

# The Demand and Supply of Esteem: An experimental analysis\*

**Paul Blacklow**

Tasmanian School of  
Business and Economics  
University of Tasmania  
Hobart 7000 TAS  
Australia

[Paul.Blacklow@utas.edu.au](mailto:Paul.Blacklow@utas.edu.au)

**Amy Beth Corman**

Department of Economics  
University of Melbourne  
Parkville 3010 VIC  
Australia

[abcorman@unimelb.edu.au](mailto:abcorman@unimelb.edu.au)

**Hugh Sibly**

Tasmanian School of  
Business and Economics  
University of Tasmania  
Hobart 7000 TAS  
Australia

[Hugh.Sibly@utas.edu.au](mailto:Hugh.Sibly@utas.edu.au)

April 2020

## **Abstract**

We use a laboratory experiment to identify some determinants of variation across individuals in the demand and supply of esteem. Some participants can choose, at a cost, buy 'credits' so as to appear to others to have a better performance on a real effort task than they really do: this purchase reflects esteem seeking. The provision of esteem, via a partner's feedback on performance, is also recorded. We find: (i) those with a low willingness to supply effort have a lower probability of purchasing credits, (ii) participants matched with low scoring partners buy more credits (iii) feedback is improved for those purchasing credits.

**JEL Codes: C91 D91**

**Keywords: Esteem, Image, Laboratory Experiment**

---

\* We would like to thank Tim Cason, Yan Chen, Lata Gangadharan, Jim Murphy, Katya Sherstyuk, John Tisdell and Tom Wilkening for comments on earlier versions of this work. Thanks are also due to the editor and referees, whose advice improved the paper. This work was conducted using a grant from the Tasmanian School of Business and Economics.

# 1 Introduction

Casual empiricism suggests that, in a wide range of domains, people judge the actions of others. Furthermore, people care about these judgements, and consequently take actions that will generate favourable opinions. For instance, people tend to buy more fashionable or different style of clothes than required for utilitarian purposes in the anticipation of receiving praise (or avoiding disapproval). Such concerns are not misplaced: it is also common for people to form opinions and to comment on other's clothing choice. Given the multitude of other similar examples<sup>1</sup>, this type of judgement and approval seeking behaviour (or disapproval avoidance) appears to be ubiquitous. The aim of this paper is to provide an experimental study of this type of behaviour, which we refer to as esteem seeking when individuals undertake actions designed to gain approval and esteem provision when they grant such approval.

The recent literature also points to the presence of esteem or status seeking behaviour (Tran and Zeckhauser, 2012; Charness et. al., 2014; Dwenger et. al. 2016; Riedl et. al., 2018; Burstyn, et. al., 2018; Kirchler and Palan, 2018; Friedrichsen and Engelmann, 2018). Although apparently ubiquitous, the strength of esteem seeking, and esteem provision, appears to differ across individuals. For example, some people are less concerned about clothing choice, and some are less judgemental about other's clothing choice. Such differences may be partly due to individual's characteristics but may also be partly due to social influences.

In this paper we report the results of a laboratory experiment designed to identify some of the key determinants of individuals' demand and supply of esteem. The design of the experiment is intended to detect key determinants of the type of esteem seeking, and esteem provision, that is described in the clothing example. Importantly we are concerned with those actions designed to improve others' opinion of oneself, when there is no public good dimension to those actions.<sup>2</sup> We concurrently investigate if and when such self-benefiting actions are effective in generating esteem.

Our experiment allows some participants to esteem seek through influencing the reporting of their performance on a real effort task. To create the potential, and to measure, the demand for esteem, we follow Charness et. al. (2014) in allowing some participants in 'matched stages' of our experiments to buy costly 'credits'. Doing so allows these participants to appear to another randomly matched participant, to have a better performance on the real effort task than they actually had. To mirror the clothing example

---

<sup>1</sup> For example, in recent years many people have spent considerable time on social media taking actions designed to garner positive reactions from others. Similarly, people may buy a more luxurious or stylish car than they would want for purely transportation purposes. Or they may stay longer each day at their job than is necessary to achieve the required work.

<sup>2</sup> As esteem, in our definition, is driven by utility derived from the thoughts of others, it is distinct from self-esteem, which is utility driven by one's own internally driven thoughts. Self-esteem is analysed in an economic framework by Benabou and Tirole (2003).

discussed above, the matched participant can not identify how much of the participant's score they view is due to effort or due to the purchase of credits.

A person's esteem seeking behaviour may be influenced by their personal characteristics and/or their social environment. We find that participants who, in the absence of social comparisons, exhibit a low willingness to supply effort (and thus to earn income) are the ones more likely to buy credits. This finding is consistent with the related study of Friedrichsen and Engelmann (2018), which is discussed in greater detail later in the paper. Gender is also found to influence the purchase of credits in a way consistent with the findings of the literature on competitiveness (Niederle and Vesterlund, 2007).

If people undertake social comparisons in valuing esteem, then a key social determinant of the willingness to buy credits will be the performance of the people they interact with. To observe this effect, our experiment is designed to ensure that each participant interacts with both a high and low performer. We find that participants are more likely to purchase credits if their matched participant is a low performer.

To quantify this potential supply of esteem, we also elicit feedback on a participant's performance from their matched participant. We observe that the positiveness of feedback increases with the score observed by the matched participant. This suggests that the purchase of credits does indeed buy a participant esteem. This is the case even though the matched participant knows that the score they observe could be inflated by the purchase of credits. We also find that gender influence the nature of feedback.

We are interested in identifying individual's differences in their propensity to esteem seek because, if present, they could have important implications. For example, if people use social comparisons in evaluating esteem (Festinger, 1954; Tesser and Campbell, 1980, 1982; Frank 1985; Bernheim, 1994; Akerlof 2017), there will be an externality associated with esteem seeking: one person's esteem seeking will raise the marginal benefit of esteem seeking for another. The presence of such an externality can make market outcomes inefficient. In particular, Frank (2005) argues that if people undertake rank-based social comparisons, a positional externality is created which can be efficiently ameliorated by an appropriate tax. However, if there is a different propensity to esteem-seek across individuals, a tax on esteem-seeking activities could have a differential impact on welfare beyond the readily quantifiable financial effects.

For example, Charles et. al. (2009) find that variation in conspicuous consumption across racial groups can be attributed largely to differences in income, with low income groups more likely to engage in conspicuous consumption. If the poor are more likely to engage in esteem seeking through fashionable clothing choice than the rich, then the impact of a new tax on clothing would, by reducing the esteem seeking externality, be somewhat ameliorated

for the poor. Indeed, in this example, the tax will cause the poor to consume more efficiently, while the rich would under consume.<sup>3</sup> As illustrated by this example, identifying differences in the propensity to esteem seek is important as these differences can have both distributional and efficiency implications in response to policy actions.

In our experiment the purchase of credits only raises a participant's score, and provides no benefit to anyone else. In contrast, much of the original theoretical interest in esteem-seeking stems from how it might be involved in the private provision of public goods. Brennan and Pettit (2004) argue that an individual's concern for esteem is fundamental to the functioning of the economy and emphasise the role esteem plays in decisions in all aspects of life. Subsequently, Benabou and Tirole (2006) develop a formal signalling model of a market consisting of individuals who are able to gain esteem by engaging in a pro-social activity. Provision of the public good acts as a signal to an individual's intrinsic pro-sociality. Ellingsen and Johannesson (2008) utilise a signalling model of esteem to explain a number of experimental findings arising from esteem seeking through the provision of public goods.

Other authors have studied theoretically the role of esteem seeking in relation to private actions (those without a public good dimension). Cowen and Glazer (2006) analyse how some labour market institutions arise in order to ensure that poor workers avoid outcomes that would result in them receiving substantial disesteem. More recently, Akerlof (2017) developed a model using esteem to explain if and how individuals coordinate their values of tasks. He finds that the heterogeneity amongst people, specifically differences in ability at tasks, plays a critical role in the nature of esteem seeking. His findings illustrate how the heterogeneity in a population's characteristics could potentially play an important role in the manifestation of esteem seeking.

Friedrichsen and Engelmann (2018) present a lab experiment of social image based on the public versus private purchase of fair-trade chocolate over conventional chocolate. Their paper differs in three key respects from ours. First, the purchase of fair-trade chocolate is interpreted as a pro-social activity, whereas our experiment involves esteem seeking that does not have a pro-social dimension. The only advantage to our participants of buying credits is that it allows them to be perceived by their matched partner as a high performer. Second, Friedrichsen and Engelmann focus on esteem seeking from public/group disclosure decision whereas we focus on both the esteem seeking and its supply between two anonymous individuals.

Thirdly, Friedrichsen and Engelmann's experiment is framed in the specific context of ordinary versus fair-trade chocolate. In contrast our experiment is relatively context free:

---

<sup>3</sup> Frank (1997) advocates a consumption tax be progressive, for the usual equity reasons. However a progressivity (or regressivity) of consumption tax could, in principle, also be adjusted to take account of differences in the propensity to esteem seek across income groups.

participants undertake an intrinsically pointless real effort task, which itself is designed to minimise the impact of pre-existing skill or knowledge on performance (the slider task by Gill and Prowse, 2012). The only purpose to the participant of buying credits is, by raising their score, to create the impression they are a high performer. The motivation for this design choice is that, by observing esteem seeking and provision under these unfavourable circumstances, the underlying determinants of this behaviour in a wide variety of contexts can be identified.

Friedrichsen and Engelmann's key finding is that those "with low intrinsic motivation to buy fair trade react positively to image building opportunities, while those with high intrinsic motivation do not" (p. 74). They attribute this finding to the nature of psychological preferences in their subjects, namely a negative correlation between individuals' intrinsic motivation to provide public goods and their utility from social image. We compare our findings to theirs in the conclusion.

