses in early 2015. Each captain received an email with a survey link on January 29, 2015, and the survey closed three weeks later.

Sample. Our data consists of 339 captains, of which 333 are male and 6 are female. A total of 216 of these captains at least partially completed the study debrief survey and 189 completed it. Of the respondents, 97 classified their training as military and 102 as civilian (the remaining declined to state). At the start of the experiment the average age in the captain population was 48 years. Eleven captains are "trusted pilots" whom the study team consulted regarding study feasibility and communications, and 62 captains are "trainers" who are responsible for keeping captains and first officers up to date in the latest techniques. An average captain had been an employee of Virgin Atlantic for over 17 years and flew approximately 5.4 flights each month.

III.D. Follow-Up Survey

The follow-up survey serves two main purposes: 1) to elicit preferences over risk and ambiguity; and 2) to understand subjects' altruism and satisfaction with their jobs. Risk and ambiguity preferences are relevant due to the personal and subjective nature of captains' decision-making, especially with respect to fuel uptake. For instance, we may conjecture that captains who are highly risk averse will be more inclined to take extra fuel "just in case" bad weather or other unexpected obstacles are encountered. We would expect that more altruistic captains who are given prosocial incentives would be inclined to put in additional effort. Finally, we include job

satisfaction as we are interested to see if our treatments as well as captains' performance influence job satisfaction.

Preferences over risk and ambiguity. To accomplish our primary survey objective of eliciting preferences over risk, we asked the subjects a series of three incentive compatible risk questions derived from the Binswanger (1980) method of risk preference elicitation (see Appendix III). Each question asks subjects to choose one of five possible gambles, each with two possible outcomes that remain constant across the three questions. Gamble 1 is the most certain (£25 vs. £25), while Gamble 5 is the most risky (£105 vs. -£15). Both Gambles 4 and 5 hold potential for losses (i.e. the second outcome is negative). In the first question, both outcomes are associated with a 50% probability of selection.

In the second and third questions—whose orders were randomized—the first and second outcomes are associated with a 75% (25%) and a 25% (75%) probability of selection. Subjects were told that one of the three questions would pay out at the end of the survey. The latter questions allow for assessment of probability weighting, which indicates an individuals' degree of sensitivity to changes in probability. We use the midpoint of the associated range of a particular gamble as our risk parameter (see Appendix III). The CRRA range is the range of r in the utility function $U = x^{(1-r)}/(1-r)$—which is characterized by constant relative risk aversion—for which an individual would select each of the five gambles. The gambles used in the survey are scaled up from those used by Eckel & Grossman (2008); all outcomes are multiplied by 2.5. Therefore, the CRRA ranges are identical.

Following the above incentive-compatible risk questions, we ask a series of questions that require self-assessment. These three questions are taken directly from the Need for Closure Scale and are meant to assess decisiveness and aversion to ambiguity (Webster & Kruglanski, 1994). In these questions, subjects are asked to rate on a scale from 0 to 10 how much they agree with the following statements: "*I don't like situations that are uncertain*"; "*I would describe myself as indecisive*"; "*I usually make important decisions quickly and confidently.*" We use the responses to determine whether uncertainty aversion and decisiveness affect the captains' responses to our treatments.

Altruism and job satisfaction. To assess captains' altruism, we asked two questions regarding individuals' private charitable donations in 2013 and in 2014 outside of the context of the study. Captains indicated their donation behavior by selecting one of ten multiple-choice intervals ranging from £0 to £200+. The information provided further allows us to ascertain whether our prosocial incentive treatment crowds out (or in) private giving.

Finally, we asked captains to rank their job satisfaction on a seven-point scale to help us to assess the impact of the interventions on Captains' job satisfaction. The question comes directly from the British Household Panel Survey, which also allows us to compare the satisfaction of VAA captains compared to the UK national average. Most importantly, the responses allow us to look at both whether job performance influences job satisfaction—as we propose in our theory—as well as whether our treatments contribute positively or negatively to captains' satisfaction with their jobs.

III.E. Econometric Specification

Our first specification uses pooled OLS regressions of ZFW, EF, and RET attainment on the three treatment dummy variables, as well as controls for weather[25] (W'_{it}) on departure and arrival and the number of engines (E_{it}) on the plane:

$$Y_i = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta'_4 W'_{it} + \beta_5 E_{it} + \varepsilon_i \qquad (3)$$

Despite having randomized and balanced on pre-experimental attainment of three behaviors, we also use a difference-in-difference (DiD) identification strategy below to tease out the treatment effects using all of the data. In addition to the controls in Equation 3 above, the model includes both individual and time fixed effects:

$$Y_{it} = \beta_0 P_t + \beta_1 T_{1,i} + \beta_2 T_{2,i} + \beta_3 T_{3,i} + \beta_4 (P_t \times T_{1,i}) + \beta_5 (P_t \times T_{2,i}) + \beta_6 (P_t \times T_{3,i}) + \beta'_7 W'_{it} + \beta_8 E_{it} + \varepsilon_{it} \qquad (4)$$

We use panel data from January 2013 through September 2014, and we treat the first day of the experiment as February 1, 2014.[26]

[25] Weather Underground is the source of our continuous weather data. As controls, we include temperature at time and location of departure and arrival and dummy variables for the weather condition (clear, cloudy, light precipitation, heavy precipitation, and haze/fog) upon departure and arrival.

We include results from Newey-West regressions due to the long panel nature of our data and the relatively small number of clusters (i.e. captains) in our sample. Newey-West estimators correct for autocorrelation and heteroskedasticity by "fixing" the estimated standard errors from OLS, allowing for error terms to be correlated with those in *m* preceding or subsequent periods. It is conceivable here that there may be autocorrelation in the error term for consecutive flights, since a captain may, for example, fly with the same First Officer on a round trip or be similarly (un)familiar with a given destination airport, neither of which we capture in our regressions.

Since lag optimization is somewhat subjective—especially when a panel is unbalanced—we run Newey-West regressions using two different lags (m=1 and m=4). To determine our "optimal" lag length, we use the first step of the formula recommended by Newey & West (1994) – $m(T)=\text{floor}(T/100)^{2/9}$ – and calculate the lag for T_{min} (5 flights, m=1), T_{max} (113 flights, m=4), and T_{mean} (67 flights, m=1) in our sample.

V. RESULTS

V.I. Main results

We provide estimates of the effects of our interventions on captains' execution of the three selected fuel-efficient behaviors. Table 1 provides summary statistics for captains' behavior before and after we begin experimental monitoring. In the 13 months leading up to the experiment, captains successfully performed a ZFW adjustment on approximately 42.3% of their flights, which increased to 46.2% of flights after monitoring began. We see much larger changes in behavior for EF and RET, even among captains in the Control Group. Average attainment of an efficient flight among all captains increased from 31.8% to 49.3%, and similarly from 35.5% to 56.9% for reduced-engine taxi in. These changes appear to be driven by both experimenter effects and treatment effects.

The interventions—especially the personalized targets—appear to have a drastic effect on captains' behavior, in particular on the implementation of reduced-engine taxi in. As shown in Figures 1-3, attainment of captains in the Control Group is several percentage points below even the

[26] Our results are robust to use of receipt of the first feedback report (March 15, 2014) as our start date.

least effective treatment groups for each behavior, with captains in the Targets Group achieving the behaviors most frequently. It is clear from the figures as well that *all* captains increased their implementation of these behaviors during the experimental period, a phenomenon which we attribute to experimenter effects owing to the framed nature of the study. Nevertheless, the raw before-and-after data indicates that captains receiving treatments were indeed more likely to achieve the metrics.

Table 2 reports the results of post-intervention OLS regressions of ZFW, EF, and RET[27] attainment on the three treatment dummy variables, as well as controls for weather on departure and arrival and the number of engines on the plane. According to the standard OLS specification—and noting the limited sample size and, therefore, statistical power—targets are the only intervention with a significant treatment effect on ZFW, increasing implementation by almost four percentage points, i.e. an 8.7% treatment effect (or 0.08 standard deviations (σ)). We find a similar effect (9.1% effect, 0.09σ), for targets on the prevalence of EF, while information (5.7% effect, 0.05σ) and charitable incentives (7.7% effect, 0.07σ) also significantly increase the implementation of this behavior. Most striking is the effect of the interventions on the occurrence of reduced-engine taxi in, which occurred on over 10% more flights for those receiving targets alone (20.5% effect, 0.21σ) and approximately 8% more flights for those receiving information (15.3% effect, 0.15σ) or charitable incentives (16.5% effect, 0.17σ).

