Optimism, Information, and Optimal Disinformation

Gigi Foster and Paul Frijters *

Keywords: behavioral economics, optimism, self-esteem, information policy
JEL classification: D83, D84, D02, D60

September 28, 2012

Corresponding author's details:
Gigi Foster
School of Economics, University of New South Wales
Sydney, NSW 2052, Australia
ph +612 9385 7472; fax +612 9313 6337

Abstract

Over-optimism is a ubiquitous aspect of real-life expectations. We suggest that over-optimism is related to both utility and choices, rather than being merely cheap talk. Our enhanced utility model includes an explicit role for expectation formation, and features costs associated with self-delusion. We predict and then verify using unique survey data that work effort, caring about the future, and over-optimism are all positively related. We derive the deep parameters of the utility function from social evolutionary processes. Finally, we explore the implications of our theory for the optimal information-revelation policy of fully-informed principals.

*University of New South Wales (gigi.foster@unsw.edu.au; corresponding author) and University of Queensland (p.frijters@uq.edu.au), respectively. We thank the participants in the 2012 Behavioral Economics workshop at UNSW, Bill von Hippel, and David Gill for helpful comments. We thank the University of South Australia and the University of Technology Sydney for their participation in the data extractions and surveys of students, as approved under the relevant Human Research Ethics Committee agreements. All errors remain ours. The authors also acknowledge the Australian Research Council (grants DP0878138 and DP0987840) for its support of this work.

1 Introduction

Why do individuals paint rosy pictures of the future, in which they will get promoted, become famous, or find complete happiness; in which their preferred sports team will dominate the tournaments; in which their children will grow up to become doctors and lawyers; and in which their country will win the next war? Similarly, why do our leaders tell us that we will win the next war, we will all get jobs, our national teams will win, and our country will go down in history as one of the greatest countries on earth?[1] Do instances of unrealistic optimism and optimistic political spin make any difference to our behavior, or are these phenomena irrelevant to our economic choices?

In traditional economic models, clearly unrealistic optimism is pure cheap talk which carries no cost and thus in no way should alter actual choices. Expectations themselves are not a choice, but the outcome of an unemotional calculus, so unrealistic optimism should not reflect 'actual' expectations. By the same token, politicians and leaders who spin optimistic stories that are rationally judged to be unlikely should be ignored.

A different approach has been taken in recent years by behavioral economists. The key argument has been that unrealistic optimism does reflect actual expectations of outcomes, and that expectations—whether or not they materialize—have a direct effect on experienced utility. The plausibility of this latter proposition is most clearly seen in the case of negative expectations: fear of a bad event can have negative consequences now, even if the feared event does not arise (Ohman & Mineka 2001). Recent evidence also shows that optimistic expectations about the future directly affect our current mood (Wulfert, Roland & Hartley 2005), and that the neural pathways that activate when we imagine a positive future are the same as those that activate when we experience a positive event now (Carlsson, Petrovic, Skare, Petersson & Ingvar 2000, Breiter, Aharon, Kahneman, Dale & Shizgal 2001). The essence of the proposition is that we are capable of directly consuming an imagined future.[2]

Yet, if the imagined future affects our utility, a set of follow-on questions arises. What

[1] The U.S. President's January 2012 'State of the Union' address contained the following illustrative examples: 'These reforms will help people get jobs that are open today'; 'America remains the one indispensable nation in world affairs—and as long as I'm President, I intend to keep it that way'. The address made by Pakistani President Asif Ali Zardari to the joint session of the Pakistani parliament on March 17, 2012, included the following examples: 'In spite of all the difficulties, the economy will grow by four per cent in 2012'; 'Together we are creating history'.

[2] This asserted capacity naturally requires an ability to imagine something that is not presently observable. Children who play at being superheroes and adults who play online fantasy games are obvious examples of this human ability to conjure up and draw pleasure from something that is unreal.

stops individuals from living completely in a perfect fantasy world? Why did evolution equip us with the option of consuming an imagined future? What is the relation between the over-optimism that this trait would support and other important traits and behaviors, such as talent and work effort?

In this paper, we take up the challenge of answering these questions. We suggest a reduced-form utility model in which individuals derive pleasure from their endogenous expectations of the future. Expectations themselves arise from a combination of objective signals and deliberate signal distortions, where distortions can result either from manipulation by others or from (rational) self-delusion. We derive the optimal level of distortions and illustrate the link between the importance of the imagined future to our utility today, and the tradeoffs that our model implies for inter-temporal effort choices. We compare the direct predictions of this very simple model with empirical findings from economics and psychology regarding optimism and self-delusion.

As a second step, we seek evolutionary reasons for engineering an individual to exhibit a tendency for over-optimism. We derive the circumstances in which Nature would optimally have programmed people to care more about the future, as well as the optimal costs of self-delusion that are borne by individuals. We argue that the central reason Nature would have programmed an individual with the ability to derive pleasure from an imagined future is that this motivates him to devote more effort towards obtaining a better future outcome. The central mechanism that yields trade-offs with regard to self-delusion is that a higher expectation on the part of the individual is also, via social interaction, a signal about the individual's potential that is relayed to others. A more favorable internal expectation about one's outcomes then gives rise to a more favorable evaluation by others, which in turn leads to better outcomes through social interactions, such as via mating opportunities. The optimal marginal cost of self-delusion that Nature would have chosen derives from comparing the socially-mediated benefits of possessing a higher self-evaluation to the personal costs of self-delusion.

As a third step, we consider the manipulation of individuals from the perspective of external parties. We ask what information policy would be optimally chosen by a principal who cares either about the utility of the individual, or else about the actual level of outcomes produced. The optimal amount of disinformation propagated by the principal arises from a trade-off between the benefits arising from the higher effort elicited if the individual is fed overly-optimistic signals about market payoffs to that effort, compared to the costs to the individual of exerting the effort.

We illustrate our arguments with existing results from the economic and psychology literatures, as well as original analysis of data from specially-designed questionnaires that allow us to measure and predict the self-delusion of university students facing different circumstances. Using these survey data, we directly validate some of the key predictions of the model, such as that abler individuals both self-delude more and exert more effort. We end by discussing the relation between over-optimism and the wider literature on self-esteem, and reflect on individual heterogeneity and the life-cycle aspects of self-delusion.

2 Optimism at the individual level

An individual i at time t is presumed to care about his imagined current outcome and his imagined future outcome, as follows:

$$U_{it} = \widetilde{Y_{it}} + \gamma \widetilde{Y_{it+1}} - e_{it} \tag{1}$$

Here, U_{it} is utility at time t, Y_{it} is the individual's actual outcome at time t, $\widetilde{Y_{it+1}}$ is the imagined outcome of the next period, and e_{it} is effort expended this period.[3] The main innovation in this simple two-period formulation is that the imagined present and the imagined future appear as direct consumption goods. This allows the possibility that the individual rationally self-deludes about both the present and the future. The actual outcomes we have in mind are the consumption levels an individual can pretend, both to himself and others, that he is currently enjoying and will enjoy in the future.

We interpret γ as the degree to which the imagined future matters to the individual, and $\gamma \widetilde{Y_{it+1}}$ as the contribution of future-savoring to today's utility. We interpret the inclusion of $\widetilde{Y_{it}}$ in the utility function mainly as a reflection of self-esteem: the better you believe your outcomes to be, the better you feel about yourself. Savoring the future is then a form of self-esteem that operates through imagining one's future.

To foreshadow our discussion of the possible objectives of Nature, one obvious evolutionary rationale for putting an imagined future into the utility function is that this provides a mechanism to motivate us to change our future, and thus to exert the effort required to avoid

[3] Y_{it}, the individual's actual outcome at time t, does not appear in this equation: utility rewards are based entirely on people's beliefs regarding reality, rather than on reality itself. This is consistent with the philosophical approach to understanding sensation, as described in Robinson (2006) (p. 750, italics in original): 'Our sensations, we may briefly say, *are* appearances. They may mislead about things other than themselves, but a "false appearance" of a sensation is really just *another* appearance'.

danger and seek advantages. The idea is that humans motivate themselves on a minute-by-minute level to make investments for the future by imagining themselves consuming the fruits of those efforts. Rather than needing to program us with some form of future-self-referential altruism, via which we care about our pleasure levels in the future because we consciously care about our future selves, Nature motivates us to work for the future by enabling us to partially consume that future now.