Several previous experimental papers have looked for verification of status seeking. Status, which represents a person's relative social ranking, is an important avenue through which people can receive esteem. Charness et. al., (2010, 2014) examine, both experimentally and theoretically, status seeking in a setting where participants are paid using piece rates. Repeated status contests within groups of three are conducted, where there are explicit winners and losers from each contest. They find that their subjects increase effort when informed about their relative performance and, importantly, pay to increase their apparent performance.

Like Charness et. al. (2010, 2014), we allow participants the opportunity (by purchasing 'credits') to boost their apparent performance ('score'). We use pairwise interactions between participants to identify the determinants of their conferring and receiving of esteem. We allow esteem-seeking to endogenously arise rather than construct status within groups. Not providing an experimenter determined signal of status might be expected to lower participant benefit from esteem seeking, resulting in less of this activity. Again, this aspect of our design makes the experimental test for the presence of esteem seeking activities more stringent. Nonetheless, like Charness et. al., (2010, 2014) we observe that some participants will pay to boost their apparent performance. However we do not find strong evidence that participants increase their effort when their score is made public.

Other previous experimental studies, however, show an effect on behaviour of exposing participants to measures of their relative performance. Azmat and Iriberry (2010, 2016) find that informing subjects of their relative performance on a real effort task increases their performance. Gill et. al. (2017) report an experiment that assesses the impact of making participants' performance rank on a real effort task observable on their subsequent effort. They find participants increase effort to retain a first place and avoid retaining last place.

Several field experiments have shown that status and esteem seeking motivate individuals in real world settings. Tran and Zeckhauser (2012) show communicating rank in a test increased student performance. Burstyn, et al (2018) demonstrate Platinum (high status) credit cards are used for social signalling. Kirchler and Palan (2018) find sales staff respond positively to either compliments or monetary tips. Unlike these field studies, our lab experiment is relatively context free. In spite of this, we find some individuals esteem seek. This indicates that that esteem seeking behaviour is likely to be ubiquitous, and not necessarily reliant on the presence of institutional structures that confer status.

## 2 The Experiment

### 2.1 Experimental Design

Our experiment's treatments are designed to assess how the propensity to esteem seek (buy credits) is influenced by the strength of a participant's preference to earn monetary income by supplying effort (a personal characteristic) and whether the participant is interacting with high or low performing individuals (the social environment). To this end, the experiment was conducted in five stages.

The first activity undertaken by participants in each of the five stages is a real effort task: the well-known 'sliders' real effort task developed by Gill and Prowse (2012). This task has the advantage that it is designed to minimise the role of skill in undertaking the task, and as such could be considered a task that reflects effort more than skill. Nonetheless, some dimensions of participants innate skill may play a role in determining performance on the task. For example, some participants may have superior hand-eye coordination which makes placing sliders relatively easy for them compared to their peers. A participant's relative performance would thus reflect either their exertion on the task or their skill level.

Stages 1, 3 and 5 of the experiment differ to stages 2 and 4. Stages 1, 3 and 5 were designed to identify a participant's willingness to supply effort in the absence of social influences. In each of these 3 stages all participants undertook only the slider task. Participants were paid a piece rate of 1 Exp\$ for each slider correctly placed, where 1 Exp\$ is equal to \$0.10 Australian dollars. In each of these 3 stages participants are only aware of their own effort (number of correctly placed sliders) and income, and not that of any other participant.

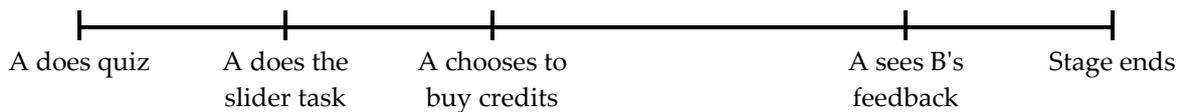
We are interested in how people's esteem seeking activities respond to the social environments in which they interact, particular with respect to the relative performance of their peers. Stages 2 and 4 of our experiment are designed to identify such effects. The design utilises a random allocation of all participants in our experiment into one of two equal-sized groups: A and B, whose members are called role A and role B participants

respectively. This allocation is made prior to the beginning of the experiment and does not change throughout the experiment.

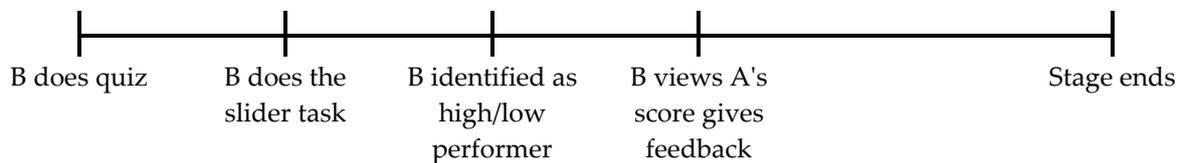
In stages 2 and 4, each role A participant was randomly paired with a role B participant. Figure 1 summarises the order of events in stages 2 and 4 for each matched pair of participants. This timeline is now discussed in detail.

**Figure 1: Order of events in stages 2 and 4**

Role A Participant



Role B Participant



After undertaking the real effort task in stages 2 and 4, role B participants were divided into two distinct groups: high and low performers. Those role B participants who scored in the top half of performers on the real effort task in that stage – the high performers - were given a score and payment of 40. Those in the lower half – the low performers - were given a score and payment of 10. This design was adopted so that, in stages 2 and 4, all role A participants were facing the same signal when matched with a high performing role B participant and also when matched with a low scoring role B participant. In effect, all role A participants face the same social setting when matched with their high and low performing partners. We are therefore able to directly compare esteem seeking responses across role A participants to the performance level of their matched role B participant.

If a role A participant was matched with a high scoring role B participant in stage 2 they would be matched with a different low scoring role B participant in stage 4, and vice-versa. Role A participants were informed that they would be matched with high or low scoring role B participant in stages 2 and 4 prior to commencing the stage.

In stages 2 and 4, role A participants were paid a piece rate for performance on the slider task, as in stages 1, 3 and 5. However, in contrast to stages 1, 3 and 5, after completion of the slider task the role A participants were given the opportunity to purchase credits. Each credit costs the participant 1 Exp\$, and they could spend up to the total amount that they had earned in that stage. The credits increased the participants' score and reduced their income as follows:

Score = the number of sliders correctly placed + any credits bought

Payment = the number of sliders correctly placed - any credits bought

Role B participants were shown the score of their matched role A participant. They were, however, not shown the number of sliders placed or the numbers of credits bought. The role B participant was asked to give feedback on their matched role A participant's score on a 5-point Likert scale ranging from very bad to very good.

The matching of each role A participant with both a high and low scoring role B participant was designed to test whether and how the different performance of the matched participant influences the participant's choice of effort/income and credits purchased.

After the fifth stage was completed, participants were asked a set of questions related to the outcome in stages 2 and 4. Participants in role A were asked, on a 5-point Likert scale and for each of stages 2 and 4 (i) whether the prospect of feedback increased their effort, (ii) whether the prospect of feedback caused them to buy credits, and (iii) how fair the feedback they received was. Participants in role B were asked on a 5-point Likert scale for both stages 2 and 4: (i) whether the prospect of giving feedback caused them to increase their effort, and (ii) whether they thought their feedback was fair.

At the conclusion of the experiment all participants were then asked a series of demographic questions (age, gender, citizenship, ethnicity, enrolment degree) and some questions about their psychology. Specifically, participants were asked to answer three questions included in the Self-Assessment Manikin (SAM) developed by Lang (1980); the big 5 personality traits; locus of control (John and Srivastava, 1999) and risk attitude.<sup>4</sup> This post experimental information is not utilised in the analysis in the paper. Blacklow, Corman and Sibly (2018) provide a discussion of the impact of these demographic and psychological variables on esteem seeking.

## 2.2 Experimental Procedures

The experimental sessions lasted approximately, on average, 1.5 hours. The experiment was conducted in the Experimental Economics Laboratory at the University of Melbourne between November 2015 and April 2016 using z-Tree (Fischbacher, 2007). The 132 participants were students from the University of Melbourne recruited randomly using ORSEE (Greiner, 2004) from a pool of more than 6000 volunteers. Subjects had a variety of academic backgrounds.

---

<sup>4</sup>These questions are drawn from those used for the German Socio-Economic Panel questionnaire in 2005, available at: <http://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/GSOEP.aspx>

Upon arrival at the laboratory, participants were seated at a partitioned computer terminal. Individuals were then provided a set of instructions, both printed and on the computer screen, which explained in detail the experiment. The instructions are provided in Appendix A of this paper. After reading the instructions, each participant answered a quiz on the structure of the experiments which aimed to help them understand how earnings were determined in the experiment. Participants were then given a practice round of the sliders task to ensure they were familiar with it.

Prior to each stage of the experiment, participants were provided with a review of the instructions that were relevant for that stage, and undertook a quiz related to the conduct of that stage. The slider task in each of the five stages of the experiment (and the practice stage) lasted four minutes.

### 2.3 Conjectures

In this section we present three conjectures to be tested with the data from our experiment. We also provide an informal economic motivation for the adoption of these conjectures. Their derivation from a formal economic model is presented in Blacklow et. al. (2018).

A key aim of our experiment is to identify how the social environment affects esteem seeking. Our within-subject treatments are designed to assess whether people respond differently to high and low performing environments: specifically, whether their esteem seeking in our experiment is influenced by being matched with high or low scoring role B participants. We would expect that a role A participant has an increased propensity to buy credits when they are more sensitive to the esteem given by their matched role B partner. However, a priori, we do not know if Role A participants are more sensitive to esteem when matched with high performing role B participants compared to low performing ones. Thus we adopt the following conjecture:

**Conjecture 1:** The number of credits purchased by role A participant is not affected by the type of their matched role B participant.