Since each treatment builds upon the last—e.g., feedback in the Targets Group builds upon the information provided in the Information Group by adding personalized exogenous targets, holding everything else constant—we "control" for the contents of previous treatments and are therefore able to make comparisons across treatments as well. While information and charitable incentives appear to have a positive effect on the incidence of fuel-efficient behaviors compared to the control group, it is clear from the above regressions that motivating captains with personalized

[27] RET data is physically stored on QAR cards inside the aircraft, which are removed every 2-4 days to pull data. These cards can corrupt or overwrite themselves, and also can reach full memory capacity before being removed. Therefore, data capture for RET is not complete—approximately 35% of flights are missing data for this metric. The reason for the missing data is entirely random and cannot be influenced by captains, so this phenomenon should not affect the results beyond reducing the power of the estimates.

targets *without* charitable incentives is the most effective strategy for motivating a shift toward more fuel-efficient behaviors.[28]

Table 3 contains the results from three DiD regressions: a fixed-effects regression with errors clustered at the captain level and a Newey-West regression where errors may be correlated up to a lag of four periods (the latter will be discussed here). The Targets Group are the only group to have improved their attainment of ZFW significantly by 3.3 percentage points (7.4% effect, 0.066σ, $p<0.10$), while information ($\beta=0.008$, 1.7% effect, 0.02σ) and charitable incentives ($\beta=0.026$, 5.8% effect, 0.05σ) have positive but non-significant effects. Both targets ($\beta=0.041$, 8.7% effect, 0.08σ, $p<0.05$) and charitable incentives ($\beta=0.031$, 6.6% effect, 0.06σ, $p<0.10$) significantly increased attainment of EF; information did not have a significant effect, though its effect was again positive ($\beta=0.021$, 4.4% effect, 0.04σ). As predicted in the theory, the largest improvements were seen in the incidence of RET. All three groups—Information ($\beta=0.091$, 18.1% effect, 0.18σ, $p<0.01$), Targets ($\beta=0.114$, 22.7% effect, 0.23σ, $p<0.01$), and Charitable Incentives ($\beta=0.108$, 21.5% effect, 0.22σ, $p<0.01$)—responded strongly and favorably to their treatments on this metric.

The reported DiD estimates confirm the effectiveness of the exogenous targets provided to captains in the Targets Group, providing solid basis to conclude that targets are associated with increased implementation on approximately 3%, 4%, and 10% of flights for ZFW, EF, and RET, respectively. Charitable incentives are also highly effective in encouraging captains to fly efficiently and turn off engines while taxiing. However, since the Charitable Incentives Group received both targets *and* incentives, we cannot conclude that the incentives themselves are impacting behavior. In fact, since the effect sizes are similar across these two groups, one might speculate that the targets are responsible for most (if not all) of the behavioral effect. Similarly, according to the DiD estimates, we can only conclude that informational feedback on its own is effective in encouraging captains to turn off at least one engine while taxiing in. That is, since all treatment groups significantly improve RET outcomes and to similar degrees, and all groups receive informational feedback on the number of flights for which the behavior was implemented, we

[28] When comparing the effects for all three behaviors across treatment groups, we find a significant difference between ZFW attainment in the targets group and both the information group ($\beta=0.029$, $p=0.05$) and the charitable incentives group ($\beta=0.030$, $p=0.05$) using the OLS specification. Targets are also marginally more effective than information in encouraging implementation of RET ($\beta=0.023$, $p=0.10$). Targets do not significantly outperform the other treatments for EF, and they do not outperform charitable incentives for RET.

cannot conclude that targets or charitable incentives provide additional impetus to turn off engines after landing.

V.II. Heterogeneity

Here we examine whether the treatments are particularly (in)effective for particular types of individuals based on various observable characteristics of our sample. We do so by interacting each of the following five variables with the DiD estimators reported above, creating triple interactions between the post-treatment indicator, the treatment groups, and the following variables. First, we examine whether "trainer" captains are more responsive to our treatments. While some might conjecture that trainers are less likely to change their behavior due to lack of direct control over the aircraft (i.e. since a trainee is often in control), it is also possible that trainer captains are in a better position to improve behavior due to the attention they must pay to—and the influence they have on—other pilots' behaviors. Second, we look at the average number of engines in planes flown by any given captain. We do so because the chosen behaviors may be inherently more or less difficult to implement depending on the size of the plane.

Third, we look at the age of the captain, since both operating procedures and aircraft are constantly evolving. Captains across the age spectrum may have been trained differently or may hold systematically variable degrees of reluctance to changes in their habits. Fourthly, we look at the seniority of the captain. Captains with tenure may feel less pressure to respond to encouragement from internal operations personnel due to high job security. Conversely, they may strive to set strong examples for younger pilots in the airline. Finally, we look at the salience of the treatment in terms of time lapsed since receiving the feedback form. Many behavioral studies have demonstrated the importance of salience in effecting behavior change.[29] We look at whether captains respond particularly strongly in the first week after receipt of feedback throughout the study duration.

In Table 5 we examine impacts according to whether the captain holds "trainer" status. Trainers are responsible for training both captains and first officers in new techniques and implementing spot checks to ensure that captains continue to fly to the airline's standards. The

[29] Many studies have addressed the issue of saliency (or lack thereof) to prices – for an example, see Chetty et al. (2009).

Trainers should be some of the "best" captains in the fleet, and should ideally perform these three behaviors as optimally and frequently as possible.[30] The results in Table 5 indicate that trainers respond less well to information and targets, implementing EF on 11.7% ($p<0.01$) and 17.4% ($p<0.01$) fewer flights than non-trainers. Additionally, trainers in the Targets Group perform RET on 14.3% ($p<0.05$) fewer flights than non-trainers. Coefficients are negative for EF and RET in all treatment groups, while coefficients are (slightly) positive for ZFW for all groups. This finding is consistent with our theory if we assume that trainers are already performing these behaviors at maximal effort levels prior to treatment, as we would expect.

In Table 6, we look at results according to whether the captain flew a two- or a four-engine plane. Coefficients for the Information (β=-0.039) and Targets Groups (β=-0.050, $p<0.10$) are considerably negative for ZFW, indicating that captains are less likely to implement ZFW in these groups if they fly a plane with four engines. In contrast, captains who receive these treatments and fly four-engine planes are more likely to implement RET (β_{info}=0.024; $\beta_{targets}$=0.064, $p<0.10$). Captains in four-engine planes are more likely to implement EF, especially in the Charitable Incentives Group (β=0.057, $p<0.05$).

Table 7 examines how captains' age and seniority interacts with the treatments. Both exogenous targets and charitable incentives have a lesser impact on older individuals, decreasing the treatment effect on ZFW by 13.6 percentage points ($p<0.01$) and 11.6 percentage points ($p<0.05$), respectively. However, older captains are more likely to respond to information in implementing EF by 8.0 percentage points ($p<0.10$). Although one would expect age and seniority to be highly correlated, seniority appears to have no impact on treatment effects.

To examine whether there is a saliency effect, we create an indicator variable that equals 1 if the flight took place within one week of receiving the feedback report and equals 0 otherwise. We interact this indicator variable with the DiD estimates, as above, and find that saliency does not increase performance significantly. However, a vast majority of the coefficients on these interactions hover around two to four percent, and none are below a tenth of a percentage point

[30] Prior to February 1, 2014, trainer captains' properly performed ZFW, EF, and RET on 52.7%, 39.8%, and 37.8% of flights (resp.), while non-training captains performed these behaviors on 40.8%, 30.6%, and 35.2% of flights. In other words, prior to intervention, based on summary stats in the 13 months prior to treatment, it is clear trainers outperformed non-training captains.

below zero (see Table 8). These numbers indicate that saliency may have a slightly positive impact on performance, though our results do not allow us to make strong conclusions on this front. In sum, we cannot definitively argue that there is a decay effect of treatment for the behavior of treated captains.

V.III. Survey Results

We report the results from the study's follow-up survey below. The survey elicited 216 partial responses (and 189 complete responses) out of a total of 339 captains[31], which represents an impressive 64% (56%) response rate. We find that there are no statistically significant differences between the captains who participated in the survey and those who did not from any of the observables in our data (performance on the three behaviors, age, seniority, treatment group, number of flights during experimental period).

We first examine the impact that risk aversion has on our treatment effects. Table 9 provides the output from regressions based on the difference-in-difference econometric specification with additional triple interactions between the post-treatment indicator variable, each treatment group, and our risk aversion parameter *r*. We do not find that different levels of risk aversion influence captains' behavioral responses to the treatments apart from the ZFW adjustment, which increasingly risk-averse individuals are more likely to implement in both the Information and Charitable Incentives Groups ($p<0.10$). Specifically, an increase of one in *r* increases the treatment effect of information and prosocial incentives by 2.5 percentage points (5.6% effect, 0.05σ, $p<0.10$) and 3.3 percentage points (7.4% effect, 0.07σ, $p<0.10$), respectively. However, the triple interactions are positive for all treatment groups, so in general we infer that more risk-averse individuals are more likely to respond to the treatments.

However, if we control for aversion to losses, we find that levels of risk aversion do matter—and in the opposite direction—when considering the effect of charitable incentives. In Table 10, we run the same specification as in Table 9, except that we control for whether individuals chose a gamble with or without negative outcomes, using this binary variable as a proxy for loss aversion. When we include a triple interaction of the post-treatment indicator with the

[31] This response rate was achieved after sending each captain up to three emails within four weeks offering incentives up to £105.

treatment group dummies as well as a dummy indicating whether the selected gamble includes a negative outcome (i.e. our loss aversion proxy), an increase of 1 in a captain's *r* value decreases attainment of the EF and RET metrics by 4.5 percentage points (9.6% effect, .09σ, $p<0.05$) and 5 percentage points (10.0% effect, 0.1σ, $p<0.10$), respectively. That is, increasing risk aversion appears to decrease treatment response to charitable incentives for those two behaviors. Additionally, the impact of loss aversion cancels out the impact of risk aversion on the treatment effect of prosocial incentives for RET, increasing attainment by 21.5 percentage points (42.8% effect, 0.4σ, $p<0.01$). One could interpret that as aversion to losses of the charitable incentives with which they have been 'endowed'.