Equation 1 is also consistent with behavioral realism. A central tenet of psychology is that individuals experience the world as a mental state. Any distinction between present and future events in our minds is purely artificial: our internal representations of all events, no matter their chronology, can all be consumed with equal ease. This is consistent with neuro-scientific evidence that the pathways via which we conjure up and evaluate images of the present are the same as those that we use to conjure up and consider images of the future (Carlsson et al. 2000, Breiter et al. 2001).

Notably, Brunnermeier & Parker (2005) also present a model in which individuals obtain 'felicity' from an imagined future and will thus choose to over-estimate their next-period income. However, unlike in Brunnermeier & Parker (2005), our framework does not require that if one is overly optimistic about one thing, then one must also be overly pessimistic about another (which is implied by their condition that subjective probabilities must sum up to one). In our framework, it is perfectly possible that people think they obtain above-average outcomes in everything. Also unlike these previous authors, we include in our model a direct cost of maintaining self-delusion (detailed below), which is borne when one puts in effort to screen out information that runs counter to beliefs. More importantly, in our model individuals have optimistic expectations as a means of self-motivating for the future, whereas in the Brunnermeier & Parker (2005) model, agents think they will be lucky in the future and thus under-invest today. While we will therefore predict that over-optimism and effort move together, consistent with a large psychological literature, Brunnermeier & Parker (2005) predict the opposite. Our framework is also much more parsimonious, allowing us to explore evolutionary mechanisms for our model and address questions of optimal manipulation.

Equation 1 has the added appeal that it can be interpreted as a reduced-form variant of a simplified mainstream economic model in which the individual maximizes a discounted utility stream. Our formulation is equivalent to assuming that instantaneous utility in any period is a simple function of the perceived outcome that period plus a mean-zero error, such as when $U_{it} = u_{it} + \gamma u_{it+1} - e_{it}$ and $u_{is} = \widetilde{Y_{is}} + v$ where v has mean zero. The key point of distinction from normal discounted utility models is that in our model, the actual

outcomes Y_{it} do not directly appear in utility; instead, individuals can derive actual pleasure from things that not only have not yet happened, but that quite possibly will never happen at all. Also, imagined future instantaneous utility (u_{it+1}) directly yields experienced utility today, which is counter to the usual assumption that actual experienced utility depends on actual outcomes in the present period.

We argue that including imagined outcomes in the utility function is one way for Nature to give us forward-looking motives. The key difficulty Nature faces however is that the future has not yet happened, and individuals will therefore have an incentive to exaggerate their image of that future rather than work to make it better. Without a mechanism that provides bounds on our illusions, we can 'run off the rails' and consume an infinitely pleasurable imagined future at the expense of our actual future. How does Nature prevent us from just choosing $\widetilde{Y_{it+1}} = \infty$ and maximally indulging our laziness? To answer this, we need to specify how outcomes arise, how expectations arise, and what actually constitutes effort.

In our model, the outcome in any period t (referred to for convenience as 'output') is a function of previous work effort and talent:

$$Y_{it} = \alpha_i w_{it-1} + v_{it}$$

Here, $\alpha_i > 0$ is a measure of the marginal payoff to work (i.e., talent) for individual i, and v_{it} is a random error term with expectation zero. In any given time period t, w_{it} is the amount of work contributed which directly creates real (not imagined) output the next period. Hence, real output next period results from a combination of work contributed this period, plus influence from a random and mean-zero factor which is unaffected by work level. Individuals are furthermore presumed to know that this is the mechanism by which real output is created. For the moment, we will also presume that each individual i knows his true value α_i.

After choosing his optimal w_{it}, an individual is presumed to receive continuous signals about both his current reality and his expected future outcome. For now, we presume that these signals are unbiased. An individual at time t is therefore continuously bombarded with the true expected level of Y_{it} as a signal s^t_{it}, and with his true expected Y_{it+1} as a signal s^t_{it+1}. We argue that individuals can distort these signals, but that this distortion takes mental effort, somewhat akin to the notion of Compte & Postlewaite (2004). Individuals' internal representations of their present and future outcomes, which may be distorted away from reality, take the following forms:

$$\begin{aligned} \widetilde{Y_{it}} &= s_{it}^t + c_{it}^A = \alpha_i w_{it-1} + c_{it}^A \\ \widetilde{Y_{it+1}} &= s_{it+1}^t + c_{it}^B = \alpha_i w_{it} + c_{it}^B \end{aligned}$$

The mental mechanism we have in mind here is that an individual filters incoming signals by his choices of c_{it}^A and c_{it}^B, 'slanting' the truth in order to increase his utility. This enables the individual to believe in an imagined present and/or future that is better than reality would suggest.

Psychologists have a long history of documenting the existence of self-delusion that results in the biased processing of signals about the self. Various words are used to describe the psychological processes that support and give rise to this type of self-delusion, including cognitive dissonance, self-affirmation, and homeostasis (Festinger 1957, Steele, Spencer & Lynch 1993, Sherman & Cohen 2002, Walker 1956). Psychologists have argued that the unconscious mind, rather than the conscious mind, plays the key role in maintaining self-delusion. Working like an invisible blockade, the unconscious flatly refuses to process information that is damaging to the self-image, and will actively motivate the conscious mind to invent reasons why things are better than sensory data would suggest. The unconscious in this way continuously repairs the disconnect between the preferred mental image of the world, and the signals about that world as they come in (exactly in line with the process described in Diekhofa, Kipshagena, Falkaia, Dechentb, Baudewigb & Grubera (2011)).

Yet, self-delusion is not easy. Constructing and maintaining a favorable but unrealistic mental image of reality, which may involve rationalizations, over- or under-interpretations of sensory data, and the creation of convoluted storylines in which the self is the victor or praiseworthy figure, requires effort. The model is thus closed by specifying the determinants of total effort, e_{it}:

$$e_{it} = w_{it}^2 + \eta[(c_{it}^A)^2 + (c_{it}^B)^2]$$

η denotes the marginal mental cost of disinformation felt by the individual. Together with the parameter γ from Equation 1, it is one of the deep parameters of the model for which we will seek an evolutionary rationale. For now, we merely presume that $\eta > 0$.

2.1 Solutions

The model has simple optimal solutions for the three different varieties of effort chosen by the individual:

$$
\begin{aligned}
w_{it} &= \frac{\alpha_i \gamma}{2} \\
c_{it}^{A} &= \frac{1}{2\eta} \\
c_{it}^{B} &= \frac{\gamma}{2\eta}
\end{aligned}
$$

The first interpretation we can make is that the more his imagined future matters to an individual (i.e., the higher is γ) and the lower the marginal cost of self-delusion (i.e., the lower is η), the more the individual will delude himself that his future is going to be good. The more an individual (believes he) can improve the future at lower cost (i.e., the higher is α_i), the more that individual will invest in the future.[4] Self-delusion about the present increases as the cost of that self-delusion falls, and does not relate to the importance to the individual of the imagined future.

2.2 Empirical verification

This simple model above gives rise to the following predictions:

1. The utility an individual experiences this period increases with the imagined future, even conditional on any action taken today.

2. An individual engages in positive (i.e., optimistic) self-delusion about future outcomes, and this tendency increases the more important is the imagined future to his utility.

3. An individual has an optimistic interpretation of the present.

4. Real work effort increases with ability (or believed self-efficacy), α_i, and also increases with the degree to which the future matters (γ).

These predictions are all subject to empirical scrutiny, and extensive literatures already exist that are pertinent to some of them. We supplement our review of prior findings with original analysis of recent data from student surveys. These surveys were designed specifically to help test these hypotheses and related hypotheses about other possible associations between students' perceived social context and beliefs on their choices and outcomes. Our

[4]This is consistent with the hypothesis that individuals' locus of control and future-investment decisions are linked (Caliendo, Cobb-Clark & Uhlendoff 2010).

data are drawn from six surveys, administered in the middle of the first and second university semesters in 2008, 2009, and 2010 to two large populations of Australian undergraduate students enrolled in business-related programs of study. Each survey was comprised of a standard set of questions (including expected outcomes in each course in which the student was currently enrolled, self-rated ability, and self-rated effort) plus a selection of modules relevant to expectations and psychology. These include modules on self-esteem, savoring, and direct expectations of future outcomes. Actual outcomes, in the form of final percentages earned in different courses, were also merged into the survey data once each semester was complete. We draw on these data as needed in the empirical verifications below. Further description of the data is available from the authors upon request.