This conjecture thus addresses one of the key issues of the paper: whether the status of people an individual associate with, affects that individual's esteem seeking.

We define a participant's lone effort as their average effort over stages 1, 3, and 5. Lone effort provides an estimate of the effort a participant would undertake in the absence of any social influences. Lone effort reflects a key individual characteristic: their willingness to supply effort, specifically on the slider task. An individual's willingness to supply effort will be larger for a high marginal utility of income and/or a low marginal cost of providing effort on the slider task. Those with a relatively low marginal utility of income will be less concerned with 'the cost' of buying credits than those with a high marginal utility of income, and will

be more inclined to increase esteem by buying more credits than increasing effort. Similarly, those who have a high marginal cost of effort will be more inclined to esteem seek through buying credits than supplying relatively costly effort. Thus we hypothesise that those with a lower willingness to undertake effort, as measured by lone effort, will be associated with a higher propensity to purchase credits:

**Conjecture 2:** Those role A participants with low lone effort are more likely to buy credits in stages 2 and 4.

Conjecture 2 also parallels the empirical findings of Charles et. al. (2009) that it is the low-income individuals who will engage in esteem seeking (conspicuous consumption in their study). Note, however, that there are plausible reasons that Conjecture 2 might not hold in the context of our experiment. For example, those with a high lone effort may feel sufficiently wealthy that they could 'afford' to engage in esteem seeking by buying credits.

Now consider the feedback which is provided by Role B participants to role A participants. To formulate a straightforward testable hypothesis, assume that participants believe that a higher score deserves better feedback. If participants exhibit lying aversion (Mazar, Amir, and Ariely, 2008; Fischbacher and Föllmi-Heusi, 2013) then Role B participants would achieve higher utility by providing accurate feedback. Thus as the role A participant's score increases, role B participants will provide better feedback as they will not want a substantial gap between their feedback and the deserved feedback.

Both matched participants are asked about the fairness of the feedback provided by the role B participant. Again, to formulate a straightforward testable hypothesis, suppose that all participants have a common understanding on the deserved feedback for a given score. In this event, by comparing the feedback with the commonly-understood deserved feedback, both the role A and role B participant will come to similar judgements about the fairness of the feedback. Thus we conjecture:

**Conjecture 3:** (i) Role B participants and their matched role A role A participant agree on the fairness of the role B participant's feedback. (ii) The feedback role B participants give is related to their matched role A participant's score.

If Conjecture 3(ii) holds then role B participants do not dismiss score as a signal of performance. In particular they would not be completely discounting a participant's score in their assessment, even though it could be substantially inflated due to credit purchases.

### 3 Results

#### 3.1 Credits, Effort and Score

In this section we present results of the experiment that relate directly to participant payments: first those relating to the credits purchased, then effort and finally score. Table 1 shows the proportion of participants who bought credits (Buy) and the average credits purchased (Credits) disaggregated by those that do and do not buy credits. The table is presented by stage (stage 2 and 4) and gender. The data by stage is further disaggregated according to whether the role A participants were matched with a high or low role B participant in stage 2. Recall that role A participants who have a high scoring partner in stage 2 are matched with a low scoring participant in stage 4, and vice versa. The two rows labelled “S2: Hi S4: Lo” and “S2: Lo S4: Hi” thus each represent two distinct halves of role A participants.

**Table 1 Credits by Treatment, Stage, Partner Performance and Gender**

	Treatment	Stage			Partner Type	Gender		Total
		2	4	Average		F	M	
Buy	S2: Hi S4: Lo	30.3%	24.2%	27.3%	High	20.5%	22.2%	21.2%
n		(8.1%)	(7.6%)	(7.2%)		(6.6%)	(8.2%)	(5.1%)
Credits		33	33	66		39	27	66
		+2.8‡	8.7‡	5.4		+3.0	+4.2	**3.5
n		(0.90)	(4.51)	(2.30)		(0.07)	(0.08)	(0.80)
		10	8	18		8	6	14
Buy	S2: Lo S4: Hi	21.2%^	12.1%^	16.7%	Low	17.9%	29.6%	22.7%
n		(7.2%)	(5.8%)	(5.6%)		(6.2%)	(9.0%)	(5.2%)
Credits		33	33	66		39	27	66
		+7.0	5.3	6.4		+10.4	+5.8	**7.9
n		(1.75)	(1.44)	(1.05)		(0.06)	(0.09)	(2.47)
		7	4	11		7	8	15
Buy	Total	25.8%^	18.2%^	22.0%	Total	19.2%	25.9%	22.0%
n		(5.4%)	(4.8%)	(4.6%)		(5.7%)	(7.7%)	(4.6%)
Credits		66	66	132		78	54	132
		4.5	7.6	5.8		6.5	5.1	5.8
n		(1.00)	(3.01)	(1.47)		(2.66)	(1.25)	(1.47)
		17	12	29		15	14	29

Figures in parentheses ( ) are standard errors clustered by participant. The \*\*, ^ and + denote statistically lower/higher at the 5% significance level and \*, +, †, and ‡ at the 10% significance level. Symbols before estimates, indicate testing over rows, the figures after the estimates indicate testing across columns.

On average 22% of role A participants esteem seek by buying credits. This result is similar in magnitude to the 15% of subjects in the experiment by Charness et. al., (2014) who paid to increase their performance measure. There is a decline in the proportion of participants who purchase credits with rounds, with the proportion in stage 4 being 18.2%, significantly less

than the 25.8% in stage 2 ( $p=0.048$ ).<sup>5</sup> Thus some participants had a reduced willingness to esteem seek through the purchase of credits as the experiment proceeded.

The right-hand column of Table 1 indicates that there is not a significant difference in the proportion who buy credits when matched with a high or low performing role B participant. However, the average credits purchased by those role A participants who purchase them is significantly different when matched with a high performing role B participant compared to a low performing participant ( $p=0.082$ ). Specifically, those who purchase credits buy on average 7.9 when matched with a low performing role B participant, which is significantly greater than the average of 3.5 credits when matched with a high performer ( $p=0.041$ ). Furthermore, the average credits purchased by those who purchased them in the S2:Hi S4:Lo group in stage 4 (when matched with a low performer) is greater than in stage 2 (when matched with a high performer) ( $p=0.098$ ).

These findings suggest that those role A participants who purchase credits, buy more credits when matched with low performing partners and they use this purchase to boost their score beyond that of those who do not buy credits. Thus Conjecture 1 is not supported:

**Observation 1:** If they buy credits, role A participants buy more credits when matched with low performing role B participants.

Observation 1 addresses a key aim of our analysis: whether the social environment affects esteem seeking. It suggests that the social environment does indeed matter, and that people more extensively esteem seek in an environment when others are low performing.

Table 1 also disaggregates the credit purchase decision by gender. None of the gender differences reported in the table are statistically significant. For instance, the proportion of females who bought credits is less than males, but not significantly so ( $p=0.242$ ). This relationship is also found when disaggregating the data by the type of matched B participant. When females did buy credits, on average, they bought more than males though not significantly more ( $p=0.316$ ).

There is a possibility that participants' purchase of credits is not driven by utility originating from the thoughts/judgements/opinions of others (i.e. esteem) but rather driven entirely by utility originating from their own feelings and judgements (which are independent of esteem). For example, credit purchase might be driven by a desire for domination (Charness and Grosskopf, 2001; Abbink and Sadrieh, 2009) or the pure joy of out-performing others (Dohmen et al. 2011). To address these concerns, in the post-experiment survey role A participants were asked, using a Likert scale, whether knowing they would receive feedback

---

<sup>5</sup> All tests of "difference" are two tailed  $t$ -tests with unequal variances and all tests of "lower/higher" are one-tailed  $t$ -tests with unequal variances, both using standard errors clustered by participant.

on their score caused them to buy credits in stages 2 and 4. In Table 9 in Appendix B we calculate the Spearman rank correlation between the decision to buy credits and the answer to this post-experiment survey question. We find a statistically significant correlation between the decision to purchase credits and the tendency to answer in the direction of strongly agree to this question. This provides support for the interpretation that the decision to purchase credits is, at least in part, designed to elicit esteem.<sup>6</sup>

Define a participants' partnered effort as their average effort in stages 2 and 4. Table 2 compares participants' average lone effort with their average partnered effort disaggregated in two ways. First it is divided by those who bought credits in either stage 2 or stage 4 or both, and second, by those who did not ever buy credits. The average partnered effort is significantly lower amongst those who bought credits than those who do not ( $p=0.002$ ). This is consistent with Conjecture 2. Note that the average partnered effort is not significantly different to lone effort. Thus, in contrast to the findings of the literature cited in the introduction, we do not observe participants raising their effort to improve their score (and consequently feedback/esteem) in partnered rounds. Similarly, there is no evidence that role A participants are trying to signal they are 'too cool' for the real effort task by deliberately performing badly.

Table 2 also compares the average lone effort with average partnered effort divided by gender. Males have a significantly higher average lone and partnered effort than females both  $p<0.001$ ). However, both males and females show no significant difference between their lone and partnered effort.

**Table 2 Average Effort of Role A Participant's Who Bought and Did Not Buy Credits**

	Not Buy	Buy	Female	Male	Total
Lone Effort Stages 1, 3, and 5	38.9*** (0.80)	34.3*** (1.28)	35.7+++ (0.80)	40.3+++ (1.16)	37.6 (0.72)
Partnered Effort Stages 2 and 4	38.4^^ (0.89)	34.2^^ (1.04)	35.1+++ (0.79)	40.2+++ (1.17)	37.2 (0.73)
n	94	38	78	54	132

Figures in parentheses ( ) are standard errors clustered by participant. The \*\*\*, ^^, +++ and +++ denote statistically lower/higher at the 1% level of significance.