Tables 11-13 display treatment effects based on ambiguity preferences. We find that generally those captains in the targets group with less aversion to uncertainty and more confidence in their own decision-making are more responsive to targets. Specifically, an increase of one (on a scale of 11) in self-reported uncertainty aversion reduces implementation of EF and RET by approximately 1.4 percentage points (3.0% effect, 0.03σ, $p<0.10$) and 2.5 percentage points (5.0% effect, 0.05σ, $p<0.01$), respectively.[32] An increase of one on self-reported indecisiveness decreases the treatment effect of targets on ZFW by 2.9 percentage points (6.5% effect, 0.06σ, $p<0.10$), though the significance disappears when we add controls for weather and number of engines on the plane. Finally, an increase of one on self-reported confidence in decision-making increases the treatment effect of targets on ZFW and EF by 1.9 percentage points (4.3% effect, 0.04σ, $p<0.05$) and 2.0 percentage points (4.3% effect, 0.04σ, $p<0.01$), respectively. Apart from weak significance for this measure on the treatment effect of prosocial incentives on EF by 1.5 percentage points (3.2% effect, 0.03σ, $p<0.10$), these three measurements of (dis)comfort with uncertainty only influence the treatment effects with respect to targets. One interpretation of these results is that our treatments encourage those captains who usually make strong decisions in the "wrong" direction to make strong decisions in the "right" direction, whereas those who are less certain of circumstances or themselves might not make the assertive and reactive (to treatments) decisions to improve fuel efficiency.

[32] Significance disappears for EF when we include controls for weather and the number of engines on the plane, though the magnitude of the effect is similar. The magnitude of the effect on RET increases by 0.5 percentage points when we include controls.

We next move on to the impact of social preferences. We use self-reported donations to charity in 2013 (i.e. before the study began) as a proxy for altruism to determine whether subject's (perceived) prosocial preferences influence their response to our treatments. As expected, we find that charitable donations increase treatment response to the two most demanding behaviors—ZFW and EF—by 3.0 percentage points (6.7% effect, 0.06σ, $p<0.05$) and 1.1 percentage points (2.3% effect, 0.02σ) per £1 donated, respectively, and has no effect on treatment response in the information only group. Interestingly, we also find that a £1 increase in donations to charity in 2013 decreases attainment in the targets-only group by 1.4 percentage points (3.0% effect, 0.03σ, $p<0.10$) and 2.3 percentage points (4.6% effect, 0.05σ, $p<0.05$) for EF and RET, respectively. In other words, altruism has a positive effect on treatment responses to prosocial incentives and a negative effect on treatment responses to targets.

Finally, in Table 16, we assess whether job performance is a predictor of job satisfaction, as proposed in our theory. We argued in our theory that a captain whose performance exceeds her target will achieve higher utility than a captain who does not achieve her target in accordance with reference dependency theory. To test this conjecture, we aggregated the number of targets met over the course of the experiment for all captains. That is, for those in groups that did not receive a personalized target (i.e. Control and Information Groups), we created 'artificial' targets that mimic those provided to captains in the Targets and Charity Treatment Groups. For each captain, we summed up the number of (hypothetical) targets met during the course of the study and regressed job satisfaction on this job performance variable. For the Targets and Charity Groups only, we find that job satisfaction increases by 0.048 points ($p<0.10$) for every target met during the study; this effect disappears when we restrict the regression to groups who did not receive a target (i.e. the Control and Information Groups). We interpret this finding as evidence of reference dependency in the context of job performance. Furthermore, since treatment assignment is random, we attribute causality to the relationship between job performance and satisfaction.

To further refine this estimate, we run a similar regression that instead divides the independent variable into three separate variables capturing the number of targets met for each of the three fuel-efficient behaviors – see Table 17. The results of this regression indicate that RET attainment is the driver of the effect of performance on satisfaction, increasing self-reported job satisfaction by 0.129 points for every successful attainment of the monthly target. One potential

reason behind this interesting job satisfaction result is that RET is arguably the easiest of the behaviors to achieve, though we (experimenters, VAA) signal them as equally desirable through the equal weighting of charitable incentives across behaviors.[33]

V.IV Fuel savings

We use engineering estimates to understand the denoted impact of our interventions on fuel use. The airline has calculated that if correct execution of the ZFW adjustment would save 250kg per flight. For this airline, we had treatment effects of 3.3% and 2.6% for the targets and charity groups, respectively. With 6,608 flights over the eight-month intervention period, our interventions increased the implementation of correct ZFW adjustment by 193 flights, equivalent to a savings of 48,250kg of fuel over an eight-month period.

For EF, we estimate that an efficient flight uses 500kg less fuel than an inefficient flight. Our effect sizes for the three groups were 2.1%, 4.1%, and 3.1% respectively. Our estimate of fuel savings from these three groups aggregate to 325 extra flights being efficient over the eight-month period, which translates to 162,500kg in fuel savings.

For RET, we use an average estimate of fuel wastage of 9kg per minute if RET is not performed. Given an average taxi-in time of 20 minutes, fuel savings per flight are approximated to be 180kg. We estimate that our treatment groups led to and additional 1,013 extra flights having met RET, which led to a fuel savings over an eight-month period of 182,000kg. Summing these savings together, our interventions led to around 392,750kg of fuel saved over an eight-month period. Using the airline's standard conversion of 3.15kg of CO_2 per kg of fuel burned and the January 2015 IATA jet fuel price of \$993/1000kg, CO_2 savings amount to 1,237,163kg. The experiment saved the airline approximately \$390,000 in fuel costs and \$8,127 (\$6.57/tonne) in carbon costs.

[33] We will add a section on persistence in the middle of April 2015, when we will receive six additional months' worth of data on captains' behavior in the aftermath of the study.

VI. Discussion

We use a framed field experiment to explore the opportunity to change the fuel efficiency of captains in the airline industry. We randomized three interventions to understand the impact on employee performance of basic informational feedback, exogenous targets associated with that information, and charitable incentives associated with these same targets and information. We find that the interventions were successful in increasing fuel-efficient behaviors and conclude that our inexpensive strategies are both feasible and cost-effective to induce airline captains to behave more efficiently.

While our workplace setting is complex with myriad competing incentives at play, clear measurement of the behavior enables innovative strategies to provide the right set of interventions to improve employee productivity and firm performance. Future research should aim to identify additional behavioral motivators to improve the fuel efficiency of airline captains and other transport personnel as a cost-effective means to simultaneously reduce the operation costs of firms in the transportation industry and prevent unnecessary emissions of globally detrimental greenhouse gas emissions.

REFERENCES

Air Transport Action Group (2015), 'Facts & Figures', March 2015, http://www.atag.org/facts-and-figures.html.

Al-Ubaydli, O., Andersen, S., Gneezy, U., and List, J., 'Carrots that Look Like Sticks: Toward an Understanding of Multitasking Incentives Schemes', *Southern Economic Association*, forthcoming.

Anderson, S., Parry, I.W.H., Sallee, J.M., and Fischer, C. (2011). Automobile Fuel Economy Standards: Impacts, Efficiency, and Alternatives. *Review of Environmental Economics and Policy*, **5**(1), 89-108.

Andreoni, J. (1989) 'Giving with impure altruism: applications to charity and Ricardian equivalence', *The Journal of Political Economy*, 1447-1458.

Andreoni, J. (1990), 'Impure altruism and donations to public goods: a theory of warm-glow giving', *The Economic Journal*, 464-477.

Anik, L., Aknin, L. B., Norton, M. I., Dunn, E. W., & Quoidbach, J. (2013), 'Prosocial bonuses increase employee satisfaction and team performance', *PloS one*, **8**(9), e75509.

Ashraf, N., Bandiera, O., and Jack, K. (2014a), 'No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery', Manuscript.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014b), 'Awards Unbundled: Evidence from a Natural Field Experiment', *Journal of Economic Behavior & Organization*, **100**, 44-63.

Baker, G. P. (1992), 'Incentive contracts and performance measurement', *Journal of Political Economy*, 598-614.

Bandiera, O., Barankay, I., and Rasul, I. (2007), 'Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment', *Quarterly Journal of Economics*, **122**(2), 729-73.

Bandiera, O., Barankay, I., and Rasul, I. (2009), 'Social Connections and Incentives in the Workplace: Evidence from Personnel Data', *Econometrica*, **77**(4), 1047-94.

Bandiera, O., Barankay, I., and Rasul, I. (2010), 'Social Incentives in the Workplace', *Review of Economic Studies*, **77**(2), 417-58.

Bandiera, O., Barankay, I., & Rasul, I. (2013), 'Team incentives: evidence from a firm level experiment', *Journal of the European Economic Association*, **11**(5), 1079-1114.

Barankay, I. (2010), 'Rankings and Social Tournaments: Evidence from a Field Experiment', University of Pennsylvania, CEPR and IZA Working Paper.

Barankay, I. (2012), 'Rank Incentives Evidence from a Randomized Workplace Experiment', Working Paper, The Wharton School, University of Pennsylvania.

Becker, G. (1974), 'A theory of social interactions'. NBER Working Paper.

Bellemare, C., and Shearer, B. S. (2009), 'Gift Giving and Worker Productivity: Evidence from a Firm-Level Experiment', *Games and Economic Behavior*, 67, 233-44.

Binswanger, H. P. (1980), 'Attitudes toward risk: Experimental measurement in rural India', *American Journal of Agricultural Economics*, **62**(3), 395-407.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013a), 'Does Management Matter? Evidence from India', Quarterly Journal of Economics, **128**(1), 1-51.

Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2013b), 'Does Working from Home Work? Evidence from a Chinese Experiment', NBER Working Paper 18871.

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2013), 'Employee Recognition and Performance: A Field Experiment', CESifo Working Paper 4164.

Burkhardt, U., & Kärcher, B. (2011), Global radiative forcing from contrail cirrus, *Nature climate change*, **1**(1), 54-58.

Burks, S., Carpenter, J., and Goette, L. (2009), 'Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment', *Journal of Economic Behavior & Organization*, **70**(3), 458-69.