2.2.1 Prediction 1: Does utility increase with the imagined future?

Several recent papers by economists confirm that real welfare effects, as measured at the individual level by self-reported happiness or satisfaction questions, derive from the mere possession of beliefs that the future will be rosy. In cross-country data (Senik 2008), within particular countries (Frijters, Liu & Meng 2008), and with regard to specific types of expectations (Foster, Frijters & Johnston 2012, Wulfert et al. 2005), a more positive imagined future has been found to yield direct happiness gains today. The imagined positive future outcome—mainly, expected future income—is found to be more important (in terms of magnitude) than current outcomes, which can interpret as tentative support for the possibility that $\gamma > 1$.

Using our own survey data, we augment these previous findings by running a simple regression of agreement with the statement, 'Overall, I am very happy' (on a scale from 1 to 11) on average expected marks[5] in the university courses in which respondents were enrolled that semester. In this regression, we control for self-assessed general ability, self-assessed mathematical and verbal ability, and self-reported effort.[6] Results are shown in Table 1. We can see that the respondent's average expected mark in his classes has a statistically significant association with his overall happiness, whether or not we control for ability and effort.

[5]'Mark' is the Australian term for the final percentage earned in a university course, which varies from 0 to 100.

[6]Full descriptions of the variables used for this and the forthcoming original analysis in this paper can be found in the Appendix.

Table 1: Predicting overall happiness

	(1)	(2)	(3)	(4)
	Happy?	Happy?	Happy?	Happy?
Average	0.059***	0.053***	0.049***	0.044***
expected mark	(0.01)	(0.01)	(0.01)	(0.01)
Self-assessed		1.134***		0.717***
ability=average		(0.25)		(0.25)
Self-assessed		1.004***		0.347
ability=high		(0.27)		(0.27)
Self-assessed			0.247***	0.236***
verbal ability			(0.03)	(0.03)
Self-assessed			0.066**	0.071**
math ability			(0.03)	(0.03)
Self-assessed				0.215
effort=average				(0.16)
Self-assessed				0.487***
effort=high				(0.18)
Constant	3.962***	3.443***	1.831***	1.485**
	(0.58)	(0.60)	(0.61)	(0.65)
AdjR-sq	0.054	0.072	0.120	0.135
Obs	965	964	964	962

The dependent variable is the respondent's agreement with the statement, 'Overall, I am very happy' (on a scale of 1 to 11, with a mean of 8.13 and a standard deviation of 2.08 across the sample). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.2.2 Prediction 2a: Do individuals have unrealistically positive expectations of future outcomes?

There is a substantial economics literature documenting humans' tendency to be over-confident about their future likelihood of winning, dominating, or otherwise succeeding (Caliendo & Huang 2008, Scheinkman & Xiong 2003, Camerer & Lovallo 1999, Barber & Odean 2001). Psychologists have also found ubiquitous evidence of over-confidence about the future (Weinstein 1980). This effect is exemplified in our own data, where the average expected mark was 72.08 while the average actual mark was 65.39. This is a large and statistically significant difference. This tendency to be over-optimistic is also reflected in our evaluations of how likely we think our favorite sports team is to win a competition, and of the odds we place on our country winning a war.

A substantive example is given by the first World War. At the start of the conflict, young men from all over Europe enthusiastically signed up for service in the almost universal belief that their country would quickly win. As stated on p. 59 of Johnson (2008): 'On both sides, key decisionmakers, the military, and the public exhibited positive illusions about winning. Both sides thought that the war would be short, that they would win, and that they would benefit from it.' Most interestingly, this over-optimism did not just afflict those with little information: young men from elite circles volunteered in particularly large numbers. It took millions of deaths—i.e., an avalanche of signals about reality—to force both sides to accept mentally that the other side would not be as easily defeated as originally imagined, and for volunteers to become scarce.

2.2.3 Prediction 2b: Are unrealistically optimistic expectations of the future even higher for those to whom future expectations are especially important?

In Table 2, we show the results of a regression of respondents' expected marks in their courses ($\widetilde{Y_{it+1}}$). These self-reported prospects are regressed on various measures of ability and effort, as well as a measure of the self-reported utility value of imagining future events. This measure is constructed from respondents' answers to questions about what is termed 'future savoring', meaning the extent to which the respondent draws pleasure from imagining good outcomes in the future (see the Appendix for details). Importantly, we are able to include the actual mark that the respondent eventually received for the given subject, meaning that the variation in expected marks that is predicted by other independent variables (such as future savoring) is best interpreted as the respondent's degree of over-optimism. The analysis level

is the student-course match.

In the first two columns of Table 2, we see a significant positive association of future savoring (which we interpret as a measure of γ) with expected future outcomes, even controlling for the actual mark received at the end of the semester. This effect remains significant even when our full array of ability controls is included in Column 5, although it falls by about 30%, indicating that abler students hold more optimistic expectations about their future outcomes, which also concurs with the predictions of the model. We find across the specifications that abler students, those who work harder, and those who have more optimistic images of themselves in the present (i.e., those with higher self-esteem) all expect better outcomes.[7]

2.2.4 Prediction 3: Do individuals exaggerate their outcomes in the present?

The human tendency to rate the current situation as above average has been termed the 'Lake Wobegon effect', based on the saying from an American radio program that 'all the women are strong, all the men are good-looking, and all the children are above average.'[8] As the evidence reviewed in Dunning, Heath & Suls (2004) shows, the vast majority of individuals who respond to surveys report to think of themselves as above-average drivers, lovers, students, managers, and so forth (Alicke & Govorun 2005).

While we do not directly explore this mechanism in our model, the psychological underpinning of this phenomenon is frequently connected with our need to maintain some minimum level of self-esteem. Through this pathway, our tendency towards self-delusion has the potential to explain many otherwise puzzling aspects of human culture. In particular, in search of self-esteem, people at society's lowest levels are susceptible to swallowing storylines that allow them to avoid the truth of their low standing. A good example of this type of storyline is the salvation story that most large religions propagate: some unseen entity (such as god or karma) cares even for those at the bottom of the ladder, and will reward them for their piety in an afterlife.

One might surmise that such a story would have been concocted by rulers in order to pacify individuals at the bottom, but this is not historically correct. Individuals will voluntarily adopt a salvation story without encouragement from the top. An excellent example of this is the adoption of Christianity by slaves in ancient Rome, contrary to the wishes of

[7] We refer here to the total effects of an across-the-board increase in the ability variables. We attribute the negative conditional association of verbal ability with expected mark as an outcome of including several correlated measures of ability at once.

[8] As quoted on p. 14 of Goldman (2008).

Table 2: Predicting expected mark in a course

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	E(mark)	E(mark)	E(mark)	E(mark)	E(mark)	E(mark)	E(mark)
Future savoring	0.490***	0.368***			0.245*	-0.046	-0.040
	(0.14)	(0.14)			(0.14)	(0.17)	(0.19)
Actual mark		0.241***		0.214***	0.211***	0.198***	0.198***
		(0.01)		(0.01)	(0.02)	(0.02)	(0.02)
Self-assessed			5.398***	4.723***	4.306***	3.382***	3.370***
ability=average			(0.69)	(0.83)	(0.90)	(0.83)	(0.81)
Self-assessed			9.165***	7.193***	6.693***	5.099***	5.080***
ability=high			(0.69)	(0.83)	(0.89)	(0.84)	(0.84)
Self-assessed				0.349**	0.239	0.163	0.132
math ability				(0.15)	(0.18)	(0.19)	(0.20)
Self-assessed				-0.226**	-0.215*	-0.252**	0.055
verbal ability				(0.10)	(0.12)	(0.11)	(0.14)
Self-assessed						1.706***	1.656***
effort=average						(0.54)	(0.54)
Self-assessed						2.768***	2.630***
effort=high						(0.59)	(0.61)
Self-esteem						0.633***	0.613***
						(0.15)	(0.16)
Dummy for							1.496**
intl student							(0.66)
Dummy for							0.510
NESB student							(0.57)
Dummy for							-0.657
female							(0.42)
Age							-0.000
							(0.04)
Dummy for							2.224***
new student							(0.68)
Constant	68.057***	53.332***	65.946***	52.403***	51.814***	50.569***	47.524***
	(1.35)	(1.42)	(0.64)	(1.38)	(1.75)	(1.74)	(2.18)
AdjR-sq	0.003	0.058	0.042	0.088	0.075	0.084	0.090
Obs	4610	4610	7327	5531	4586	4537	4277

The dependent variable is the expected mark in a given course; analysis is at the student-course level. Dummy variables included in the final column indicate international student status; status as a student of non-English speaking background; status as a female; and status as a student experiencing his or her first semester at the university. Discipline of the course is also controlled in the final column using eight dummy categories. See Appendix for a description of the construction of the future savoring and self-esteem variables. Sample size fluctuations are due to the fact that not all question modules appeared in every survey; all useable responses are used to estimate each specification. Standard errors are clustered at the student level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the Roman elite. Despite their purges, the Romans failed to eradicate the salvation ideology adhered to by their slaves, and eventually even adopted Christianity for themselves during the reign of Constantine in the 4th century AD.[9] Not even the threat of torture and death outweighed the self-esteem benefits to a slave of believing that someone unseen loved him.[10]

2.2.5 Prediction 4: Does real effort increase with ability and the degree to which the future matters?

To test this implication using our survey data, we employ as a measure of real effort students' responses to the question of whether the respondent works hard in class. This effort measure is regressed on a battery of individual ability measures, as well as future savoring and self-esteem measures. Results are shown in Table 3.