Table 3 shows the average effort of both role A and role B participants across all stages. Note that there is a tendency for average effort to increase across stages, particularly for Role A participants whose average effort is significantly higher in stage 4 than stage 2.

---

<sup>6</sup> Note that even if esteem is only part of the reason participants purchase credits, all that is required for our statistical analysis to be valid is that the purchase of credits is related to the utility of esteem.

**Table 3 Average Effort by Role and Stage**

Stage:	1	2	3	4	5
Role A: All	34.5 (0.78) 66	35.8 <sup>^^</sup> (0.77) 66	*38.0 (0.82) 66	**38.5 <sup>^^</sup> (0.81) 66	*40.2 (0.74) 66
Role A: Not Buy/Buy		***37.0 <sup>^^</sup> (0.92) 49	***32.5 <sup>++</sup> (1.07) 17	38.9 <sup>^^</sup> (0.92) 54	36.8 <sup>++</sup> (1.63) 12
Role B	33.7 (0.84) 66	34.7 (0.79) 66	*35.8 (1.12) 66	**36.0 (1.14) 66	*38.1 (1.08) 66

Figures in parentheses ( ) are standard errors clustered by participant. The figures directly below the standard errors are the number of observations. The <sup>^^</sup> and <sup>\*\*\*</sup> denote statistically lower/higher at the 1% level of significance, <sup>\*\*</sup>, <sup>^^</sup>, <sup>++</sup> at 5% and <sup>\*</sup> at the 10% level of significance. Symbols before estimates, indicate testing over rows or within columns, the figures after the estimates indicate testing across columns.

Table 2 indicates that lone effort is related to the credit purchase decision, but that gender is also related to lone effort. Thus, gender may both directly and indirectly affect the purchase decision. Table 3 shows that effort varies over stage, particularly amongst buyers, so participants' credit purchase decision may differ between stages 2 and 4. To disentangle these effects we estimate four probit models for all 66 role A participants' decision to buy credits over the two partnered stages. We start with model (1) which only includes dummy variables for their partner type (Low), Stage 4, and whether in stage 2 participants had a high performing partner (S2:Hi, S4:Lo). Model (2) includes role A participants' lone effort in addition to the treatment and stage dummy variables from (1). The third model (3) includes their gender (Female), their age (Age) as additional regressors. The fourth model (4) we estimate is a random effects (RE) probit to allow for additional heterogeneity in the error term for participants.

Table 4 provides the marginal effects and estimation details of the four probit models for role A participants' choice to buy credits (the estimated coefficients are contained in Table 10 in Appendix B). The estimation results of the first model (1) demonstrate that all treatment and stage dummy variables are all statistically insignificant and do not help explain the decision to buy credits. The inclusion of lone effort in model (2) considerably improves the fit to the data and demonstrates that lone effort is significantly related to the decision to buy credit at the 1% level of significance. In model (2) the average marginal effect of an additional point of effort decreases the probability of purchasing credits by 1.8%.

The third column of Table 4 shows that including gender, and age further improves pseudo R<sup>2</sup> and that females are significantly more likely to purchase credits than males at the 1% level of significance. Furthermore, the effect of lone effort on purchasing credits is much larger, when these variables, and in particular gender are controlled. The average marginal

effect of an additional point of effort in model (3) decreases the probability of purchasing credits by 2.5%. The log likelihood and pseudo R<sup>2</sup> of RE probit (4) in Table 4 demonstrate that allowing for additional heterogeneity in the error term, allows the model to better fit the data. In addition, the variance of the participant component of the error term,  $\sigma_u$ , is statistically significant at the 5% level, evidence that the RE probit (4) is the superior model. The average marginal effect of lone effort and gender in the RE probit (4) are almost identical to model (3) and both are negative and significant at the 1% level.

The results in Table 4 are consistent with Conjecture 2: a low lone effort implies a higher probability that the role A participant purchases credits. Thus:

**Observation 2:** Credit purchase is associated with a lower lone effort.

Recall that the data in Table 2 showed that there was no significant difference in effort between the partnered and unpartnered stages by those who bought credits. This indicates that those participants who bought credits chose to do so rather than increase effort. As a result, their payment from the experiment was lower in the stages with social comparisons than those without. This esteem seeking therefore results in a worse financial outcome for these participants.

**Table 4 Credits Purchase of Role A Participants – Marginal Effects of Probits**

	(1)	(2)	(3)	(4)
Low	0.027 (0.047)	0.022 (0.046)	0.023 (0.046)	0.020 (0.045)
Stage 4	-0.080* (0.047)	-0.071 (0.045)	-0.071 (0.045)	-0.075* (0.043)
S2:Hi, S4:Lo	0.109 (0.091)	0.079 (0.091)	0.106 (0.091)	0.098 (0.089)
Lone Effort		-0.018** (0.007)	-0.025*** (0.007)	-0.026*** (0.006)
Female			-0.213** (0.087)	-0.214** (0.086)
Age			-0.008 (0.014)	-0.007 (0.013)
$\sigma_u$				1.526** (0.777)
pseudo R <sup>2</sup>	0.025	0.090	0.143	0.264
log likelihood	-67.8	-63.3	-59.6	-51.2
n	132	132	132	132

Figures in parentheses ( ) are standard errors clustered by participant. The \*\*\* denote statistically different from zero at 1%, \*\* at 5% and \* at 10% level of significance.

While no gender differences were observed in Table 1, controlling for lone effort in the results in Table 4 indicate that female participants are significantly less likely to purchase credits. The average marginal effect over the sample of being female reduces the probability of purchasing credits by 21% in models (3) and (4). There is also some evidence that role A participants are less likely to buy credits in stage 4, with its estimated parameter being negative and significant at the 10% level in the basic model (1) and RE model (4). There is no evidence that the matched role B partner's performance influences the purchasing decisions, with Low and S2:Hi, S4:Lo being insignificant in all models.

Table 5 gives the average score of role A participants in stages 2 and 4 disaggregated by those who bought credits and those that did not. In line with the increase in effort over stages described in Table 5, there is also an increase in score of all classes of participant from stage 2 to stage 4 in Table 5.

**Table 5 Average Score of Role A Participants in Stages 2 and 4 by Partner Type**

Partner	Not Buy Credits			Buy Credits			All Participants		
	S2	S4	Mean	S2	S4	Mean	S2	S4	Mean
High	36.7	39.6 <sup>++</sup>	38.3	35.4 <sup>‡</sup>	43.8	*37.8	36.3 <sup>^^</sup>	40.1 <sup>++</sup>	38.2
	(1.4)	(1.3)	(1.0)	(1.8)	(1.5)	(1.7)	(1.1)	(1.1)	(0.8)
Low	23	29	52	10	4	14	33	33	66
	37.3 <sup>++</sup>	38.0	37.6 <sup>+</sup>	39.3	44.63 <sup>‡</sup>	*42.1 <sup>+</sup>	37.7 <sup>++</sup>	39.6 <sup>^^</sup>	38.7
Total	(1.2)	(1.4)	(0.9)	(2.8)	(5.8)	(3.4)	(1.1)	(1.8)	(1.0)
	26	25	51	7	8	15	33	33	66
Total	37.0 <sup>***</sup>	38.9 <sup>***</sup>	38.0	37.0 <sup>**</sup>	44.3 <sup>**</sup>	40.0	37.0 <sup>^^^</sup>	39.9 <sup>^^^</sup>	38.4
	(0.9)	(0.9)	(0.8)	(1.6)	(3.8)	(2.2)	(0.8)	(1.0)	(0.8)
	49	54	103	17	12	29	66	66	132

Figures in parentheses ( ) are standard errors clustered by participant with the number of observations directly below. The \*\*\* denote statistically lower/higher at the 1% level of significance, \*\*, ++, ^^ and ^ at 5% and \*, † and ‡, at the 10% level of significance. Symbols before estimates, indicate testing over rows, the figures after the estimates indicate testing across columns.

As shown in Table 1, role A participants who buy credits, buy on average more credits when facing low scoring partners compared to high scoring partners. When facing a high performing partner, those role A participants who buy credits buy just enough so that their score is not significantly different from those who do not buy credits ( $p=0.768$ ), but less than 40, the score of a high-performing role B participant ( $p=0.1057$ ). In contrast, when facing a low performing partner, role A participants who bought credits had a significantly higher average score than those role A participants who did not buy credits ( $p=0.095$ )<sup>7</sup>. Although

<sup>7</sup> This result is difficult to reconcile with the predictions of signalling models based on adverse selection, such as Ellingsen and Johannesson (2008). Such an empirical finding could not reflect an equilibrium: for this to be the case a high score from a role A participant would indicate that the participant was in fact a low effort participant. That is, low effort participants are expending effort to indicate they are low effort. However, they could have equally well and costlessly sent this signal without purchasing credits.

this average score is greater than 40 (the score of a high performing role B participant) this difference is not significant ( $p=0.267$ ).

Table 6 shows the average score of those who buy and do not buy credits by gender. Females who do not buy credits have a lower score than males who do not buy credits ( $p<0.001$ ). This reflects the finding in Table 2 that females exhibit lower effort than males. However, of those who buy credits, there is no significant difference between the score of males and females ( $p=0.550$ ). This implies those females who buy credits, do so to generate a score that is similar to that of the average participant. In contrast, the overall average score of females is significantly lower than that of males ( $p<0.001$ ).

**Table 6 Average Role A Participant Score by Gender and Credits Purchase Decision**

Gender	Not Buy	Buy	All
Female	35.7***	38.8	36.3^^
	(0.92)	(3.31)	(0.94)
	63	15	78
Male	41.6***	41.4	41.5^^
	(1.24)	(2.77)	(1.18)
	40	14	54

Figures in parentheses ( ) are standard errors clustered by participant with the number of observations directly below. The \*\*\* and ^^ denote significantly lower/higher at the 1% level of significance.