Chandler, D., and Kapelner, A. (2013), 'Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets', *Journal of Economic Behavior & Organization*, **90**, 123-33.

Charness, G., Cobo-Reyes, R., & Sanchez, A. (2014), 'The effect of charitable giving on workers' performance: Experimental evidence', (No. 14/06). Department of Economic Theory and Economic History of the University of Granada.

Chetty, R., Looney, A., and Kroft, K. (2009), 'Salience and Taxation: Theory and Evidence,' *American Economic Review*, **99**(4), 1145-1177.

Christen, M., Iyer, G., & Soberman, D. (2006), 'Job satisfaction, job performance, and effort: A reexamination using agency theory', *Journal of Marketing*, **70**(1), 137-150.

Cohn, A., Fehr, E., and Goette, L. (2013), 'Fair Wages and Effort Provision: Combining Evidence from the Lab and the Field', University of Zurich Department of Economics Working Paper 107.

Delfgaauw, J., Dur, R., Non, A., and Verbeke, W. (2014), 'Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment', *Labour Economics*, **28**, 1-13.

Delfgaauw, J., Dur, R., Non, A., and Verbeke, W. (2013), 'The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment', *Journal of Labor Economics*, forthcoming.

Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013), 'Tournament Incentives in the Field: Gender Differences in the Workplace', *Journal of Labor Economics*, **31**(2), 305-26.

DellaVigna, S., & John, A. List, and Ulrike Malmendier (2012), 'Testing for Altruism and Social Pressure in Charitable Giving', *Quarterly Journal of Economics*, **127**(1), 1-56.

Eckel, C. C., & Grossman, P. J. (2008), 'Forecasting risk attitudes: An experimental study using actual and forecast gamble choices', *Journal of Economic Behavior & Organization*, **68**(1), 1-17.

Englmaier, F., Roider, A., and Sunde, U. (2012), 'The Role of Salience in Performance Schemes: Evidence from a Field Experiment', CESifo Working Paper 3771.

Englmaier, F. and Leider, S. (2012), 'Managerial Payoff and Gift Exchange in the Field', CESifo Working Paper Series No. 3707.

Fryer, R. G., Levitt, S. D., List, J., and Sadoff, S. (2012), 'Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment', NBER Worker Paper 18237.

Gibbs, M., Neckermann, S., and Siemroth, C. (2014), 'A Field Experiment in Motivating Employees', IZA Discussion Paper No. 8096.

Gino, F., and Grant, A. (2010), 'A Little Thanks Goes a Long Way: Explaining Why Gratitude Expressions Motivate Prosocial Behavior', *Journal of Personality and Social Psychology*, **98**(6), 946-55.

Goerg, S., and Kube, S. (2012), 'Goals (th)at Work—Goals, Monetary Incentives, and Workers' Performance', Working Paper Series of the Max Planck Institute for Research on Collective Goods 2012/19.

Gneezy, U., and List, J. (2006), 'Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments', *Econometrica*, **74**(5), 1365-84.

Gubler, T., Larkin, I., and Pierce, L. (2013), 'The Dirty Laundry of Employee Award Programs: Evidence from the Field', Harvard Business School Working Paper 13-069.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003), 'Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation', *Journal of Political Economy*, **111**(3), 465-97.

Harrison, G. W., & List, J. A. (2004), 'Field experiments', *Journal of Economic Literature*, 1009-1055.

Hennig-Schmidt, H., Rockenbach, B., and Sadrieh, A. (2010), 'In Search of Workers' Real Effort Reciprocity—A Field and a Laboratory Experiment', *Journal of the European Economic Association*, **8**(4), 817-37.

Hölmstrom, B. (1979), 'Moral hazard and observability', *The Bell Journal of Economics*, 74-91.

Hölmstrom, B., & Milgrom, P. (1991), 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design', *Journal of Law, Economics, & Organization*, 24-52.

Hossain, T., and List, J. (2012), 'The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations', *Management Science*, **58**(12), 2151-67.

ICAO Environmental Report 2013: Aviation and Climate Change. Environment Branch, ICAO, 2013.

Imas, A. (2014), 'Working for the "warm glow": On the benefits and limits of prosocial incentives', *Journal of Public Economics*, **114**, 14-18.

International Air Transport Association (2014), 'Fact Sheet: Fuel.'

Karlan, D., and Valdivia, M. (2011), 'Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions', *Review of Economics and Statistics*, **93**(2), 510-27.

Karlan, D., Knight, R., and Udry, C. (2014), 'Consulting and Capital Experiments with Micro and Small Tailoring Enterprises in Ghana', Working Paper.

Kosfeld, M., and Neckermann, S. (2011), 'Getting More Work for Nothing? Symbolic Awards and Worker Performance', *American Economic Journal: Microeconomics*, **3**(3), 86-99.

Kosfeld, M., Neckermann, S., and Yang, X. (2014), 'Knowing that You Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition', Tinbergen Institute Discussion Paper 14-043.

Köszegi, B., & Rabin, M. (2006), 'A model of reference-dependent preferences', *The Quarterly Journal of Economics*, 1133-1165.

Kube, S., Marechal, M. A., and Puppe, C. (2012), 'The Currency of Reciprocity: Gift Exchange in the Workplace', *American Economic Review*, **102**(4), 1644-62.

Lee, D. S., Fahey, D. W., Forster, P. M., Newton, P. J., Wit, R. C., Lim, L. L., Owen, B., & Sausen, R. (2009), Aviation and global climate change in the 21st century, *Atmospheric Environment*, **43**(22), 3520-3537.

Levitt, S.D. and Neckermann, S. (2015). What Field Experiments Have and Have Not Taught Us about Managing Workers. *Oxford Review of Economic Policy*.

List, J., and Rasul, I. (2010), 'Field Experiments in Labor Economics', in D. Card, and O. Ashenfelter (eds.), (Vol. 4a), *Handbook of Labor Economics*, North Holland.

Lyons, E. (2014), 'Team Production in International Labor Markets: Experimental Evidence from the Field', Working Paper.

Metcalfe, R., & Dolan, P. (2012), "Behavioural economics and its implications for transport,' *Journal of Transport Geography*, **24**, 503-511.

Nagin, D. S., Rebitzer, J. B., Sanders, S., and Taylor, L. J. (2002), 'Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment', *American Economic Review*, **92**(4), 850-73.

Penner, J. E. (Ed.). (1999), *Aviation and the global atmosphere: a special report of IPCC Working Groups I and III in collaboration with the Scientific Assessment Panel to the Montreal Protocol on Substances that Deplete the Ozone Layer*, Cambridge University Press.

Pugno, M., & Depedri, S. (2009), 'Job performance and job satisfaction: An integrated survey', Universita Degli Studi di Trento, Dipartimento di Economia, Discussion Paper 4.

Shearer, B. (2004), 'Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment', *Review of Economic Studies*, **71**(2), 513-34.

Smith, Patrick (2013), *Cockpit Confidential: Everything You Need to Know About Air Travel*, Naperville, IL; Sourcebooks Inc.

Sustainable Aviation CO_2 Roadmap, March 2012.

Zhang, L., & Ortmann, A. (2013), 'Exploring the meaning of significance in experimental economics', UNSW Australian School of Business Research Paper 2013-32.

FIGURES

Figure 1: ZFW metric before and during the experiment across the experimental groups

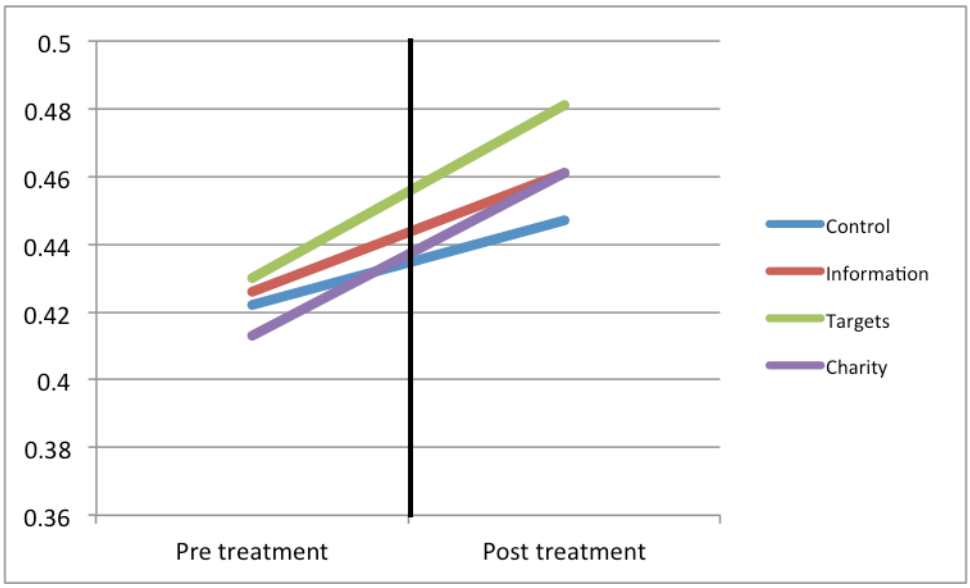


Figure 2: EF Metric before and during the experiment across the experimental groups