These results show a positive and significant association of ability with effort, along with a positive and significant association of future savoring with effort. This latter effect loses size and significance when we also include a measure of self-esteem, consistent with the argument that over-confidence is mainly about the maintenance of self-esteem (i.e., the two concepts are largely the same). It appears that those who have higher self-esteem conditional on our ability measures—in other words, those who are more optimistically deluded about their present worth or efficacy—also feel higher utility effects from optimistic views of the future.

3 Evolutionary mechanisms

We now step back and ask why, in terms of evolutionary fitness, our human psychology might have developed along the lines that we depict in our model.[11] There are two main

[9] See p. 239-240 of Ekstrand (2008) for a brief history of the lead-up to and Constantine's role in the Roman Empire's adoption of Christianity, in 313 AD.

[10] The connection of Christian belief to self-esteem amongst the poor is also explicitly discussed in Keitlen (2004) (p. 53). Morton & Olenik (2005) state, with reference to modern Japan: 'Christianity retains an appeal to marginal groups including *burakumin* or "outcasts, prostitutes, and the homeless".' (p. 280).

[11] Johnson & Fowler (2011) have a similar objective. That paper finds a different role for over-confidence, which is that it helps in competitive situations where individuals play a simple heuristic strategy (fight-or-flight) based on comparisons between point estimates of own ability versus the perceived ability of the other (which is true ability plus an error). Over-confidence then serves as a means of engaging in more conflicts, effectively producing for oneself a high level of positive measurement error in the snapshots taken by others. This is optimal when there is a large potential surplus from winning the conflict and when there is no escape from the (irrational) heuristic followed.

Table 3: Predicting hard work in classes

	(1)	(2)	(3)	(4)	(5)
	Work hard?	Work hard?	Work hard?	Work hard?	Work hard?
Self-assessed	0.692***	0.333**		0.036	-0.229
ability=average	(0.15)	(0.16)		(0.19)	(0.19)
Self-assessed	0.866***	0.285		0.107	-0.290
ability=high	(0.16)	(0.18)		(0.21)	(0.21)
Self-assessed		0.046**		0.020	-0.009
verbal ability		(0.02)		(0.03)	(0.03)
Self-assessed		0.161***		0.138***	0.115***
math ability		(0.02)		(0.02)	(0.02)
Future savoring			0.178***	0.154***	0.048
			(0.03)	(0.03)	(0.03)
Self-esteem					0.248***
					(0.04)
Constant	7.020***	5.762***	6.313***	5.190***	4.874***
	(0.14)	(0.25)	(0.26)	(0.35)	(0.35)
AdjR-sq	0.009	0.034	0.019	0.039	0.062
Obs	2888	2621	1791	1783	1764

The dependent variable is stated agreement with the statement, 'I work hard in tutorials' (on a 1-to-11 scale, with a mean of 7.71 and a standard deviation of 2.27 across the sample). See Appendix for a description of the construction of the future savoring and self-esteem variables. Sample size fluctuations are due to the fact that not all question modules appeared in every survey; all useable responses are used to estimate each specification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

phenomena for which we seek evolutionary reasons. First, why would Nature have equipped us with a forward-looking mechanism? Second, why would Nature have included costly self-delusion as a negative input to utility? To answer these questions, we take the paternalistic perspective of Nature maximizing some notion of surplus at the individual level.

Nature's social welfare function includes actual outcomes and expectations of outcomes, i.e. appearances:

$$SW = \delta(Y_{it+1} - e) + (1-\delta)(\widetilde{Y_{it+1}} - e)$$

This social welfare function is a weighted sum of actual surplus resources $(Y_{it+1} - e)$ and imagined surplus resources $(\widetilde{Y_{it+1}} - e)$. Various underlying mechanisms would logically give rise to this social welfare function, and we discuss these below. For now, we merely note that when $\delta = 1$, Nature only cares about the individual's actual surplus, while in the case that $\delta = 0$, Nature only cares about the appearance of surplus.

3.1 Solutions

Given this objective function, we first insert our optimal effort solutions at the individual level, thereby obtaining the actual level of SW:

$$SW = \delta(\frac{\alpha_i^2\gamma}{2} + v_{it}) + (1-\delta)(\frac{\alpha_i^2\gamma}{2} + \frac{\gamma}{2\eta}) - \frac{\alpha_i^2\gamma^2}{4} - \frac{1}{4\eta^2} - \frac{\gamma^2}{4\eta^2}$$

We then find the following solution equations for the key parameters η and γ by maximizing the mathematically expected level of this social welfare function:

$$\begin{aligned} \frac{\partial E\{SW\}}{\partial \eta} &= -(1-\delta)\frac{\gamma}{2\eta^2} + \frac{\gamma^2+1}{2\eta^3} = 0 \\ \frac{\partial E\{SW\}}{\partial \gamma} &= \frac{\alpha_i^2}{2} + \frac{(1-\delta)}{2\eta} - \frac{\alpha_i^2\gamma}{2} - \frac{\gamma}{2\eta^2} = 0 \end{aligned}$$

These solution equations have the appropriate negative second derivatives, and they obey the conditions that $\frac{\partial E\{SW\}}{\partial \eta}\big|^{\lim \eta \downarrow 0} > 0$ and $\frac{\partial E\{SW\}}{\partial \gamma}\big|^{\gamma=0} > 0$, meaning that there is an optimal solution where both η and δ are larger than zero. This in turn yields the solutions

$$\begin{aligned} \eta &= \frac{2(\gamma^2+1)}{\gamma(1-\delta)} \\ \gamma &= \frac{\frac{\alpha_i^2}{2} + \frac{(1-\delta)}{2\eta}}{\frac{\alpha_i^2}{2} + \frac{1}{2\eta^2}} \end{aligned}$$

which reduce to

$$\begin{aligned}\eta &= \frac{2(\gamma^2+1)}{\gamma(1-\delta)} \\ \gamma &= \frac{\alpha_i^2\eta^2+\eta(1-\delta)}{\alpha_i^2\eta^2+1} = 1+\frac{\eta(1-\delta)-1}{\alpha_i^2\eta^2+1} = 1+\frac{\frac{2(\gamma^2+1)}{\gamma}-1}{\alpha_i^2\frac{2(\gamma^2+1)}{\gamma(1-\delta)}\frac{2(\gamma^2+1)}{\gamma(1-\delta)}+1}\end{aligned}$$

While there is no closed-form analytical solution to this system of equations and uniqueness is not guaranteed, it is possible to gauge something about the comparative statics by first noting that η is always above $\frac{4}{1-\delta}$, which in turn means that $(\eta(1-\delta)-1>0)$ and that $\gamma > 1$. We can then note that in this range of values, η is increasing in γ while γ is decreasing in α_i, in turn implying that η is decreasing in α_i. Also, in this range, η increases in δ, and γ decreases in δ. We can thus derive the following comparative statics results:

$$\frac{d\eta}{d\delta} > 0, \frac{d\eta}{d\alpha_i} < 0, \frac{d\gamma}{d\delta} < 0, \frac{d\gamma}{d\alpha_i} < 0$$

Intuitively, the reasons for these relationships are as follows. When appearances matter less (δ is higher), it is optimal for Nature to minimize the effort we put into self-delusion by increasing the marginal cost of self-delusion. As talent (α_i) increases, the evolutionary payoff to appearances also decreases since the future outcome will be higher anyway. Therefore, the optimal marginal cost of self-delusion (η) is higher, and the value of future consumption (γ) is lower, and for both of these reasons the total effort spent on self-delusion is lower. At the margin, it also holds that the higher the talent of the individual and the less that appearances matter in terms of evolutionary selection (δ), the less important the imagined future becomes for the individual.