To summarise the findings on gender from Table 1, 2, 4 and Table 6, even accounting for females having lower average effort, they are less likely to buy credits. However, those that do buy credits, buy sufficient credits to ensure that their score is close to the average score for all participants of 38.4 (shown in Table 5). The lower probability of females buying credits might be due to females being less competitive than males (Niederle and Vesterlund, 2007). Those that do buy credits may be concerned with “not standing out”, particularly as a low performer. This interpretation would be consistent with the results of Jones and Linardi (2014), who find that women are more motivated than men to choose a social signal which makes them appear average. However, the data in Table 6 suggests that males are also buying credits to bring their score up to the average of the males who do not buy credits. It should be noted, though, that is not possible to conclusively verify these interpretations of the gender effects using our data.

### 3.2 Feedback and Fairness

In this section we first present data relating to the feedback given by role B participants to role A participants. The role B participants give feedback on a 5-point Likert scale, where 1 is very bad and 5 is very good. We then examine the data on how role A and role B participants view the fairness of this feedback. In the post-experiment survey, the role A

participants indicate their perception of the fairness of their matched partner's feedback, and role B participants were also asked how fair the feedback they gave was. Fairness was reported on a 5-point Likert scale, where 1 is "very unfair" and 5 is very fair.

### 3.2.1 Summary of Feedback and its Fairness

Table 7 shows the average feedback provided by role B participants to role A participants, role A participants' average assessment of the fairness of that feedback, and role B participants' average assessment of the fairness of the feedback they gave.

Feedback from high performing role B participants is smaller than from low performing role B participants ( $p=0.098$ ). High performers may have higher standards and be harsher in their assessment. Alternatively, high performers may be more sceptical than low performers, that the score of their matched role A participants reflects their effort, rather than the purchase of credits. Role A and B's perception of fairness is not statistically different across high and low performing Role B participants.

The figures in Table 7 show that on average role A participants have a significantly less favourable perception of the feedback they receive than role B participants have of the feedback they give ( $p=0.001$ ). This finding is not consistent with Conjecture 3(i): Role B participants and their matched role A role A participant agree on the fairness of the role B participant's feedback. Conjecture 3(i) is rejected at the 1% level of significance ( $p=0.002$ )

**Table 7 Average Feedback and Fairness Scores**

Role		Role B's Type		Total
		High	Low	
A	Feedback	3.92*	4.14*	4.03
		(0.13)	(0.10)	(0.08)
	Fairness	3.67 <sup>++</sup>	3.80 <sup>**</sup>	3.73 <sup>***</sup>
		(0.14)	(0.12)	(0.09)
	n	66	66	132
B	Fairness	4.08 <sup>++</sup>	4.09 <sup>**</sup>	4.08 <sup>***</sup>
		(0.10)	(0.08)	(0.07)
	n	66	66	132

Figures in parentheses ( ) are standard errors clustered by participant. The \*\*\* denote statistically lower/higher at the 1% level of significance, \*\* and ++ at the 5% level of significance and \* at the 10% level of significance.

### 3.2.2 Determinants of Role B Participants' Feedback

Conjecture 3(ii) states that the role B's feedback will be related to their matched role A participant's reported score. The feedback given by role B participants may also be influenced by their own effort (which they may use as a benchmark for assessment) or their

demographic characteristics. Summary statistics on how these determinants are correlated with feedback are provided in Table 11 in Appendix B.

To estimate these effects with appropriate controls, we model Role B's feedback using an ordered probit as a function of these determinants for the 66 Role B participants over the two partnered stages.<sup>8</sup> The output of this estimation is reported in the column labelled 'B's feedback' in Table 8 in the following section 3.2.3. The results show that a role A participant's score is significantly correlated with the feedback provided by their matched role B participant. It is also observed that on average females provide significantly higher feedback. There is some evidence that the higher role B participant's lone effort the lower is their feedback. This may reflect that high performers have a higher expectation of performance in others.

However, the results in Table 8 provide no evidence that the role B participant's partnered effort, or whether their performance is classified as high or low influences their feedback. Additionally, there is no evidence of an ordering effect in feedback, so that experience of providing feedback in stage 2 does not change the feedback in stage 4. This is consistent with role B participants having well-formed perceptions of performance.

### 3.2.3 Determinants of Role A Participants' view of Fairness

In this section we consider the determinants of the role A participant's perception of the feedback they receive.

A role A's assessment of fairness may vary systematically with either of the two components of score. In particular, there might be a systematic relationship between effort and the fairness perception of the feedback, though it is not possible to specify a direction of this effect a priori. For example, a participant who provides little effort and receives good feedback (and vice-versa) may feel the feedback is unfair. Alternatively, a participant who provides high effort and receives good feedback may feel vindicated, thus feel the feedback was fair.

---

<sup>8</sup> A likelihood ratio test that the random effects ordered probit was an improvement on the ordered probit was rejected for three models in Table 8. In addition, the participant component of the random effects error variance was not significant for the three ordered probit models.

**Table 8 Feedback and Fairness - Ordered Probits**

	<b>B's Feedback</b>	<b>A's Fairness</b>	<b>B's Fairness</b>
Low	0.047 (0.243)	-0.094 (0.177)	0.032 (0.279)
Stage 4	-0.013 (0.177)	-0.099 (0.214)	0.172 (0.173)
Lone Effort	-0.045* (0.024)	0.033 (0.035)	0.038* (0.020)
Partnered Effort	0.021 (0.022)	-0.017 (0.031)	-0.017 (0.023)
Buy		-0.421 (0.289)	
Credits		0.065** (0.032)	
A's Score	0.077*** (0.020)		-0.021 (0.014)
Female	0.516** (0.205)	0.211 (0.260)	-0.241 (0.232)
Age	0.007 (0.021)	0.047 (0.030)	0.020 (0.018)
Feedback = 1		-1.570** (0.651)	-1.893* (1.147)
Feedback = 2		-0.667 (0.487)	0.743 (0.759)
Feedback = 4		0.962*** (0.356)	0.272 (0.267)
Feedback = 5		1.720*** (0.420)	0.959** (0.386)
Cut 1	-0.071 (0.873)	0.415 (1.380)	-2.016* (1.141)
Cut 2	0.708 (0.810)	1.171 (1.352)	-1.089 (1.079)
Cut 3	1.674** (0.824)	2.009 (1.372)	-0.293 (1.045)
Cut 4	2.979*** (0.836)	3.644** (1.428)	1.526 (1.058)
pseudo R <sup>2</sup>	0.131	0.181	0.077
log likelihood	-144.2	-146.2	-127.8
n	132	132	132

Figures in parentheses ( ) are standard errors clustered by participant. The \*\*\* denote statistically different from zero at 1%, \*\* at 5% and \* at 10% level of significance.

Additionally, Role A's assessment of fairness may be influenced by the number of credits they purchased. The decision to buy credits, and/or the number bought may lead a participant to expect or anticipate high feedback or, alternatively, might make the participant feel their score is fraudulent. The former effect would make credit purchase increase the fairness perception of positive feedback, while the latter would see credit purchases reduce fairness perceptions of positive feedback. If individuals value the esteem associated with positive feedback independent of the components of their score, it would be expected that the perception of fairness is positively related to the level of feedback.

To estimate these effects with appropriate controls, we model A's perception of the fairness of the feedback they receive using an ordered probit with controls for these potential determinants. The estimation results for the 66 Role A participants over the two stages are reported in the second column of figures in Table 8. For reference, summary statistics on how these determinants are correlated with role A's perception of the fairness of feedback are provided in Table 12 in Appendix B.

These results indicate that role A participants assess their feedback as fairer when they receive "good" feedback and even more fair when they receive "very good" feedback compared to neutral feedback. Furthermore, fairness assessment is significantly lower for "very poor" feedback compared to neutral feedback. These results suggest that role A participants value good feedback per se, irrespective of whether that good feedback is justified or not.

Table 8 also provides evidence that the role A participant's assessment of fairness is significantly positively related to the quantity of credits purchased ( $p=0.042$ ). This finding reflects that, after controlling for the feedback, those who buy more credits believe their feedback is fairer. This finding suggests that the role A participants who buy credits do not feel that their score is fraudulent, rather appear happy that their purchase of credits was successful in generating positive feedback.

Summarising the data in Table 7 and Table 8, we derive the following conclusions concerning feedback and its perceived fairness:

**Observation 3:** (i) On average role A participants have a significantly less favourable perception of the feedback they receive than role B participants have of the feedback they give. (ii) Feedback provided to role A participants is significantly related to their reported score. (iii) The level of feedback and number of credits bought, and not effort, influences role A participants' assessment of the fairness of their received feedback.

Observation 3(i) is not consistent with Conjecture 3(i): role A and B participants are not observed to agree on the fairness of feedback. Observation 3(ii) is consistent with Conjecture

3(ii). Overall, Observation 3 suggests the following regarding feedback. A role A participant's score will influence role B's feedback. However, that feedback is less than would be wanted by the role A participant, who seemingly want good feedback irrespective of their effort.

### 3.2.4 Determinants of Role B Participants' View of their Own Fairness

In this section we consider the determinants of Role B participants' perception of the fairness of the feedback they gave. Role B participants' perception of fairness might be related to their demographic characteristics, their own effort, the score of their matched role A partner, or the feedback they give. We estimate an ordered probit of Role B fairness perception as a function of these variables for the 66 Role B participants over the two stages. The results of the estimation are reported in the third column of figures in Table 8. For reference, summary statistics of the absolute correlation between these variables and B's fairness are given in Table 13 in Appendix B.