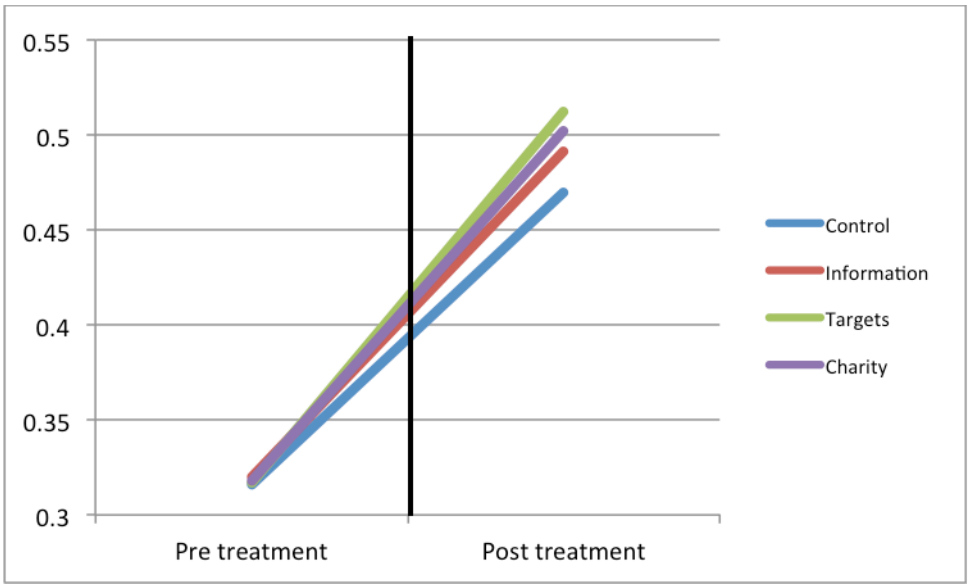
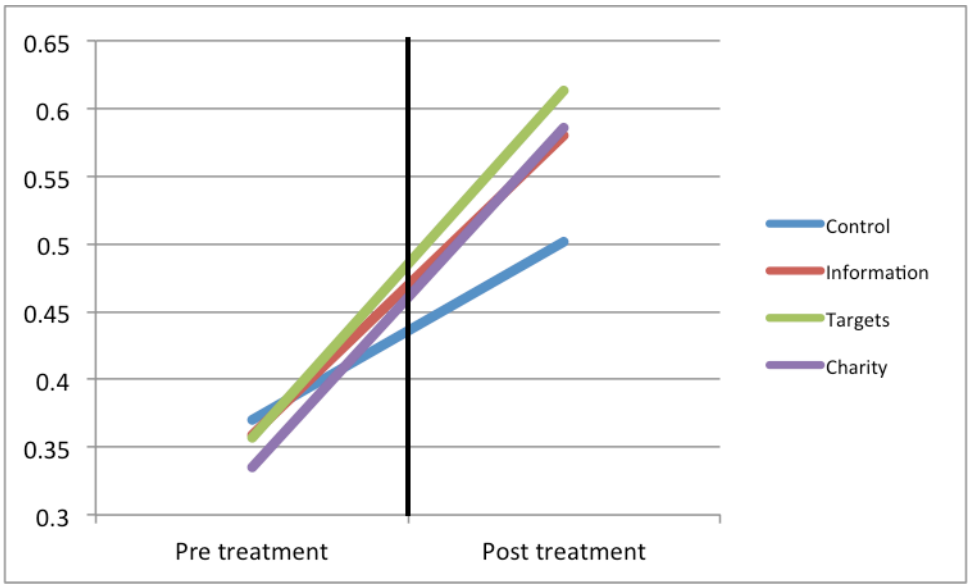


Figure 3: RET metric before and during the experiment across the experimental groups



TABLES

TABLE 1
AVERAGE ATTAINMENT OF ZFW, EF, AND RET BEFORE AND AFTER INTERVENTION BEGINS

Variable (proportion of flights)	Control	Treatment 1: Information	Treatment 2: Targets	Treatment 3: Charitable Incentives
ZFW, before	0.422 (0.494) 5528 obs	0.426 (0.495) 5718 obs	0.430 (0.495) 5452 obs	0.413 (0.493) 5406 obs
ZFW, after	0.447 (0.497) 3504 obs	0.461 (0.499) 3529 obs	0.481 (0.500) 3156 obs	0.461 (0.499) 3452 obs
EF, before	0.316 (0.465) 5528 obs	0.320 (0.467) 5718 obs	0.317 (0.466) 5452 obs	0.318 (0.466) 5406 obs
EF, after	0.470 (0.499) 3504 obs	0.491 (0.500) 3529 obs	0.512 (0.500) 3156 obs	0.502 (0.500) 3452 obs
RET, before	0.370 (0.483) 3294 obs	0.359 (0.480) 3447 obs	0.357 (0.479) 3249 obs	0.335 (0.472) 3227 obs
RET, after	0.502 (0.500) 2428 obs	0.580 (0.494) 2155 obs	0.613 (0.487) 2155 obs	0.586 (0.493) 2388 obs

Notes: The table reports the proportion of flights for which captains in a given group performed each of the three selected behaviors. Due to random memory errors, RET data is unavailable for 40.2% of pre-experimental flights and for 30.8% of post-experimental flights. This missing data is in no way systematic and therefore does not bias the results, though it moderately reduces the power of the RET estimates. Standard deviations are reported in parentheses, and *N* captures the total number of observations (flights) from which the summary statistics are calculated.

TABLE 2
OLS REGRESSION FOR POST-INTERVENTION PERIOD

	ZFW	EF	RET
T1: Info	0.015	0.027**	0.077***
	(0.012)	(0.012)	(0.014)
T2: Targets	0.039***	0.043***	0.103***
	(0.012)	(0.012)	(0.014)
T3: Charity	0.014	0.036***	0.083***
	(0.012)	(0.012)	(0.014)
Constant	0.763***	0.451***	0.348***
	(0.032)	(0.032)	(0.037)
N	*N*=339 subjects		
Obs	34,779 (22,187)		
Controls	Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Robust errors are clustered at the captain level. Standard errors are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 3
DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS FOR ZFW, EF, AND RET

	Fixed Effects			Newey-West (lag=4)		
	ZFW	EF	RET	ZFW	EF	RET
Post*Info	0.007	0.010	0.064**	0.008	0.021	0.091***
	(0.017)	(0.018)	(0.027)	(0.018)	(0.017)	(0.022)
Post*Targets	0.020	0.024	0.090***	0.033*	0.041**	0.114***
	(0.018)	(0.019)	(0.028)	(0.018)	(0.017)	(0.022)
Post*Charity	0.022	0.031*	0.075***	0.026	0.031*	0.108***
	(0.016)	(0.019)	(0.028)	(0.018)	(0.017)	(0.023)
Constant	0.575***	0.426***	0.404***	0.636***	0.348***	0.204***
	(0.026)	(0.026)	(0.027)	(0.022)	(0.021)	(0.028)
N	*N*=339 subjects			*N*=339 subjects		
Obs	34,779 (22,187)			34,779 (22,187)		
FE	Time, Individual			None		
Controls	Weather, # Engines			Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Robust errors are clustered at the captain level. Standard errors are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 4

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, WITH QUADS

	ZFW	EF	RET
Post*Info	0.008	0.018	0.074***
	(0.016)	(0.016)	(0.020)
Post*Targets	0.015	0.029*	0.103***
	(0.016)	(0.017)	(0.021)
Post*Charity	0.018	0.025	0.089***
	(0.016)	(0.016)	(0.020)
Post	0.027**	0.169***	0.137***
	(0.012)	(.011)	(.014)
Info	0.004	0.004	-0.012
	(0.010)	(0.009)	(0.013)
Targets	0.003	0.002	-0.016
	(0.010)	(0.009)	(0.013)
Charity	-0.011	0.006	-0.022
	(0.010)	(0.010)	(0.013)
Constant	0.554***	0.370***	0.204***
	(0.033)	(0.032)	(0.042)
N	*N*=339 subjects		
Obs	34,779 (22,187)		
FE	None		
Controls	Weather, # Engines, Quads		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 5

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, BY TRAINER

	ZFW	EF	RET
Post	0.031***	0.149***	0.145***
	(0.012)	(0.010)	(0.014)
Trainer	0.119***	0.092***	0.026*
	(0.012)	(0.011)	(0.015)
Post*Info	0.008	0.029**	0.079***
	(0.015)	(0.014)	(0.019)
Post*Targets	0.029*	0.057***	0.128***
	(0.016)	(0.015)	(0.019)
Post*Charity	0.007	0.030**	0.084***
	(0.015)	(0.015)	(0.019)
Post*Trainer	-0.031	0.080**	0.023
	(0.037)	(0.032)	(0.045)
Post*Info*Trainer	0.012	-0.117***	-0.021
	(0.046)	(0.040)	(0.055)
Post*Targets*Trainer	0.009	-0.174***	-0.143**
	(0.047)	(0.042)	(0.056)
Post*Charity*Trainer	0.025	-0.044	-0.015
	(0.046)	(0.040)	(0.056)
_cons	0.408***	0.306***	0.352***
	(0.004)	(0.004)	(0.006)
N	N=339 Captains		
Obs	35,745 (22,650)		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 6

DIFFERENCE-IN-DIFFERENCE ESTIMATES, BY NUMBER OF ENGINES

	ZFW	EF	RET
Post	0.016	0.129***	0.052**
	(0.019)	(0.017)	(0.020)
Four_Engines	-0.144***	0.013*	0.095***
	(0.008)	(0.007)	(0.011)
Post*Info	0.037	0.020	0.066**
	(0.023)	(0.021)	(0.027)
Post*Targets	0.074***	0.026	0.061**
	(0.025)	(0.024)	(0.027)
Post*Charity	0.014	-0.004	0.096***
	(0.024)	(0.022)	(0.027)
Post*Four_Engines	0.008	0.033	0.142***
	(0.022)	(0.021)	(0.025)
Post*Info*Four_Engines	-0.039	0.002	0.024
	(0.028)	(0.027)	(0.032)
Post*Targets*Four_Engines	-0.050*	0.019	0.064*
	(0.029)	(0.029)	(0.033)
Post*Charity*Four_Engines	-0.007	0.057**	-0.011
	(0.028)	(0.027)	(0.033)
Constant	0.524***	0.309***	0.290***
	(0.007)	(0.006)	(0.009)
N	N=339 Captains		
Obs	35,745 (22,650)		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 7
DIFFERENCE IN DIFFERENCE REGRESSION COEFFICIENTS, BY AGE & SENIORITY