It remains an open question whether Nature actually has the ability to set different η and γ values for individuals with different talents, or to differentiate across people or populations in terms of the importance of appearances (δ). Since Nature has an incentive to optimize α_i in all states of the world, there may in fact be no evolutionary mechanism through which low-talent individuals would acquire lower levels of η and γ: they would simply die out. Some of these comparative statics results are thereby more likely to hold across whole humanoid species or sub-populations. In this case we can interpret α_i as the relative importance, in terms of affecting ultimate evolutionary success, of influencing actual outcomes in the area or niche inhabited by the sub-population to which i belongs. Where the environment allows for more real manipulation (i.e., where α_i is higher), then η and γ should be lower (i.e., closer to 1) because those unable to exploit the opportunities to gain advantages presented

in such environments would have died out. Given the short length of time that human sub-populations have had time to diverge, however, significant deviations in η and γ across humans would seem unlikely.

An important question is whether Nature has other mechanisms at her disposal to give individuals an incentive to work for the future. In particular, why not make the utility cost of actual productive effort (w_{it}) lower? The answer within this model is that the cost of productive effort is the numeraire, and that all other costs are relative to that cost. The 'deep' parameters that Nature cannot increase and that thereby force a role for self-delusion are then α_i, which denotes the marginal productivity of real effort, and δ, which is the degree to which appearances matter. These arise from the competition with other species and the limits on social learning that lie outside the model.

3.2 Special cases and micro-models of Nature's welfare function

An important special case to consider is when $\delta = 1$ and Nature only cares about actual surplus, which can be viewed as the pure odds of the individual's own survival. In that case, $\eta = \infty$ and $\gamma = 1$ for any level of talent, because these values give the individual optimal incentives to invest in the future. This reinforces the central argument that the evolutionary role of the $\gamma\widetilde{Y_{it+1}}$ term in the utility function is to give the individual an incentive to invest in the future; maximization of actual surplus by itself implies no role for self-delusion.

Another important special case is when $\delta = 0$ and Nature therefore cares only about our appearances. In that case, $\eta > 4$ and $\gamma > 1$. The reason to nevertheless set a high marginal cost of signalling in this case is that there are two opportunity costs created by signalling: that borne when one self-deludes about the future, and that borne when one self-deludes about the present.[12]

This second special case captures social mating dynamics. It includes the case where, at time t, individuals have to pair up in mates and the only signal sent to potential mates about a given individual is his self-assessed surplus, i.e., $(\widetilde{Y_{it+1}} - e)$. In such a situation, and presuming that in all cases the actual surplus increases in the level of talent α_i, both we and

[12]If we change the basic utility function to $U_{it} = Y_{it} + \gamma\widetilde{Y_{it+1}} - e_{it}$, then qualitatively all the results go through, though the new boundary solution would be that $\eta > 2$. The reason why a positive cost of self-delusion remains is that there is an alternative method of signalling (via work effort), and the net total costs of signalling rise quadratically as the marginal cost η falls linearly due to the quadratic in the effort function. We have explored versions in which Nature cares about a function of α_j, but this does not change the basic results.

our potential mates would assortatively match on this imagined surplus. This corresponds to the classic case where $(\widetilde{Y_{it+1}} - e)$ denotes a costly form of signalling about own talent α_i. This case can therefore be interpreted as the reduced-form state of the world in which Nature maximizes the expected value α_j of our mate j.

This also allows for a structural interpretation of the assumption that $SW = \delta(Y_{it+1} - e) + (1 - \delta)(\widetilde{Y_{it+1}} - e)$. In particular, this framework captures the situation where there is a probability δ that our potential mates can observe our actual future outcome (or equivalently, our current outcome, from which they could perfectly infer our future outcome), and a probability $(1 - \delta)$ that they can observe our self-assessed outcome.

The crucial assumption behind the argument that $\widetilde{Y_{it+1}}$ plays a role in mating games is thus that with some positive probability, potential mates must rely on $\widetilde{Y_{it+1}}$ as a signal of our abilities. Evolutionary psychologists have engaged in long non-mathematical debates about exactly this issue, where it has also been argued that the fundamental reason for individuals to care about their self-esteem and appearances in general is because others can observe what we think about ourselves and use that information in mate selection decisions. The empathic ability of humans to gauge others' beliefs, including to sniff out lies about those beliefs, then forces us to delude ourselves in order to be able to fool others (von Hippel & Trivers 2011).

3.3 Predictions

There are again important empirical regularities predicted by these arguments that we can confront with data.

1. Self-delusion is more important in periods when mating considerations are more relevant. For humans, this means that self-delusion should peak after puberty and during early adulthood.

2. We should see more display of appearances in animal populations with limited opportunities to observe the actual outcomes of other potential mates (i.e., where δ is likely to be low), while we should see less display of appearances in animal populations with more opportunities to observe the actual outcomes of potential mates (i.e., where δ is likely to be high).

On the first of these predictions, indirect support comes from cognitive psychology research on the development of the brains of young adults. The emotional maturation process

in adolescence has been found to be affected strongly by both physical development and social interactions (Susman, Nottelmann, Inoff-Germain, Dorn, Jr., Loriaux & Chrousos 1985, Marston, Hare & Allen 2010, MacDermott, Gullone, Allen, King & Tonge 2010). Young men in particular during this period have been found to be over-confident, and this has been argued to directly relate to their attempts (peaking in the young-adult stage of life) to impress young women with acts of bravery and prowess (Wilson & Daly 1985).

On the second of these predictions, we refer to observations of mating display activities amongst animals, where the most elaborate mating rituals are seen in bird species whose female and male do not nest together (with the Bird of Paradise (Frith & Beehler 1997) being the most ridiculous example). Kroiss, Lechner & Strohm (2010) review and add to the evidence that male courtship displays are strongly related to the distribution of females across time and space, and indicate that in species where so-called 'leks' develop (mating arenas in which each of a group of males carves out a territory to which he attempts to lure the female), prolonged interaction between the sexes is rare: 'Males usually employ visual, acoustical, or chemical displays as well as physical fighting to establish a hierarchy and/or to attract females ... [m]ales do not provide any paternal care, and females visit these leks only to choose a mate based on the display and to copulate.'

4 Optimal information policy

If we take the basic utility function proposed above as given, complete with its evolutionary rationale, a question that arises in any social application is how the individual can be manipulated towards the ends of others. More particularly, how would an informed outside agent, with his own independent objective function, optimally slant the information fed to the individual for his own purposes?

Consider as the principal's welfare function the following:

$$W = \epsilon U_{it} + (1 - \epsilon) Y_{it+1} - f$$

where $0 \leq \epsilon \leq 1$ and f denotes the effort put in by a principal. This welfare function includes cases ($\epsilon = 0$) in which the principal only derives some benefit from the actual outcome Y_{it+1}, which would hold for a tax collector or the owner of the business for whom i works. This welfare function also includes cases ($\epsilon = 1$) in which the principal is only concerned with the experienced well-being of the individual U_{it}. Prime examples of this case would be friends, concerned parents, a benevolent government, or the reflective rational part of the individual

himself.

The principal is presumed to have the ability to manipulate the signal s^t_{it+1} that an individual obtains about the future in a particular way:[13]

$$s^t_{it+1} = (\alpha_i + \kappa_1)w_{it} + \kappa_2$$

This distorted signal has two distinct components. The first component, κ_1, is a distortion of the actual marginal payoff of the productive work put in by an individual, for example through deliberately exaggerating the likelihood that an individual will get a promotion if he works harder. Another example of this type of signal distortion would be a politician promising more benefit from a costly policy than is actually likely to materialize. The second distortion, κ_2, is a pure distortion of the actual outcome independent of any work, and can therefore be interpreted as the degree to which 'luck' is seen to lead to a positive outcome.

The cost of manipulation faced by the manipulating principal is the personal effort put into deliberate misinformation of each type:

$$f = d_1\kappa_1^2 + d_2\kappa_2^2$$

Here, d_1 and d_2 are not given by Nature but by the information technology present in human society. They will be high when individuals obtain information from many overlapping sources which thereby provide some defense against manipulation, and low when there are few sources of information available to individuals, leading them to be easily manipulated.