Role B participants who give 'very good' as feedback, are more likely to believe the feedback they provide is fair than those giving neutral feedback ( $p=0.013$ ). There is also some evidence that people who give 'very bad' as feedback, are more likely to believe the feedback they provide is unfair ( $p=0.099$ ).

There is some evidence in Table 8, that those with high lone effort have greater faith in the fairness of their feedback ( $p=0.057$ ). Given the evidence that role B's gave lower feedback the higher their lone effort in section 3.2.2, this result suggests that those role B participants who are high performers provide 'harsh' but, in their own estimation, 'fair' feedback.

This finding suggests a possible motivation for role A participants who we observe buying more credits when matched with a low performing role B participant: they may be hoping for more favourable feedback from their matched low performing participant. However, such hope appears to be misplaced. As reported in Table 8, neither partnered effort or Low are determinants of either role B's feedback or their perception of fairness. Nonetheless, in the structure of the experiment the categorisation of performance as High/Low is the only signal the role A participant receives about their matched participant. So, role A participants, working with what they've got, respond to that signal. Clearly, though, such conclusion cannot be definitive based on our current experiment. Further experimental investigation of this potential relationship would be useful to clarify the issue

No variable other than lone effort and Feedback=5 in the final column of Table 8 has a statistically significant correlation to role B participant's perception of their own fairness. Importantly this includes the matched role A participant's score. This finding is consistent with role B participants believing they provide fair feedback regardless of the scores

presented to them. Taken with the results in Table 7, the results in Table 8 suggest that most role B participants systematically believe their feedback is not unfair with only 5 out of 132 rating it less than 3 for fairness.

## 4 Discussion

The experiment reported in this paper was designed to elicit some of the key determinants of esteem seeking and provision when: (i) esteem seeking yields only private benefits, and (ii) the scope for participants to esteem seek is common knowledge. Approximately 22% of relevant (role A) participants in our experiment engaged in overt esteem seeking, i.e. purchased credits. Credits provide no material benefit to the participant who buys them, but their purchase is shown to elicit positive feedback (esteem) from their partner. The interpretation of credit purchases as esteem seeking is further supported by the high correlation between the participants' actual purchase of credits with their response to the post-experiment question on the motivation for buying credits.

We find that participants who exhibit a low willingness to supply effort are more likely to engage in esteem seeking. Specifically, those who exhibit low lone effort (effort chosen without social influences) when paid by piece rates are more likely to buy credits.

In spite of the differences noted in the introduction, between our experiment and that of Friedrichsen and Engelmann's, in this respect their findings parallel ours. In both experiments, the participants who could be considered the 'low' types are the ones who actively esteem seek: those with low measured intrinsic motivation to supply the public good in Friedrichsen and Engelmann's paper, and a low willingness to supply effort in ours. Both these findings are also consistent with the field evidence of Charles et. al. (2009) that it is low-income people who engage in conspicuous consumption. The conclusion that 'low' types are more likely to esteem seek thus does not depend on whether esteem is generated through providing a public good (supporting fair trade) or an activity with only private benefits (performance on a real effort task).

Our findings also suggest that those who engage in esteem seeking do so more in a low performing social environment than a high performing social environment. Specifically, role A participants purchase more credits when facing a low performing role B participant than a high performing one.

Together, these two findings suggest that it is those who choose to provide low effort that are more likely to engage in esteem seeking activities to mask their low performance, and that those participants buy more credits when facing a low scoring partner. This conclusion suggests that when people are seeking esteem, they prefer to turn to people who have a

similarly low performance on a real effort task. Such a finding is reminiscent of Festinger's (1957) definition of a person's referents as people who are "close to [their] own ability".

In our experiment role B participants' feedback is used to capture esteem provision. We find that the possibility a role A participant's score can be inflated by credit purchases does not cause their matched role B participant to dismiss their score when assessing their performance: role B feedback is positively related to their matched role A's score.

The fairness assessments of the feedback of the two groups differ somewhat. We observe that role Bs assess their own feedback as fairer than do role As. Furthermore, the role A's fairness assessment of feedback is more closely related to the positivity of that feedback than that of role B participants. We find that the higher the credit purchase by a role A participant, the fairer they view the feedback they receive. Thus those who engaged in esteem seeking (through the purchase of credits) appear satisfied with the response to their esteem-seeking: specifically, they see it as fair.

Extrapolating beyond the lab, our findings suggest that those who engage in esteem seeking would do so more actively in low esteem/status environments. For example, people would devote more effort to impress others with their fashion choices when those around them are not as likely to do so. Our experiment predicts that this effort will indeed be rewarded with additional esteem offered by others. Similarly, extrapolating from our experiment suggests that it will be those who are unwilling, or find it difficult, to engage in an esteem-generating activity who will be the ones most likely to esteem seek. For example, those who have difficulty making informed fashion choices (e.g., the time poor or the colour blind) can simply spend money to buy prestige labels and/or purchase advice. Such people are often rewarded for their expenditure by being complimented on their fashion choices.

Esteem seeking outside the lab usually occurs in a social context. When, in contrast to the lab's anonymity, people are identifiable, the benefit of esteem seeking would be expected to be higher. Engagement in esteem seeking would therefore also be higher. Similarly, the type of activity may influence peoples' extent of esteem-seeking. For example, a person might be expected to be more willing to esteem seek when buying new clothes than when buying credits to improve their score on a real effort task.<sup>9</sup> Of course, while these hypotheses on the prevalence of esteem seeking outside the lab seem plausible, they do require empirical verification.

---

<sup>9</sup> On this topic, it is worth keeping in mind that Akerlof (2017) argues that the activities which yield esteem are endogenous.

## References

- Abbink, K. and Sadrieh, A. (2009) "The pleasure of being nasty", *Economics Letters*, vol. 105, pp. 306–308.
- Akelof, R. (2017), "Value Formation: The Role of Esteem," *Games and Economic Behavior*, vol. 102, pp. 1-19.
- Azmat, G. and Iriberry, N. (2010) "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students", *Journal of Public Economics*, 2010, vol. 94 (7-8), pp. 435-452.
- Azmat, G. and Iriberry, N. (2016) "The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction", *Journal of Economics & Management Strategy*, vol. 25, pp. 77–110.
- Bernheim, B. Douglas, (1994) "A theory of conformity", *Journal of Political Economy*, vol. 102(5), pp. 841–877.
- Benabou, R. and Tirole, J. (2003) "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*, 70, pp. 489–520.
- Benabou, R. and Tirole, J. (2006) "Incentives and Prosocial Behavior", *American Economic Review*, December, vol. 96(5), pp. 1652-78.
- Blacklow, P., Corman, A. and Sibly, H. (2018) "The Demand and Supply for Esteem: An experimental analysis", *Tasmanian School of Business and Economics Discussion Paper Series*, no. 2018(03).
- Brennann, G. and Pettit, P. (2004) *The economy of esteem: An essay on civil and political society*, Oxford and New York: Oxford University Press.
- Burszтын, L., Fermany, B., Fiorinz, S., Kanx, M., Rao, G. (2018), "Status Goods: Experimental Evidence from Platinum Credit Cards" *Quarterly Journal of Economics*, vol. 133(3), pp. 1561–1595.
- Charles, K. K., Hurst, E., Roussano, N., (2009) "Conspicuous Consumption and Race", *Quarterly Journal of Economics*, vol. 124(2), pp. 425–467.
- Charness, G, and Grosskopf, B. (2001) "Relative payoffs and happiness: an experimental study", *Journal of Economic Behavior & Organization*, vol. 45, pp. 301–328.
- Charness, G, Masclet, D. Villeval, M.C. (2010) "Competitive Preferences and Status as an Incentive: Experimental Evidence", *IZA DP*, no. 5034.
- Charness, G, Masclet, D. Villeval, M.C. (2014) "The Dark Side of Competition for Status" *Management Science*, vol. 60, pp. 38-55.
- Cowen, T. and Glazer, A. (2007) "Esteem and ignorance", *Journal of Economic Behavior & Organization*, vol. 63, pp. 373–383.
- Dohmen, T., Falk, A., Fliessbach, K. Sunde, U., Weber, B. (2011) "Relative versus absolute income, joy of winning, and gender: Brain imaging evidence", *Journal of Public Economics*, vol. 95, pp. 279–285.

- Dwenger, N., Kleven H., Rasul, I., Rincke, J.,(2016) "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany", *American Economic Journal: Economic Policy*, vol. 8(3), pp. 203–232.
- Ellingsen, T. and Johannesson, J., (2008) "Pride and Prejudice: The Human Side of Incentive Theory", *American Economic Review*, vol. 98(3), pp. 990-1008.
- Festinger, L., (1954) "A Theory of Social Comparison Processes", *Human Relations*, vol. 7, pp. 117-140.
- Fischbacher, U., (2007) "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments", *Experimental Economics*, vol. 10(2), pp. 171-178.
- Fischbacher, Urs, and Franziska Föllmi-Heusi. (2013) "Lies in Disguise: An Experimental Study on Cheating." *Journal of the European Economic Association*, vol. 11 (3), pp. 525–47.
- Friedrichsen, J. and Engelmann, D. (2018) "Who cares about social image?", *European Economic Review*, vol. 110, pp. 61-77.
- Frank, Robert H., (1985) *Choosing the Right Pond: Human Behavior and the Quest for Status*, Oxford University Press.
- Frank, Robert H., (1997), "The Frame of Reference as a Public Good", *The Economic Journal*, vol. 107, No. 445, pp. 1832-1847.
- Frank, Robert H. (2005) "Positional Externalities Cause Large and Preventable Welfare Losses", *The American Economic Review*, vol. 95(2), pp. 137-141.
- Gill, D., Kissova, Z., Jaesum, L. and Prowse, V. (2017) "First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision", forthcoming in *Management Science*.
- Gill, D. and Prowse, V. (2012) "A Structural Analysis of Disappointment Aversion, in a Real Effort Competition", *American Economic Review*, vol. 102(1), pp. 469–503.
- John, O., and S. Srivastava (1999): "Big Five Trait Taxonomy. History, Measurement, and Theoretical Perspectives," in *Handbook of Personality. Theory and Research*, ed. by L. Pervin, and O. John, pp. 102–139. Guilford, New York.
- Jones, Daniel and Linardi, Sera. (2014). "Wallflowers: Experimental Evidence of an Aversion to Standing Out", *Management Science*, vol. 60, pp. 1757-1771.
- Kirchler, M. and Palan, S. (2018) "Immaterial and monetary gifts in economic transactions: evidence from the field" *Experimental Economics*, vol. 21, pp. 205-230.
- Lang, P. J. (1980) "Behavioral treatment and bio-behavioral assessment: computer applications", In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex.
- Mazar, Nina, On Amir, and Dan Ariely. 2008 "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research*, vol. 45(6), pp. 633–44.
- Niederle, M. and Vesterlund, L. (2007) "Do Women Shy away from Competition? Do Men Compete too Much?", *Quarterly Journal of Economics*, vol. 122(3), pp. 1067-1101.
- Rotter, J. (1966) "Generalized expectancies for internal versus external control of reinforcement", *Psychological Monographs: General & Applied*, vol. 80 (1), pp. 1–28.