	Age			Seniority		
	ZFW	EF	RET	ZFW	EF	RET
Post	-2.724***	0.732	-0.777	0.011	0.118***	0.153***
	(0.956)	(0.873)	(1.150)	(0.038)	(0.036)	(0.049)
Age	0.006	-0.017	0.003			
	(0.012)	(0.011)	(0.016)			
Age^2	-0.000	0.000*	0.000			
	(0.000)	(0.000)	(0.000)			
Post*Info	0.938	-1.912*	1.148	-0.011	0.008	0.070
	(1.144)	(1.071)	(1.399)	(0.048)	(0.046)	(0.059)
Post*Targets	3.319***	-0.446	1.921	0.138***	0.078	0.136**
	(1.178)	(1.073)	(1.404)	(0.053)	(0.049)	(0.061)
Post*Charity	2.979**	-1.389	2.191	-0.049	0.068	0.042
	(1.228)	(1.192)	(1.515)	(0.050)	(0.049)	(0.065)
Post*Age	0.109***	-0.024	0.032			
	(0.037)	(0.034)	(0.045)			
$Post*Age^2$	-0.001***	0.000	-0.000			
	(0.000)	(0.000)	(0.000)			
Post*Info*Age	-0.035	0.080*	-0.035			
	(0.044)	(0.042)	(0.055)			
Post*Targets*Age	-0.136***	0.019	-0.068			
	(0.046)	(0.042)	(0.054)			
Post*Charity*Age	-0.116**	0.056	-0.078			
	(0.048)	(0.046)	(0.059)			
$Post*Info*Age^2$	0.000	-0.001**	0.000			
	(0.000)	(0.000)	(0.001)			
$Post*Targets*Age^2$	0.001***	-0.000	0.001			
	(0.000)	(0.000)	(0.001)			
$Post*Charity*Age^2$	0.001**	-0.001	0.001			
	(0.000)	(0.000)	(0.001)			
SN				0.000*	-0.001***	-0.001***
				(0.000)	(0.000)	(0.000)
SN^2				-0.000	0.000***	0.000***
				(0.000)	(0.000)	(0.000)
Post*SN				-0.000	0.001	0.000
				(0.000)	(0.000)	(0.000)
$Post*SN^2$				0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)
Post*Info*SN				0.001	-0.000	-0.000
				(0.001)	(0.000)	(0.001)
Post*Targets*SN				-0.001	-0.001	-0.000
				(0.001)	(0.001)	(0.001)
Post*Charity*SN				0.001*	-0.000	-0.000
				(0.001)	(0.000)	(0.001)
$Post*Info*SN^2$				-0.000*	0.000	0.000
				(0.000)	(0.000)	(0.000)
$Post*Targets*SN^2$				0.000	0.000*	0.000
				(0.000)	(0.000)	(0.000)
$Post*Charity*SN^2$				-0.000*	0.000	0.000
				(0.000)	(0.000)	(0.000)
Constant	0.387	0.746***	0.207	0.390***	0.380***	0.409***
	(0.303)	(0.277)	(0.405)	(0.014)	(0.013)	(0.017)
N	N=339 Captains			N=339 Captains		
Obs	35,745 (22,650)			35,745 (22,650)		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 8
DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, BY SALIENCE

	ZFW	EF	RET
Post	0.022	0.166***	0.139***
	(0.018)	(0.017)	(0.021)
Salient	-0.003	0.020	-0.010
	(0.020)	(0.019)	(0.023)
Post*Info	0.002	0.023	0.079***
	(0.017)	(0.016)	(0.020)
Post*Targets	0.026	0.034**	0.107***
	(0.017)	(0.016)	(0.020)
Post*Charity	0.004	0.022	0.071***
	(0.017)	(0.016)	(0.020)
Post*Info*Salient	0.039	-0.009	-0.004
	(0.028)	(0.027)	(0.032)
Post*Targets*Salient	0.025	0.024	0.017
	(0.028)	(0.028)	(0.033)
Post*Charity*Salient	0.034	0.036	0.046
	(0.028)	(0.027)	(0.033)
Constant	0.427***	0.298***	0.366***
	(0.021)	(0.020)	(0.024)
N	N=339 Captains		
Obs	35,745 (22,650)		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Salient is a dummy variable that takes on a value of 1 if the flight took place within a week of receiving feedback, and 0 otherwise. Newey-West standard errors are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 9

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, BY RISK AVERSION

	ZFW	EF	RET	ZFW	EF	RET
Post	0.073***	0.129***	0.193***	0.057**	0.140***	0.211***
	(0.026)	(0.023)	(0.028)	(0.025)	(0.023)	(0.027)
Post*Info	-0.048	0.017	0.021	-0.033	0.025	0.004
	(0.111)	(0.027)	(0.034)	(0.030)	(0.027)	(0.032)
Post*Target	-0.002	0.066**	0.125***	0.006	0.070**	0.106***
	(0.032)	(0.029)	(0.035)	(0.031)	(0.030)	(0.032)
Post*Charity	-0.043	0.059**	0.059*	-0.033	0.064**	0.039
	(0.030)	(0.027)	(0.035)	(0.029)	(0.028)	(0.033)
Post*Risk	-0.023	0.010	-0.038**	-0.021	0.009	-0.042***
	(0.015)	(0.013)	(0.017)	(0.015)	(0.013)	(0.016)
Post*Info*Risk	0.025*	0.010	0.025	0.023	0.011	0.034
	(0.020)	(0.018)	(0.023)	(0.019)	(0.018)	(0.022)
Post*Target*Risk	0.025	0.013	0.010	0.028	0.013	0.012
	(0.020)	(0.018)	(0.023)	(0.019)	(0.018)	(0.022)
Post*Charity*Risk	0.033*	-0.023	0.008	0.028	-0.020	0.019
	(0.019)	(0.017)	(0.022)	(0.018)	(0.017)	(0.021)
Constant	0.442***	0.319***	0.361***	0.660***	0.348***	0.244***
	(0.005)	(0.005)	(0.006)	(0.026)	(0.021)	(0.032)
N	*N*=216 subjects			*N*=216 subjects		
Obs	22,691 (14,379)			22,691 (14,379)		
Controls	None			Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 10

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS FOR ZFW, EF, AND RET

	ZFW	EF	RET
Post	0.038	0.137***	0.249***
	(0.032)	(0.030)	(0.028)
Post*Info	-0.015	0.025	0.030
	(0.036)	(0.034)	(0.034)
Post*Target	0.009	0.060	0.109***
	(0.041)	(0.039)	(0.036)
Post*Charity	-0.016	0.049	-0.029
	(0.036)	(0.034)	(0.036)
Post*Risk	-0.033*	0.007	-0.018
	(0.018)	(0.016)	(0.022)
Post*LossAv	0.045	0.007	-0.091**
	(0.044)	(0.041)	(0.046)
Post*Info*Risk	0.035	0.007	0.022
	(0.027)	(0.025)	(0.032)
Post*Target*Risk	0.028	0.005	0.019
	(0.024)	(0.023)	(0.030)
Post*Charity*Risk	0.037	-0.045**	-0.050*
	(0.025)	(0.023)	(0.029)
Post*Info*Loss Av	-0.044	0.009	0.060
	(0.058)	(0.053)	(0.064)
Post*Target*LossAv	-0.004	0.025	-0.017
	(0.058)	(0.056)	(0.061)
Post*Charity*LossAv	-0.037	0.067	0.215***
	(0.056)	(0.053)	(0.061)
Constant	0.660***	0.344***	0.244***
	(0.026)	(0.021)	(0.032)
N	*N*=216 subjects		
Obs	22,691 (14,379)		
Controls	Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 11

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, BY UNCERTAINTY AVERSION

	ZFW	EF	RET	ZFW	EF	RET
Post	0.110*	0.067	0.074	0.122**	0.083	0.053***
	(0.059)	(0.050)	(0.067)	(0.060)	(0.052)	(0.067)
Post*Info	-0.005	0.060	0.118	-0.052	0.106*	0.170**
	(0.071)	(0.064)	(0.083)	(0.070)	(0.061)	(0.080)
Post*Target	-0.065	0.183***	0.337*	-0.060	0.180***	0.355***
	(0.072)	(0.064)	(0.079)	(0.071)	(0.066)	(0.077)
Post*Charity	-0.123*	0.084	0.135	-0.177**	0.089	0.194**
	(0.071)	(0.063)	(0.085)	(0.071)	(0.063)	(0.083)
Post*UncAv	-0.007	0.010*	0.007	-0.010	0.009	0.011
	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)	(0.008)
Post*Info*UncAv	-0.003	-0.011	-0.005	0.004	-0.010	-0.013
	(0.008)	(0.007)	(0.010)	(0.008)	(0.007)	(0.009)
Post*Target*UncAv	0.009	-0.014*	-0.025***	0.010	-0.013	-0.030***
	(0.009)	(0.008)	(0.010)	(0.009)	(0.008)	(0.009)
Post*Charity*UncAv	0.012	-0.006	-0.005	-0.020**	-0.006	-0.014
	(0.009)	(0.008)	(0.010)	(0.008)	(0.008)	(0.010)
Constant	0.446***	0.317***	0.374***	0.669***	0.331***	0.261***
	(0.006)	(0.005)	(0.007)	(0.028)	(0.026)	(0.034)
N	*N*=193 subjects			*N*=216 subjects		
Obs	20,321 (12,888)			19,766 (12,614)		
Controls	None			Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 12