4.1 Solutions

We start with computing the actual outcomes U_{it} and Y_{it+1} as a function of the choices κ_1 and κ_2, which follow from simply applying the steps in the original model with these updated signals:

$$\begin{aligned} U_{it} &= \widetilde{Y_{it}} + \gamma\widetilde{Y_{it+1}} - e_{it} = \widetilde{Y_{it}} + \gamma\{\frac{(\alpha_i + \kappa_1)^2\gamma}{2} + \kappa_2 + \frac{\gamma}{2\eta})\} - \frac{(\alpha_i + \kappa_1)^2\gamma^2}{4} - \frac{1}{4\eta^2} - \frac{\gamma^2}{4\eta^2} \\ Y_{it+1} &= \alpha_i w_{it} + u_{it} = \frac{\alpha_i(\alpha_i + \kappa_1)\gamma}{2} + u_{it} \end{aligned}$$

If we now presume that the principal is rational, and hence find the optimal solution by taking the expected value of W, we recover

[13] It is also possible to allow the principal to manipulate the signal an individual receives about the present, but that case turns out to be equivalent to a special case of manipulating the signal about the future.

$$E\{W\} = \epsilon[\frac{(\alpha_i + \kappa_1)^2\gamma^2}{2} + \gamma\kappa_2 - \frac{(\alpha_i + \kappa_1)^2\gamma^2}{4}] + (1-\epsilon)[\frac{\alpha_i(\alpha_i + \kappa_1)\gamma}{2}] - d_1\kappa_1^2 - d_2\kappa_2^2 + \sigma(\widetilde{Y_{it}}, \gamma, \alpha_i, \eta)$$

where $\sigma(\widetilde{Y_{it}}, \gamma, \alpha_i, \eta)$ is a residual term that does not depend on κ_1 or κ_2.

If we maximize this function over possible choices of κ_1 and κ_2,[14] we recover

$$\begin{aligned} \kappa_1 &= \frac{(1-\epsilon)\frac{\alpha_i\gamma}{2} + \epsilon\frac{2{\alpha_i}^2\gamma^2}{4}}{2d_1 - 2\epsilon\frac{\gamma^2}{4}} \\ \kappa_2 &= \frac{\epsilon\gamma}{2d_2} \end{aligned}$$

This tells us that the principal will only apply a slant to information ($\kappa_2 \neq 0$) in proportion to the degree to which he is concerned with the actual utility level of the individual (through ϵ). More interestingly, the principal will deliberately bias information about payoffs to work investments even when he only cares about the utility of the individual (i.e., when $\epsilon = 1$), in order to increase the degree to which an individual receives utility from an improved imagined future. It is now clear that this principal could be equally thought of as the rational part of an individual himself, which would actively encourage self-delusion of other parts of the individual in order to 'self-medicate'. Similarly, friends and benevolent parents would not merely filter out negative information about us, but would actively encourage us to believe that our efforts have higher payoffs than they actually have.

By contrast, a more selfish principal who only has a stake in the outcome (i.e., for whom $\epsilon = 0$) would not slant information in a positive direction (hence $\kappa_2 = 0$), but would exaggerate the perceived payoffs of work up until the point where $\kappa_1 = \frac{\alpha_i\gamma}{4d_1}$.

4.2 Predictions

The stylized predictions from this analysis are as follows.

1. Even on reflection, an individual would rationally choose to be misinformed by others about both the level of his future outcomes and the impact of his efforts in bringing about those outcomes. Yet, this disinformation is not infinite, since there is an added cost to believing that one's work efforts have more impact than they do in reality.

[14] We do not consider here the out-of-bounds case where $4d_1 < \epsilon\gamma^2$, in which case the principal would infinitely bias information for the benefit of the individual.

2. Principals with a stake in our outcomes would misinform us about the payoffs to our efforts, whereas those with a stake in our feelings would misinform us both about effort payoffs and about the degree to which factors outside our control will favor us. It is our friends, family, and rational selves who should be observed to assure us that 'luck is with us'.

As evidence in support of the first of these predictions, we use results from two questions on our student surveys asking respondents to choose between two situations. In the first question, Situation A involved receiving signals that a favored sports team or artist was likely to win a competition, followed by that favored team or artist losing. Situation B involved receiving signals that the favored team or artist was not going to win, followed by the predicted loss. In the second question, Situation A again involved receiving positive signals, but the outcome too was a win; Situation B involved receiving negative signals instead, in advance of an unexpected win.

Because each respondent answered one pair of these scenario questions (either two questions dealing with sports, or two questions dealing with arts, both of which appear in full in the Appendix), we observe two choices for each respondent based on this question set. Our measure of the degree to which people derive pleasure from anticipation of a positive event is a simple binary indicator that divides respondents into those who chose the high-expectations option (Situation A), versus those who chose the low-expectations option (Situation B). We find that 77 percent of respondents prefer to be fed an optimistic version of the future when the future outcome is actually going to be bad (in other words, they prefer to be lied to); and 63 percent of respondents prefer to be fed optimistic signals when the outcome is actually going to be good. This latter statistic shows that on average in our sample, the pleasure from being pleasantly surprised that may result from holding lower expectations than are warranted by reality is dwarfed by the pleasure of consuming positive expectations today.

In support of our second prediction regarding manipulation, the benefits of flattery (a type of 'ingratiation' in the terminology of social psychology) in producing desired behaviors are well-documented. Not only adults (Seiter 2007) but even children as young as four years old (Fu, Evans, Wang & Lee 2008) employ disinformation techniques that involve the deception of others.

4.3 Further reflections

Some further points can be made about how the basic mechanisms in this paper relate to other literatures.

4.3.1 Self-delusion, self-esteem and levels of consciousness

As indicated previously, one way to interpret the 'imagined future' is to equate it with self-esteem. A large literature has argued for the importance of self-esteem in politics, business, science, and personal life (Brennan & Pettit 2004). While that literature has to our knowledge not made connections to evolutionary arguments and the optimal strategies of manipulators, it has pointed to the incentives for self-delusion when self-esteem is important.

A crucial point of difference between this existing literature and our model is that in models of self-esteem, it is necessary to invoke different levels of rationality wherein a 'Self 0' knows more about the actual traits of an individual than a 'Self 1' (see, e.g., Benabou & Tirole 2006, Carrillo & Mariotti 2000). Such constructions have an inherent difficulty, as they require either the disappearance of intelligence over time, or at a minimum, different layers of consciousness. Our model, on the other hand, allows for individuals to have constant mental traits over time. The key feature of our model that allows for this is that we do not presume that individuals are maximizing discounted streams of utility. Instead, we depict individuals as maximizing current utility, with imagined future outcomes included as part of that utility.

This crucial choice means that it is not necessary, or optimal, for an individual to know at any time what 'the truth' actually is. An individual is a continuous self-deluder, continuously distorting incoming signals without necessarily ever being aware at some 'higher level' that the self-delusion is happening. Our approach also more naturally accommodates the notion that rational self-reflection must be learned, and is unusual from an evolutionary perspective. Indeed, from an evolutionary perspective, learning the true psychology of humans is both unfortunate from a utility perspective and costly in an evolutionary sense, to the extent that the knowledge itself makes it harder to subsequently self-delude.

4.3.2 Over-confidence of managers

Following the arguments above, we should expect to see that managers and other professionals in charge of producing predictions of future outcomes will exhibit over-confidence, particularly when there is little other available information. Managerial over-confidence has been well-documented (Billett & Qian 2008). Like male sexual displays in the wild, over-confidence in management may serve as a substitute signal of own ability, and therefore something that will attract rewards. Jorgensena, Teigenb & Molokkena (2004) hypothesize in this vein: 'the observed problems of [over-confidence]... may be caused by the estimators' desire to show competence and to provide useful information'.

The best evidence that managerial over-confidence is greater in situations where outside information is weaker comes from studies of the relationship between certainty and over-optimism, such as those reviewed in Barros & da Silveira (2008). In cases where a process is highly uncertain, managers tend to exhibit more over-confidence about the outcome of that process than in cases where the process is simpler, and hence where its outcome is easier to foresee based on information other than the manager's own prediction.

5 Discussion

In this paper, we suggest a simple micro-model whose solution implies that over-optimism is related to both experienced utility and actual choices, rather than being merely cheap talk. We include the imagined future as a consumption good, thus creating a primary role for expectation formation. Costs associated with the self-delusion required to sustain unrealistically high expectations generate the prediction that work effort, caring about the future, and over-optimism move together. We study our model as the result of evolutionary processes wherein individuals face a tradeoff between costly self-delusion and the social signalling advantage of self-belief. We supplement our discussion with supporting survey evidence, real-world examples, and a discussion of related literatures in social science. We conclude with a discussion of our model's implications for the optimal information-revelation policy of fully-informed principals.