- Riedl A, Smeets P. (2017), "Why Do Investors Hold Socially Responsible Mutual Funds?", *The Journal of Finance*, vol. 72(6), pp. 2505-49.
- Tesser, Abraham, Campbell, Jennifer, (1980) "Self-definition: the impact of the relative performance and similarity of others", *Social Psychology Quarterly*, vol. 43(3), pp. 341-347.
- Tesser, Abraham, Campbell, Jennifer, (1982) "Self-evaluation maintenance and the perception of friends and strangers", *Journal of Personality*, vol. 50 (3), pp. 261-279.
- Tran A. and Zeckhauser, R. (2012) "Rank as an inherent incentive: Evidence from a field experiment", *Journal of Public Economics*, vol. 96, pp. 645-650.
- Veblen, Thorstein (1899) *The Theory of the Leisure Class*. Project Gutenberg.

## Appendices

The Demand and Supply for Esteem: An  
experimental analysis

Intended for Online Publication

## Appendix A.

### Summary of Instructions

These printed instructions contain identical information to the instructions on the computer screen and are provided for your reference.

#### About this experiment:

All participants receive the same information about the way this experiment is conducted. As with all economics experiments, we will not use deception of any kind. Everything stated in these instructions is true.

You will receive payment for participating in this experiment. The amount of money you actually leave with will depend on your actions in the experiment. You will be paid privately in cash at the end of the session.

#### Payment:

Throughout the experiment we will refer to payments in Exp\$ or Experimental Dollars. At the end of the experiment these will be converted into Australian Dollars at the rate of

$$10 \text{ Exp\$} = \text{AU\$}1.00$$

The experiment has five stages, and you will be paid your earnings from each stage. You will also be paid a show up fee of AU\$10.00 which is paid to all participants.

#### Experimental Rules:

If you do not comply with these rules you will not be paid your earnings from this experiment. You will not be allowed to participate in future sessions.

1. Talking is not permitted during the experiment. You must not share any information with others during the experiment.
2. Only the experiment windows are permitted to be open during the experiment. You are not permitted to operate other software such as email or internet during the experiment. This also applies to your own devices including mobile phones.
3. If you have a question, please raise your hand and an experimenter will come answer your question.

Experimenters can answer questions about procedures but cannot provide you with advice about decisions. You must make decisions by yourself.

## Slider Task:

In this experiment you must undertake a series of "slider tasks" which works as follows:

Each slider is initially positioned at 0 and can be moved as far as 100.

Each slider has a blue number to the left showing the target value.

Each slider has a black number to its right showing its current position.

You can use the mouse in any way you like to move each slider. You can readjust the position of each slider as many times as you wish.

You need to make the current position match the target value for each slider. Everyone has the same targets.

## Trial Slider Task:

On the screen are 3 sliders. The purpose of this trial is to ensure that you understand how to place the slider at the target. This trial will not end until you have successfully placed all three sliders.

You will not be paid for the trial.

## Structure of the experiment:

### Roles:

You will be randomly assigned a role for the entire experiment: Role A or Role B.

Half of the participants will be randomly assigned Role A and the other half will be assigned Role B.

You will find out which role you have been assigned after you complete the first quiz

### Stages:

This experiment has five stages.

Each stage will begin with a brief review of the instructions relevant to that stage and a short quiz.

In each stage participants then undertake a slider task. Performance on the slider task will play a role in how much you are paid.

In Stages 1, 3 and 5 participants do not interact with any other participant.

In Stages 2 and 4 each participant in Role A is randomly matched with a participant in Role B. You will be matched with a different participant in Stage 2 and Stage 4. You will never learn who the person is that you are matched with.

The experimental procedures are different for those in roles A and role B. These are now described.

If you are randomly assigned to Role A your payment is determined by:

In Stages 1, 3 and 5 you will be paid 1 Exp\$ for each slider you correctly place in the available time.

In Stages 2 and 4 you get 1 Exp\$ for each slider you correctly place in the available time but you will have the opportunity to spend your Exp\$ to buy credits to increase your score as explained below.

In Stages 2 and 4 your matched player in Role B will see your score for that stage and give you feedback on your score.

Before your score is shown to your matched Role B participant, you can buy **credits** to increase the score that your matched participant sees.

Each credit will cost 1 Exp\$. You can spend up to the total amount of Exp\$ that you have earned in that stage.

Your matched Role B participant will see your score which is calculated by:

$$\text{Score} = \text{the number of sliders correctly placed} + \text{any credits bought}$$

Your matched Role B participant will only see your score, and not how many sliders you correctly placed or if you have bought any credits.

Your payment in stages 2 and 4 is calculated by:

$$\text{Payment} = \text{the number of sliders correctly placed} - \text{any credits bought}$$

The Grid below summarises Role A:

Role A	Payment	Matching/Info Shared	Feedback
Stage 1	1 Exp\$ for each slider correctly placed	Score is private	None
Stage 2	1 Exp\$ for each slider correctly placed less 1 Exp\$ for each credit purchased	Score (number of correctly placed sliders plus any credits bought) is told to your matched Role B	Your matched Role B will give you feedback on your score
Stage 3	1 Exp\$ for each slider correctly placed	Score is private	None
Stage 4	1 Exp\$ for each slider correctly placed less 1 Exp\$ for each credit purchased	Score (number of correctly placed sliders plus any credits bought) is told to your matched Role B	Your matched Role B will give you feedback on your score
Stage 5	1 Exp\$ for each slider correctly placed	Score is private	None

The number of sliders you place and the number of credits you purchase are never revealed to other participants. In stage 2 and stage 4, only your score is reported to your matched partner.

If you are randomly assigned to Role B your payment is determined by:

In all five stages the computer will **rank** each Role B participant based on the number of correct sliders in that stage.

The **highest performing half of participants in Role B will be assigned a payment of 40 Exp\$** for that stage.

The **lowest performing half of participants in Role B will be assigned a payment of 10 Exp\$** for that stage.

If two or more participants have completed the same number of sliders (tied) the relative ranking of these participants will be determined randomly.

This means that each participant in Role B could be paid 10 Exp\$ or 40 Exp\$ in any of the five stages. The participant's score will be equal to their payment.

Payment in each stage is determined independently from the other stages.

In Stages 2 and 4 you will see the score of your matched Role A participant and they will see your payment for that stage.

You will be asked to provide feedback to your matched Role A after you have seen their score.

The Grid below summarises Role B:

<b>Role B</b>	<b>Payment</b>	<b>Matching/Info shared</b>	<b>Feedback</b>
Stage 1	10 or 40 Exp\$ based on the number of correct sliders in Stage 1 relative to other Role B participants	Payment is not seen by other participants	None
Stage 2	10 or 40 Exp\$ based on the number of correct sliders in Stage 2 relative to other Role B participants	Matched Role A will see your payment (10 or 40 Exp\$) and you will see their score	You will give feedback to your matched Role A on their score
Stage 3	10 or 40 Exp\$ based on the number of correct sliders in Stage 3 relative to other Role B participants	Payment is not seen by other participants	None
Stage 4	10 or 40 Exp\$ based on the number of correct sliders in Stage 4 relative to other Role B participants	Matched Role A will see your payment (10 or 40 Exp\$) and you will see their score	You will give feedback to your matched Role A on their score
Stage 5	10 or 40 Exp\$ based on the number of correct sliders in Stage 5 relative to other Role B participants	Payment is not seen by other participants	None

## Summary of Matching and Information:

The following table summarises the matchings made in each stage, and the information provided to each participant.

	<b>Both roles: Role A &amp; Role B</b>
Stage 1	Participants are not matched. No information is available to any participant about the performance on the slider task of other participants.
Stage 2	Each participant is matched with a participant in a different role. The participant in Role A knows whether their matched participant in Role B has a score of 10 or 40 (depending on the matched participant's performance on the slider task in Stage 2). The participant in Role B sees the score in Stage 2 of their matched participant and provides feedback on that score.
Stage 3	Participants are not matched. No information is available to any participant about the performance on the slider task of other participants.
Stage 4	Each participant is matched with a participant in a different role (who is different from the matched participant in Stage 2). The participant in Role A knows whether their matched participant in Role B has a score of 10 or 40 (depending on the matched participant's performance on the slider task in stage 4). The participant in role B sees the score of their matched participant and provides feedback on that score.
Stage 5	Participants are not matched. No information is available to any participant about the performance on the slider task of other participants.

## ID:

You will be randomly assigned an ID for the entire experiment.