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, BY INDECISIVENESS

	ZFW	EF	RET	ZFW	EF	RET
Post	0.064**	0.133***	0.133***	0.058**	0.142***	0.142***
	(0.028)	(0.027)	(0.037)	(0.028)	(0.027)	(0.036)
Post*Info	-0.007	-0.005	0.037	-0.001	0.005	0.032
	(0.036)	(0.034)	(0.046)	(0.036)	(0.034)	(0.044)
Post*Target	0.081*	0.011	0.094*	0.061	0.012	0.104**
	(0.047)	(0.045)	(0.054)	(0.471)	(0.046)	(0.052)
Post*Charity	-0.058*	0.056	0.142***	-0.064*	0.068*	0.141***
	(0.035)	(0.035)	(0.046)	(0.035)	(0.035)	(0.044)
Post*Indec	-0.003	0.007	-0.001	-0.006	0.006	0.002
	(0.010)	(0.009)	(0.012)	(0.010)	(0.009)	(0.011)
Post*Info*Indec	-0.007	0.007	0.014	-0.007	0.006	0.012
	(0.012)	(0.011)	(0.014)	(0.012)	(0.011)	(0.014)
Post*Target*Indec	-0.029*	0.025	0.023	-0.017	0.026	0.013
	(0.017)	(0.017)	(0.020)	(0.017)	(0.017)	(0.019)
Post*Charity*Indec	0.014	-0.011	-0.022	0.018	-0.012	-0.026*
	(0.012)	(0.012)	(0.015)	(0.012)	(0.012)	(0.015)
Constant	0.446***	0.317***	0.374***	0.668***	0.333***	0.262***
	(0.006)	(0.005)	(0.007)	(0.028)	(0.026)	(0.034)
N	*N*=193 subjects			*N*=193 subjects		
Obs	20,321 (12,888)			19,766 (12,614)		
Controls	None			Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 13

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, CONFIDENCE IN DECISION-MAKING

	ZFW	EF	RET	ZFW	EF	RET
Post	0.215***	0.274***	0.134*	0.147***	0.294***	0.230***
	(0.053)	(0.050)	(0.070)	(0.051)	(0.052)	(0.068)
Post*Info	-0.193***	-0.029	0.010	-0.144**	-0.027	-0.057
	(0.074)	(0.068)	(0.093)	(0.071)	(0.070)	(0.090)
Post*Target	-0.151**	-0.094	0.146*	-0.101	-0.087	0.043
	(0.069)	(0.063)	(0.083)	(0.067)	(0.066)	(0.081)
Post*Charity	-0.107	-0.070	0.170*	-0.057	-0.086	0.109
	(0.067)	(0.065)	(0.093)	(0.066)	(0.066)	(0.088)
Post*ConfDec	-0.019***	-0.015**	0.000	-0.013**	-0.016***	-0.010
	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)	(0.008)
Post*Info*ConfDec	0.020**	0.005	0.008	0.015*	0.005	0.015
	(0.009)	(0.008)	(0.011)	(0.008)	(0.008)	(0.011)
Post*Target*ConfDec	0.019**	0.020***	0.001	0.015*	0.020**	0.011
	(0.008)	(0.008)	(0.010)	(0.008)	(0.008)	(0.010)
Post*Charity*ConfDec	0.009	0.012	-0.010	0.004	0.015*	-0.004
	(0.008)	(0.008)	(0.011)	(0.008)	(0.008)	(0.010)
Constant	0.446***	0.317***	0.374***	0.668***	0.330***	0.261***
	(0.006)	(0.005)	(0.007)	(0.028)	(0.026)	(0.034)
N	*N*=193 subjects			*N*=193 subjects		
Obs	20,321 (12,888)			19,766 (12,614)		
Controls	None			Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 14

DIFFERENCE-IN-DIFFERENCE REGRESSION COEFFICIENTS, BY SELF-REPORTED DONATIONS IN 2013

	ZFW	EF	RET	ZFW	EF	RET
Post	0.103**	0.092**	0.009	0.110**	0.104**	0.009
	(0.049)	(0.043)	(0.057)	(0.047)	(0.044)	(0.059)
Post*Info	-0.035	-0.027	0.165**	-0.060	-0.025	0.176**
	(0.064)	(0.053)	(0.071)	(0.062)	(0.054)	(0.071)
Post*Target	-0.087	0.182***	0.332***	-0.093	0.179***	0.321***
	(0.076)	(0.065)	(0.078)	(0.073)	(0.067)	(0.078)
Post*Charity	-0.246***	-0.048	0.060	-0.256***	-0.042	0.046
	(0.056)	(0.053)	(0.072)	(0.054)	(0.055)	(0.072)
Post*Donations13	-0.006	0.007	0.015**	-0.009	0.007	0.017**
	(0.005)	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)
Post*Info*Donations13	0.002	0.005	-0.011	0.006	0.006	-0.014
	(0.008)	(0.006)	(0.009)	(0.007)	(0.007)	(0.009)
Post*Target*Donations13	0.012	-0.014*	-0.023**	0.014	-0.013	-0.024**
	(0.009)	(0.008)	(0.010)	(0.009)	(0.008)	(0.010)
Post*Charity*Donations13	0.030***	0.011*	0.005	0.032***	0.011*	0.006
	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.009)
Constant	0.448***	0.318***	0.373***	0.672***	0.335***	0.251***
	(0.006)	(0.005)	(0.007)	(0.028)	(0.027)	(0.035)
N	*N*=193 subjects			*N*=193 subjects		
Obs	20,321 (12,888)			19,766 (12,614)		
Controls	None			Weather, # Engines		

Notes: The dependent variables in these regressions indicate the proportion of flights for which a given behavior (ZFW, EF, or RET) was successfully performed. Newey-West standard errors (lag=4) are reported below estimates in parentheses. Total flight observations are provided, followed by reduced observations for RET in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 15

JOB SATISFACTION AND JOB PERFORMANCE

	Groups: Control & Info		Groups: Targets & Charity	
	Job Satisfaction		Job Satisfaction	
ZFW Targets Met	0.030	-	-0.003	-
	(0.048)		(0.059)	
EF Targets Met	-0.073	-	-0.027	-
	(0.057)		(0.056)	
RET Targets Met	0.032	-	0.129**	-
	(0.037)		(0.052)	
Overall Targets Met	-	-0.004	-	0.048*
		(0.023)		(0.025)
Constant	5.751***	5.727***	5.399***	5.386***
N	*N*=103 subjects		*N*=99 subjects	
Obs	103		99	
Controls	None		None	

Notes: The dependent variable in these regressions is a 7-point scale of job satisfaction, where self-reported job satisfaction increases in the scale. Robust standard errors are reported below estimates in parentheses. The independent variables indicate the number of targets met per behavior and overall over the course of the study. *significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 16

JOB SATISFACTION AND TREATMENT GROUP

	Job Satisfaction
TG1: Info	0.212
	(0.224)
TG2: Targets	0.242
	(0.249)
TG3: Charity	0.365*
	(0.223)
Constant	5.58***
	(0.174)
N	N=202 subjects
Obs	202
Controls	None

Notes: The dependent variable in this regression is a 7-point scale of job satisfaction, where self-reported job satisfaction increases in the scale. Robust standard errors are reported below estimates in parentheses.

APPENDIX I: EXPERIMENTAL DESIGN MATERIALS

TREATMENT GROUP DESIGN

	Information	Targets	Charity
Control			
Treatment Group 1	✓		
Treatment Group 2	✓	✓	
Treatment Group 3	✓	✓	✓

BALANCE CHECK

Groups	gender	seniority	age	trainer	trusted pilot	weekend	plan ramp fuel	number of engines	flights per month	above average PRF	above average NOE	above average FPM	ZFW	EF	RET
C, T1	1.00	0.272	0.735	0.687	0.700	0.309	0.831	0.749	0.889	0.539	0.645	1.000	0.920	0.613	0.391
C, T2	0.560	0.678	0.430	0.838	0.650	0.273	0.860	0.755	0.844	0.443	0.878	0.878	0.899	0.683	0.390
C, T3	0.560	0.962	0.918	0.551	0.650	0.879	0.719	0.898	0.637	0.167	0.759	0.759	0.960	0.512	0.603
T1, T2	0.560	0.519	0.627	0.843	0.406	0.957	0.969	0.991	0.726	0.878	0.759	0.878	0.817	0.925	0.995
T1, T3	0.560	0.299	0.653	0.846	0.406	0.219	0.892	0.851	0.516	0.443	0.878	0.759	0.960	0.880	0.779
T2, T3	1.000	0.715	0.370	0.695	1.000	0.187	0.858	0.858	0.785	0.539	0.878	0.645	0.859	0.811	0.775
Test	χ^2	t-test	t-test	χ^2	χ^2	t-test	t-test	t-test	t-test	χ^2	χ^2	χ^2	t-test	t-test	t-test

APPENDIX II: FUEL AND CARBON EFFICIENCY REPORTS

Treatment Group 1: Information



Fuel and carbon efficiency report for Capt. John Smith

Below is your monthly fuel and carbon efficiency report for **Month 2014**

1. ZERO FUEL WEIGHT	2. EFFICIENT FLIGHT	3. REDUCED ENGINE TAXY IN
Proportion of flights for which the ZFW calculation was completed and fuel load adjusted as necessary	*Proportion of flights for which actual fuel use is less than planned fuel use (e.g. optimised speed, altitude etc)*	*Proportion of flights for which at least one engine was shut off during taxy in*
RESULT: **XX% of flights**	RESULT: **XX% of flights**	RESULT: **XX% of flights**

We will continue to keep you updated on your monthly performance for the next **X months**, John.