The two main arguments underpinning the model in this paper are as follows. First, the capacity for direct consumption of an imagined future is the most likely way through which Nature equipped us with an incentive to care about the future. Second, through our social interactions, our own estimate of our future is an important signal received by others about our abilities, giving us a socially-grounded reason to be over-confident. We argue that the combination of these two factors rationalizes the pervasive over-optimism people display both about their future and about their present, and explains why poor people are prone to believe in salvation stories; why self-delusion must occur at the unconscious level (i.e., so that we are unaware of the delusion, which enables it to have a social signalling benefit); why more talented individuals are less over-confident; why effort and self-delusion go hand in hand; and why both our reflective selves and our leaders encourage us to think better of ourselves than is rational.

While they may improve our understanding of society, the policy implications arising from our paper are not clear at all, even though we have given explicit rules for optimal

disinformation depending on the utility function of the principal. The key difficulty is the unconscious element of self-delusion: the benefits of self-delusion arise only as long as an individual truly believes the illusion. How can policy then be arrived at via open discussion of how to optimally misinform people? Merely discussing the question of how best to lie to people defeats the purpose of lying, which is to allow people benefits they can only enjoy if they do not know they are being lied to. This is a general conundrum raised by the strong likelihood that populations derive great psychological and evolutionary benefits from untruths. In what forum can the benevolent manipulator discuss manipulation, and does it even make sense to do so, since there will be an inevitable loss of long-run reputation?

This issue gets particularly fraught if we consider the other side of this debate from a philosophical perspective. It is normally the recognized role of science to uncover and denounce misinformation: the normal moral stance of scientists, from which much of our standing and social usefulness derives, is that the truth is better than a lie even if the truth is unpopular. To maintain this role, we would argue that social scientists should not discuss the actual manipulation of individuals in any particular case, but should take as our limit a discussion of the mechanism and general rules for optimal disinformation. Particular principals can then implement this knowledge in particular cases, without social science being seen to approve.

6 Appendix

In this Appendix we briefly describe the survey questions that we used to construct the psychological measures and outcomes for the original analysis in this paper.

6.1 Ability and effort measures

One vector of innate academic abilities is constructed from students' self-rated levels of fluency with English and mathematics separately, on 1-to-11 scales of agreement to disagreement with positive statements. We also capture self-rated relative academic ability and effort on 4-level comparative scales, where the comparison group is other students at the respondent's university. Finally, as an alternative and more continuous measure of effort, we have responses to a question about how hard the student claims to work in discussion sections (termed 'tutorials' for interpretability by this population). The relevant questions are as follows:

My overall fluency in English - speaking, writing, reading, and understanding - is excellent.

My overall fluency with mathematics, statistics, and numbers generally - interpretations, manipulations, and illustrations of them - is excellent.

I work hard in tutorials.

Overall, with respect to academic work, would you rate yourself as(choose one)

- Not as smart/capable as other UniSA/UTS students
- About as smart/capable as other UniSA/UTS students
- More smart/capable than other UniSA/UTS students
- Much more smart/capable than other UniSA/UTS students

Overall, with respect to academic work, would you rate yourself as(choose one)

- Not as hardworking as other UniSA/UTS students
- About as hardworking as other UniSA/UTS students
- More hardworking than other UniSA/UTS students
- Much more hardworking than other UniSA/UTS students

For these latter two questions, because very few students self-rated themselves as 'much more' able or hardworking than others, we aggregate responses in categories 3 and 4, yielding a three-level measure of self-rated relative ability and a three-level measure of self-rated relative effort.

6.2 Savoring

We measure heterogeneity in individuals' direct utility benefits from (1) the anticipation of good events, (2) good events in the present, and (3) good events in the past, based on their answers to a battery of questions adapted from Bryant & Veroff (2006) to measure past, present, and future 'savoring'. The list of questions, each of which was answered on a 1 (strongly agree) to 11 (strongly disagree) scale, is as follows:

1. I feel a joy of anticipation when I think about upcoming good things.
2. I can make myself feel good by remembering pleasant events from my past.
3. When I reminisce about pleasant memories, I often start to feel sad or disappointed.
4. I can't seem to capture the joy of happy moments.
5. It's hard for me to get very excited about fun times before they actually take place.
6. I feel fully able to appreciate good things that happen to me.
7. I can make myself feel good by imagining what a happy time that is about to happen will be like.
8. I don't enjoy things as much as I should.
9. It's easy for me to rekindle the joy from pleasant memories.
10. When I think about a pleasant event before it happens, I often start to feel uneasy or uncomfortable.
11. It's easy for me to enjoy myself when I want to.
12. For me, once a fun time is over and gone, it's best not to think about it.

After reverse-coding questions 1, 2, 6, 7, 9, and 11, we take the average of responses to questions 1, 5, 7, and 10 to measure future-savoring; the average of responses to questions 4, 6, 8, and 11 to measure present-savoring; and the average of responses to questions 48, 49, 55, and 58 to measure past-savoring. We include all of these in various empirical specifications, with particular emphasis on interpretations of the future-savoring variable which we treat as a proxy for the parameter γ in our model.

6.3 Expectations of future outcomes

Several variables in the survey data are relevant to the level of expectations that the marginal value (compared to an alternative study path) in terms of future outcomes that the student believes is associated with his choice to study at the institution where he is currently enrolled. Another measure is the expected mark that the student expects in each course in which he is enrolled.[15] The measures we used are constructed from the following questions.

A degree from <respondent's institution> will help me to get a better job than I could get with a degree from TAFE or another vocational-training institution.

Please list the courses you are enrolled in at ¡respondent's institution¿ this semester and the final course marks (percentages out of 100) that you expect in each. (PLEASE LIST FULL NAMES OF COURSES AND SEPARATE COURSES WITH SEMICOLONS—For example: Business Statistics 80; Microeconomics 70; Financial Accounting 75.)

We also use answers to the following more involved questions.

Please choose one of the following two scenario sets:

SCENARIO SET 1 (if you prefer sports to performing arts) Imagine the next important game your favourite sports team is going to play. Which of the following would you prefer to experience in the build-up to the game?

A) You are told by all commentators and fellow supporters of all the strengths of your team and the many ways in which it is going to play better than the opponents, making you believe your team might well pull off a famous victory over a difficult opponent. During the game in question, the outcome is often close but your team does eventually lose.

B) You are told by all commentators and fellow supporters of all the weaknesses of your teams and the many ways in which it is going to be outplayed by the opponents, making you believe your team has no chance of victory. During the game in question, the outcome is often close but your team does eventually lose.

Now, suppose the same game is coming up. Which of the following would you prefer to experience in the build-up to the game?

[15] Multiple values for this variable are captured for each student on a given survey, depending upon how many courses the student was enrolled in at the time. Expected and actual marks are the only survey response variables which vary within student-semester in our data.

A) You are told by all commentators and fellow supporters of all the strengths of your team and the many ways in which it is going to play better than the opponents, making you believe your team might well pull off a famous victory over a difficult opponent. During the game in question, the outcome is often close but your team does eventually win.

B) You are told by all commentators and fellow supporters of all the weaknesses of your teams and the many ways in which it is going to be outplayed by the opponents, making you believe your team has no chance of victory. During the game in question, the outcome is often close but your team does eventually win.

SCENARIO SET 2 (if you prefer performing arts to sports) Imagine that an important international arts awards ceremony is coming up, and it is possible that your favourite performing artist may get a particular award. Which of the following would you prefer to experience in the build-up to the ceremony?

A) You are told by all art lovers and fellow fans of all the strengths of your favourite artist and the many reasons why his or her work is exceptional, making you believe your artist might well succeed in getting the award. During the awards ceremony, many commentators remark about how your artist is a contender for the award, but in the end s/he does not get it.

B) You are told by all art lovers and fellow fans of all the weaknesses of your favourite artist and the many reasons why his or her work is not particularly outstanding, making you believe your artist has no chance of getting the award. During the awards ceremony, many commentators remark about how your artist is a contender for the award, but in the end s/he does not get it.

Now, suppose the same arts awards ceremony is coming up. Which of the following would you prefer to experience in the build-up to the ceremony?

A) You are told by all art lovers and fellow fans of all the strengths of your favourite artist and the many reasons why his or her work is exceptional, making you believe your artist might well succeed in getting the award. During the awards ceremony, many commentators remark about how your artist is a contender for the award, and in the end s/he gets it.

B) You are told by all art lovers and fellow fans of all the weaknesses of your favourite artist and the many reasons why his or her work is not particularly outstanding, making you believe your artist has no chance of getting the award. During the awards ceremony, many commentators remark about how your artist is a contender for the award, and in the end s/he gets it.