Role A participants will be randomly assigned a colour from the following list:

GREEN  
YELLOW  
WHITE  
RED  
PURPLE  
CYAN  
BROWN  
ORANGE  
MAGENTA  
BLACK  
GREY  
BLUE

Role B participants will be randomly assigned a shape from the following list:

RECTANGLE  
TRIANGLE  
HEXAGON  
CIRCLE  
PENTAGON  
TRAPEZOID  
DIAMOND  
SQUARE  
MOON  
OCTAGON  
STAR  
OVAL

You will never learn which participants are assigned to which IDs.

## Feedback:

In Stages 2 and 4:

Each Role A participant will see the payment for their matched Role B participant.

Each Role B participant will see the score for their matched Role A participant.

Each Role B participant will then provide feedback to the Role A participant on their score by answering the following question:

*You must now select the feedback you would like to send GREEN about their score:*

*Very Good*

*Good*

*Neither good nor bad*

*Bad*

*Very Bad*

The Role A participant will then see this feedback as follows:

*RECTANGLE thinks your score in*

## Appendix B.

### Additional data analysis

Table 9 summarizes the responses by role A participants in to the post-experiment survey question which asked them whether knowing they would receive feedback on their score caused them to buy credits in stages 2 and 4. Table 9 includes the calculation of the Spearman rank correlation between the decision to buy credits and the answer to this post-experiment survey question. The spearman rank correlation between the decision to buy and feedback on credits is 0.6285 (0.643 in stage 2 and 0.602 in stage 4) and significant at the 1% level of significance.

**Table 9 Proportion who Bought Credits by Feedback on Credits**

	"Do you think that knowing you would get feedback from RECTANGLE caused you to buy Credits in Stage <i>r</i> "				
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
Stage 2	8%	11%	15%	26%	41%
Stage 4	9%	8%	8%	33%	42%
Total	8%	9%	11%	30%	42%

Table 10 on the following page, provides the estimated coefficients and estimation details of the four probit models for role A participants' choice to buy credits as outlined in section 3.1. The marginal effects from this regression are provided in Table 4 in section 3.1.

**Table 10 Credits Purchase of Role A Participants – Estimated Coefficients of Probits**

	(1)	(2)	(3)	(4)
Low	0.094 (0.163)	0.080 (0.174)	0.090 (0.183)	0.195 (0.434)
Stage 4	-0.276* (0.163)	-0.265 (0.175)	-0.281 (0.185)	-0.714 (0.467)
S2:HiS4:Lo	0.377 (0.316)	0.290 (0.331)	0.423 (0.367)	0.939 (0.924)
Lone Effort		-0.067** (0.029)	-0.101*** (0.028)	-0.243** (0.100)
Female			-0.844** (0.353)	-2.033** (1.005)
Age			-0.032 (0.054)	-0.069 (0.128)
Constant	-0.893*** (0.247)	1.600 (1.110)	3.968** (1.847)	9.478* (5.092)
$\sigma_u$				1.527** (0.777)
pseudo R <sup>2</sup>	0.025	0.090	0.143	0.264
log likelihood	-67.8	-63.3	-59.6	-51.2
n	132	132	132	132

Figures in parentheses ( ) are standard errors clustered by participant. The \*\*\* denote statistically different from zero at 1%, \*\* at 5% and \* at 10% level of significance.

To provide greater insight into feedback and fairness in this section we present summary statistics and Spearman's rank correlation of feedback with their determinants identified in section 3. This measure of absolute correlation is then contrasted with partial correlations of feedback with its determinants as given in Table 8 (where we control for the determinants).

For each of the five levels of feedback, Table 11 presents the average score of role A's who received that feedback, and the average of B's lone and partnered effort, the proportion of B's who were low performers, female and who gave that that feedback in stage 4. The second last column provides the average for all feedback levels and the final column provides the Spearman rank correlation between feedback and each of the possible determinants. The feedback that Role B participants give is most strongly positively correlated with A's Score, significantly positively related to being female and significantly negatively related to their lone effort.

**Table 11 Potential Determinants of Role B's Feedback**

<b>Feedback:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Average</b>	<b>Correlation</b>
A's Score	38.5 (9.5)	34.1 (3.9)	32.9 (1.1)	37.7 (0.6)	42.6 (1.2)	38.4 (0.6)	0.469*** (0.080)
Lone Effort	45.2 (2.8)	41.0 (4.1)	36.3 (1.3)	35.8 (1.1)	34.5 (1.5)	35.8 (0.9)	-0.162* (0.096)
Partnered Effort	42.5 (2.5)	40.3 (4.1)	35.4 (1.1)	34.6 (1.2)	35.2 (1.4)	35.4 (0.9)	-0.044 (0.103)
Low	0% (0%)	29% (26%)	48% (11%)	55% (7%)	51% (8%)	50% (5%)	0.077 (0.092)
Female	0% (0%)	0% (0%)	57% (11%)	53% (8%)	66% (8%)	55% (6%)	0.219** (0.093)
Stage 4	50% (50%)	71% (12%)	43% (10%)	42% (5%)	60% (6%)	50% (0%)	0.086 (0.078)
n	2	7	23	53	47	132	132

Figures in parentheses ( ) are standard errors clustered by participant. . The \*\*\*, \*\* and \* denote correlation is significantly different form zero at the 1%, 5% and 10% level of significance, respectively.

**Table 12 Potential Determinants of Role A's Fairness of Feedback**

<b>Fairness:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Average</b>	<b>Correlation</b>
Feedback	2.4 (0.3)	3.4 (0.3)	3.4 (0.2)	4.3 (0.1)	4.6 (0.1)	4.0 (0.1)	0.581*** (0.078)
Score	36.6 (3.5)	39.0 (2.0)	32.9 (1.4)	38.7 (0.8)	42.2 (2.1)	38.4 (0.8)	0.290*** (0.097)
Lone Effort	35.4 (3.3)	38.2 (1.9)	33.9 (1.3)	38.1 (0.9)	39.5 (1.3)	37.6 (0.7)	0.254** (0.103)
Partnered Effort	36.4 (3.5)	37.9 (1.8)	32.2 (1.3)	37.6 (0.9)	39.9 (1.5)	37.2 (0.7)	0.272*** (0.097)
Buy	14% (12%)	36% (15%)	23% (9%)	23% (7%)	17% (7%)	22% (5%)	-0.070 (0.084)
Low	29% (16%)	45% (16%)	55% (7%)	52% (3%)	50% (6%)	50% (0%)	0.038 (0.069)
Female	29% (20%)	45% (16%)	77% (12%)	65% (8%)	47% (12%)	59% (6%)	-0.046 (0.103)
Stage 4	57% (17%)	64% (15%)	50% (7%)	44% (3%)	57% (6%)	50% (0%)	-0.013 (0.070)
n	7	11	22	62	30	132	132
Credits	1.0 (0.0)	3.0 (0.8)	3.2 (1.1)	5.0 (1.0)	13.8 (6.9)	5.8 (1.5)	0.337 (0.199)
n	1	4	5	14	5	29	29

Figures in parentheses ( ) are standard errors clustered by participant. \*\*\* denote the correlation is statistically different from zero at the 1% level and \*\* at the 5% level of significance.

Table 12 presents the averages or proportions of the potential determinants of A's fairness for each of the five fairness scores, the average over all scores and their Spearman rank correlation with fairness. The data in Table 12 show a positive and significant correlation between role A's score and their perception of the fairness of the feedback they receive. However, we noted above in regard to Table 11 that score is significantly related to feedback. For this reason, to identify the separate effect of score, it is necessary to estimate ordered probits, reported in Table 8 for the feedback B's provide and how fair role A and B participants' feel the feedback is. This analysis indicates that that A's score has only a small and insignificant effect on fairness. after controlling for credit purchases on fairness. In addition, the positive and significant correlation between partnered effort and A's perception of fairness in Table 12, does not exist when we control for feedback and other determinants with the ordered probits in Table 8.

Table 13 presents the averages or proportions of the potential determinants of B's fairness for each of the five fairness scores, the average over all scores and their Spearman rank correlation with fairness. Only the feedback that B provides is significantly correlated with how fair B feels their feedback was in Table 13. When we control for feedback and the other determinants in section 3.2.4 with the ordered probit, we also found some evidence that B's lone effort is positively related to how fair they believe their feedback was.

**Table 13 Potential Determinants of Role B Assessment of Fairness**

<b>Fairness</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Average</b>	<b>Correlation</b>
Feedback	1.0 (0.0)	4.3 (0.3)	3.8 (0.3)	4.0 (0.1)	4.3 (0.2)	4.0 (0.1)	0.201* (0.109)
A's Score	48.0 (0.0)	42.0 (4.8)	38.6 (1.3)	37.6 (0.9)	39.5 (0.9)	38.4 (0.6)	0.035 (0.076)
Lone Effort	42.3 (0.0)	39.3 (1.9)	35.8 (2.2)	34.8 (1.2)	37.6 (1.5)	35.8 (0.9)	0.074 (0.101)
Partnered Effort	40.0 (0.0)	38.0 (3.0)	35.6 (1.8)	34.6 (1.2)	36.6 (1.6)	35.4 (0.9)	0.054 (0.095)
Low	0% (0%)	50% (29%)	43% (13%)	55% (7%)	44% (10%)	50% (5%)	-0.018 (0.095)
Female	0% (0%)	75% (25%)	79% (12%)	49% (8%)	56% (11%)	55% (6%)	-0.066 (0.103)
Stage 4	100% (0%)	0% (0%)	50% (11%)	49% (3%)	56% (4%)	50% (0%)	0.087 (0.066)
n	1	4	14	77	36	132	132

Figures in parentheses ( ) are standard errors clustered by participant. \* correlation is significantly different from zero at the 10% level.