Please see reverse side for further details of the three behaviours.

Questions? We are here to help! Please email us at project.uoc@fly.virgin.com.

All data gathered during this study will remain anonymous and confidential. Safety remains the absolute and overriding priority. This study will be carried out within Virgin's existing and highly robust safety standards, using our existing fuel procedures and policies. Captains retain full authority, as they always have done in VAA, to make decisions based on their professional judgment and experience.

Treatment Group 2: Targets

virgin atlantic

Fuel and carbon efficiency report for Capt. John Smith

Below is your monthly fuel and carbon efficiency report for **Month 2014**

1. ZERO FUEL WEIGHT

Proportion of flights for which the ZFW calculation was completed and fuel load adjusted as necessary

TARGET: **XX% of flights**

RESULT: **XX% of flights**

You ACHIEVED/MISSED your target.

2. EFFICIENT FLIGHT

Proportion of flights for which actual fuel use is less than planned fuel use (e.g. optimised speed, altitude etc)

TARGET: **XX% of flights**

RESULT: **XX% of flights**

You ACHIEVED/MISSED your target.

3. REDUCED ENGINE TAXY IN

Proportion of flights for which at least one engine was shut off during taxy in

TARGET: **XX% of flights**

RESULT: **XX% of flights**

You ACHIEVED/MISSED your target.

WHAT WAS YOUR OVERALL OUTCOME?

You achieved X of your 3 targets last month.

WELL DONE! We will continue to keep you updated on your monthly performance for the next **X months**, John.

Please continue to fly efficiently next month to achieve your targets.

Please see reverse side for further details of the three behaviours.

Questions? We are here to help! Please email us at project.uoc@fly.virgin.com.

All data gathered during this study will remain anonymous and confidential. Safety remains the absolute and overriding priority. This study will be carried out within Virgin's existing and highly robust safety standards, using our existing fuel procedures and policies. Captains retain full authority, as they always have done in VAA, to make decisions based on their professional judgment and experience.

Treatment Group 3: Charitable Incentives



Fuel and carbon efficiency report for Capt. John Smith

Below is your monthly fuel and carbon efficiency report for **Month 2014**

1. Zero Fuel Weight

Proportion of flights for which the ZFW calculation was completed and fuel load adjusted as necessary

TARGET: **XX% of flights**

RESULT: **XX% of flights**

You ACHIEVED/MISSED your target and earned/missed out on £10 in donations to Charity Name.

2. Efficient Flight

Proportion of flights for which actual fuel use is less than planned fuel use (e.g. optimised speed, altitude etc)

TARGET: **XX% of flights**

RESULT: **XX% of flights**

You ACHIEVED/MISSED your target and earned/missed out on £10 in donations to Charity Name.

3. Reduced Engine Taxy In

Proportion of flights for which at least one engine was shut off during taxy in

TARGET: **XX% of flights**

RESULT: **XX% of flights**

You ACHIEVED/MISSED your target and earned/missed out on £10 in donations to Charity Name.

What Was Your Overall Outcome?

Due to your fuel and carbon efficient decision making last month, you achieved X of your 3 targets and secured £XX of a possible £30 for your chosen charity, Charity Name.

WELL DONE! For the next **X months**, you still have the ability to donate **£X to Charity Name.** Please continue to fly efficiently next month to achieve your targets so your charity does not lose out.

Please see reverse side for further details of the three behaviours.

Questions? We are here to help! Please email us at project.uoc@fly.virgin.com.

All data gathered during this study will remain anonymous and confidential. Safety remains the absolute and overriding priority. This study will be carried out within Virgin's existing and highly robust safety standards, using our existing fuel procedures and policies. Captains retain full authority, as they always have done in VAA, to make decisions based on their professional judgment and experience.

All Treatment Groups: Reverse Side of Report

THE THREE BEHAVIOURS WE ARE MEASURING

Behaviour 1: Zero Fuel Weight Adjustment (ZFW) - Pre Flight

This measure compares Actual Ramp against Plan Ramp adjusted for changes in ZFW. It captures whether a double iteration adjustment has been implemented for ZFW in line with Plan Burn Adjustment and any further amendments to flight plan fuel that have been entered into ACARS. This behaviour has a tolerance of 200kg, which ensures that rounding in the fuel request / loading procedure will not adversely affect the result.

Behaviour 2: Efficient Flight (EF) - During Flight

This measure examines the actual fuel burn per minute compared against the expected fuel burn per minute from OFF to ON (expected fuel burn is Plan Trip adjusted for ZFW). It highlights pilot technique (e.g. optimum settings are realised, optimum levels are sought, speed is optimised, etc.).

Behaviour 3: Reduced Engine Taxy In (RETI) - Post Flight

This measure observes if an engine has been shut down during taxy in. RETI is considered to have taken place if one engine burns less than 70% of the average of other engines during taxy in. If taxy in is shorter than the cool down period required, the flight is omitted, as RETI was not possible.

We hope the above information is beneficial to you. If you require more information about any of the behaviours, please email us at project.uoc@fly.virgin.com.

APPENDIX III: SURVEY MATERIALS

Risk

Below you will be asked to select from among five different gambles the one gamble you would like to play. The five different gambles are listed below.

Each gamble has two possible outcomes, both of which are associated with a 50% probability of payout (i.e. a fair coin toss).

Please note that if you should select either Gamble 4 or Gamble 5 and you incur a loss (i.e. -£5 or -£15), your losses will be deducted from your £25 compensation for completing the survey (e.g., £25 - £5 = £20).

Please select your preferred 50/50 gamble.

- ○ **Gamble 1:** £25 vs. £25
- ○ **Gamble 2:** £45 vs. £15
- ○ **Gamble 3:** £65 vs. £5
- ○ **Gamble 4:** £85 vs. -£5
- ○ **Gamble 5:** £105 vs. -£15

Below you will find a gamble with the ***same outcomes*** but with ***different probabilities*** of payout.

Each gamble has two possible outcomes, the **first of which is associated with a 75% probability of payout** and the **second of which is associated with a 25% probability of payout**.

Please note that if you should select either Gamble 4 or Gamble 5 and you incur a loss (i.e. -£5 or -£15), your losses will be deducted from your £25 compensation for completing the survey (e.g., £25 - £5 = £20).

Please select your preferred 75/25 gamble.

- ○ **Gamble 1:** £25 vs. £25
- ○ **Gamble 2:** £45 vs. £15
- ○ **Gamble 3:** £65 vs. £5
- ○ **Gamble 4:** £85 vs. -£5
- ○ **Gamble 5:** £105 vs. -£15

Below you will find a gamble with the ***same outcomes*** but with ***different probabilities*** of payout.

Each gamble has two possible outcomes, the **first of which is associated with a 25% probability of payout** and the **second of which is associated with a 75% probability of payout**.

Please note that if you should select either Gamble 4 or Gamble 5 and you incur a loss (i.e. -£5 or -£15), your losses will be deducted from your £25 compensation for completing the survey (e.g., £25 - £5 = £20).

Please select your preferred 25/75 gamble.

- ○ **Gamble 1:** £25 vs. £25
- ○ **Gamble 2:** £45 vs. £15
- ○ **Gamble 3:** £65 vs. £5
- ○ **Gamble 4:** £85 vs. -£5
- ○ **Gamble 5:** £105 vs. -£15

GAMBLES, CRRA RANGES, AND RISK COEFFICIENTS

	Probabilities	Event A	Event B	CRRA Range	selected *r*
Gamble 1	50/50	£25	£25	$r>2$	$r=2.5$
Gamble 2	50/50	£45	£15	$0.67<r<2$	$r=1.34$
Gamble 3	50/50	£65	£5	$0.38<r<0.67$	$r=0.53$
Gamble 4	50/50	£85	-£5	$0.20<r<0.38$	$r=0.29$
Gamble 5	50/50	£105	-£15	$r<0.20$	$r=0$

Ambiguity & Decisiveness

For the following three statements, please rate on a scale from 0 to 10 the extent to which you agree with each statement.

I don't like situations that are uncertain.

Strongly Disagree										*Strongly Agree*
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

I would describe myself as indecisive.

Strongly Disagree										*Strongly Agree*
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

I usually make important decisions quickly and confidently.

Strongly Disagree										*Strongly Agree*
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Altruism

To the best of your knowledge, how much (if any) did you donate to charity in **2013**?

- £0 (I did not donate to charity in 2013.)
- £1 - £10
- £11 - £20
- £21 - £30
- £31 - £40
- £41 - £50
- £50 - £75
- £75 - £100
- £100 - £200
- more than £200

How much (if any) did you donate to charity in **2014 outside of the context of the study**?

- £0 (I did not donate to charity in 2014 outside of the context of the study.)
- £1 - £10
- £11 - £20
- £21 - £30
- £31 - £40
- £41 - £50
- £50 - £75
- £75 - £100
- £100 - £200
- more than £200

Job Satisfaction

All things considered, how **satisfied** are you with...

	Not at all satisfied 1	2	3	4	5	6	Extremely Satisfied 7
...your present job overall?	○	○	○	○	○	○	○