6.4 Self-esteem and image of the present

We adopt a scale of self-esteem from Rosenberg (1965). Our measure is based on the following survey items, to each of which respondents chose an answer on a scale of 1 to 11 with 1 being "Strongly Agree" and 11 being "Strongly Disagree":

1. On the whole, I am satisfied with myself.
2. At times I think I am no good at all.
3. I feel that I have a number of good qualities.
4. I am able to do things as well as most people.
5. I feel I do not have much to be proud of.
6. I certainly feel useless at times.
7. I feel that I'm a person of worth, or at least on an equal plane with others.
8. I wish I could have more respect for myself.
9. All in all, I am inclined to feel that I am a failure.
10. I take a positive attitude toward myself.

After reverse-coding questions 1, 3, 4, 7, and 10, we take the simple average of responses across all ten of these questions to arrive at a 1-to-11 scaled measure of self-esteem.

We also use a measure of the favorableness of the respondent's view of the present, based on the following survey question.

Overall, I'm doing better than those I regularly compare myself to.

References

Alicke, M. & Govorun, O. (2005), The better-than-average effect, *in* M. D. Alicke, D. Dunning & J. Krueger, eds, 'The self in social judgment', Psychology Press, pp. 85–106.

Barber, B. M. & Odean, T. (2001), 'Boys will be boys: Gender, overconfidence, and common stock investment', *Quarterly Journal of Economics* **116**, 261–292.

Barros, L. & da Silveira, A. (2008), 'Overconfidence, managerial optimism, and the determinants of capital structure', *Brazilian Review of Finance* **6**(3).

Benabou, R. & Tirole, J. (2006), 'Incentives and prosocial behaviour', *American Economic Review* **96**(5), 1652–1678.

Billett, M. T. & Qian, Y. (2008), 'Are overconfident ceos born or made? evidence of self-attribution bias from frequent acquirers', *Management Science* **54**(6), 10371051.

Breiter, H., Aharon, I., Kahneman, D., Dale, A. & Shizgal, P. (2001), 'Functional imaging of neural responses to expectancy and experience of monetary gains and losses', *Neuron* **30**, 619–639.

Brennan, G. & Pettit, P. (2004), *The Economy of Esteem: An Essay on Civil And Political Society*, Oxford University Press, Oxford.

Brunnermeier, M. K. & Parker, J. A. (2005), 'Optimal expectations', *American Economic Review* **95**(4), 1092-1118.

Bryant, F. B. & Veroff, J. (2006), *Savoring: A new model of positive experience*, Lawrence Erlbaum Associates Incorporated.

Caliendo, F. & Huang, K. X. (2008), 'Overconfidence and consumption over the life cycle', *Journal of Macroeconomics* **30**(4), 13471369.

Caliendo, M., Cobb-Clark, D. & Uhlendoff, A. (2010), Locus of control and job search strategies. DIW Berlin Discussion Paper No. 979.

Camerer, C. & Lovallo, D. (1999), 'Overconfidence and excess entry: An experimental approach', *American Economic Review* **81**(1), 306–318.

Carlsson, K., Petrovic, P., Skare, S., Petersson, K. M. & Ingvar, M. (2000), 'Tickling expectations: Neural processing in anticipation of a sensory stimulus', *Journal of Cognitive Neuroscience* **12**(4), 691–703.

Carrillo, J. & Mariotti, T. (2000), 'Strategic ignorance as a self-disciplining device', *Review of Economic Studies* **67**, 529–544.

Compte, O. & Postlewaite, A. (2004), 'Confidence-enhanced performance', *American Economic Review* **94**(5), 1536–1557.

Diekhofa, E. K., Kipshagena, H. E., Falkaia, P., Dechentb, P., Baudewigb, J. & Grubera, O. (2011), 'The power of imagination—how anticipatory mental imagery alters perceptual processing of fearful facial expressions', *Neuroimage* **54**(2), 1703–1714.

Dunning, D., Heath, C. & Suls, J. (2004), 'Flawed self-assessment implications for health, education, and the workplace', *American Psychological Society* **5**(3), 69–106.

Ekstrand, D. (2008), *Christianity: The pursuit of divine truth*, Xulon Press, Florida.

Festinger, L. (1957), *A Theory of Cognitive Dissonance*, Stanford University Press, California.

Foster, G., Frijters, P. & Johnston, D. (2012), 'The triumph of hope over disappointment: A note on the utility value of good health expectations', *Journal of Economic Psychology* **33**(1), 206–214.

Frijters, P., Liu, A. Y. & Meng, X. (2008), Are optimistic expectations keeping the Chinese happy? NCER Working Paper Series 37, National Centre for Econometric Research.

Frith, C. B. & Beehler, B. M. (1997), 'Courtship and mating behaviour of the twelve-wired bird of paradise seleucidis melanoleuca', *EMU* **97**, 133–140.

Fu, G., Evans, A., Wang, L. & Lee, K. (2008), 'Lying in the name of the collective good: A developmental study', *Developmental Science* **11**(4), 495–503.

Goldman, B. (2008), *The Science of Settlement: Ideas for negotiators*, Ali-Aba, Philadelphia.

Johnson, D. (2008), *Overconfidence and War: The havoc and glory of positive illusions*, Harvard University Press, New York.

Johnson, D. D. & Fowler, J. H. (2011), 'The evolution of overconfidence', *Nature* **477**, 317320.

Jorgensena, M., Teigenb, K. & Molokkena, K. (2004), 'Better sure than safe? over-confidence in judgement based software development effort prediction intervals', *Journal of Systems and Software* **70**(1-2), 7993.

Keitlen, S. (2004), *The Psychology of Self-esteem*, Llumina Press, Florida.

Kroiss, J., Lechner, K. & Strohm, E. (2010), 'Male territoriality and mating system in the european beewolf philanthus triangulum f. (hymenoptera: Crabronidae): evidence for a hotspot lek polygyny', *Journal of Ethology* **28**, 295304.

MacDermott, S. T., Gullone, E., Allen, J. S., King, N. J. & Tonge, B. (2010), 'The emotion regulation index for children and adolescents (erica): A psychometric investigation', *Journal of Psychopathology and Behavioral Assessment* **32**(3), 301–314.

Marston, E. G., Hare, A. & Allen, J. P. (2010), 'Rejection sensitivity in late adolescence: Social and emotional sequelae', *Journal of Research on Adolescence* **20**(4), 959–982.

Morton, W. & Olenik, J. (2005), *Japan: Its history and culture*, McGraw Hill, New York.

Ohman, A. & Mineka, S. (2001), 'Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning', *Psychological Review* **108**(3), 483–523.

Robinson, W. S. (2006), 'What is it like to like?', *Philosophical Psychology* **19**(6), 743–765.

Rosenberg, M. (1965), *Society and The Adolescent Self-Image*, Princeton University Press.

Scheinkman, J. & Xiong, W. (2003), 'Overconfidence and speculative bubbles', *Journal of Political Economy* **111**, 1183–1219.

Seiter, J. (2007), 'Ingratiation and gratuity: The effect of complimenting customers on tipping behavior in restaurants', *Journal of Applied Social Psychology* **37**(3), 478–485.

Senik, C. (2008), 'Is man doomed to progress?', *Journal of Economic Behavior and Organization* **68**, 140–152.

Sherman, D. & Cohen, G. L. (2002), 'Accepting threatening information: Self-affirmation and the reduction of defensive biases', *Current Directions in Psychological Science* **11**(3), 119–123.

Steele, C., Spencer, S. & Lynch, M. (1993), 'Self-image resilience and dissonance: The role of affirmational resources', *Journal of Personality and Social Psychology* **64**, 885–896.

Susman, E., Nottelmann, E., Inoff-Germain, E., Dorn, L., Jr., B. C., Loriaux, L. & Chrousos, P. (1985), 'The relation of relative hormonal levels and physical development and social-emotional behavior in young adolescents', *Journal of Youth and Adolescence* **14**(3), 245–264.

von Hippel, B. & Trivers, R. (2011), 'The evolution and psychology of self-deception', *Behavioral and Brain Sciences* **34**, 1–56.

Walker, N. (1956), 'Freud and homeostasis', *The British Journal for the Philosophy of Science* **7**(25), 61–72.

Weinstein, N. D. (1980), 'Unrealistic optimism about future life events', *Journal of Personality and Social Psychology* **39**(5), 806–820.

Wilson, M. & Daly, M. (1985), 'Competitiveness, risk taking, and violence: The young male syndrome', *Ethology and Sociobiology* **6**, 59–73.

Wulfert, E., Roland, B. D. & Hartley, J. (2005), 'Heart rate arousal and excitement in gambling: Winners versus losers', *Psychology of Addictive Behaviors* **19**(3), 311–316.