Can Wishful Thinking Explain Evidence for Overconfidence? An Experiment on Belief Updating

Uri Gneezy[a] and Moshe Hoffman[b] and Mark A. Lane[c] and John A. List[d] and Jeffrey A. Livingston[e*] and Michael J. Seiler[f]

Uri Gneezy
University of California-San Diego and CREED

Moshe Hoffman
MIT Media Lab

Mark A. Lane
Cisco Systems, Inc.

John A. List
University of Chicago and NBER

Jeffrey A. Livingston*
Bentley University

Michael J. Seiler
College of William & Mary

Abstract

Recent theoretical work shows that the better-than-average effect, where a majority believes their ability to be better than average, can be perfectly consistent with Bayesian updating. However, later experiments that account for this theoretical advance still find behavior consistent with overconfidence. The literature notes that overoptimism can be caused by either overconfidence (optimism about performance), wishful thinking (optimism about outcomes), or both. To test whether the better-than-average effect might be explained by wishful thinking instead of overconfidence, we conduct an experiment that is similar to those used in the overconfidence literature, but removes performance as a potential channel. We find evidence that wishful thinking might explain overconfidence only among the most optimistic subjects, and that conservatism is possibly more of a worry; if unaccounted for, overconfidence might be underestimated.

JEL codes: D91, C91

* Contact author. email: jlivingston@bentley.edu. Bentley University, Department of Economics, 175 Forest Street, Waltham, MA 02452.

Acknowledgements: We thank David Franks and Lina Ramirez for many helpful comments and excellent research assistance. Any remaining errors are our own.

1. Introduction

Overconfidence has been used to explain seemingly anomalous outcomes in a wide variety of important economic contexts such as CEO investment decisions (Malmendier and Tate 2005a), pricing of consumer goods (Grubb 2009) and monetary policy decisions (Claussen et al. 2012). However, scholars from several disciplines have recognized that overconfidence is one of two potential drivers of the more general concept of overoptimism: the tendency to overestimate the probability that a preferred outcome will occur. The other is wishful thinking.[1] As Vosgerau (2010) explains, 'overconfidence describes people's overoptimism with respect to their own performance. Wishful thinking, in contrast, denotes people's overoptimism about future events that are unrelated to their performance.'

It is important to understand whether overoptimism is driven by overconfidence or wishful thinking (or both) because, while they are often observationally equivalent, they can result in different behaviors. For example, Ertac (2011), Eil and Rao (2011), Grossman and Owens (2012), and Charness, Rustichini, and van de Ven (2018) each provide evidence that people update beliefs about their own performance differently than they update beliefs about things that are unrelated to their own performance. Relatedly, the psychological motivations that have been proposed to explain overconfidence and wishful thinking are different. Explanations for overconfidence include ego utility (Koszegi, 2006) and

[1] There is some disagreement over the definition of these terms. Some studies, such as Vosgerau (2010) and Herz, Schunk, and Zehnder (2014), define the terms as we have here with 'overoptimism' referring to the general concept of overestimating the probability of a preferred outcome and 'wishful thinking' referring to the version of this that is independent of one's own performance. Others such as Heger and Papageorge (2018) swap the roles of overoptimism and wishful thinking, treating wishful thinking as the general concept and optimism as the specific version that is independent of one's own performance. We use the terms as Vosgerau (2010) and Herz, Schunk, and Zehnder (2014) define them.

strategically attempting to affect the belief of others about one's ability (Charness, Rustichini, and van de Ven 2018), while explanations for wishful thinking include desirability bias and allegiance bias (Vosgerau 2010). Misunderstanding how seemingly overoptimistic judgments or decisions are made based on overconfidence, wishful thinking, or both can lead to incorrect predictions about how people will behave in a given context and poor policy prescriptions.

Heger and Papageorge (2018) note that studies in this literature typically consider overconfidence and wishful thinking as separate phenomena. However, in situations where individuals predict their own performance and their payoffs depend on the outcome of their performance, overly optimistic predictions can be explained by either overconfidence or wishful thinking: people might be optimistic that good outcomes will occur, rather than (or in addition to) being optimistic that their performance will be superior to that of others. Ignoring wishful thinking when estimating overconfidence can thus conflate the two, leading to upward bias in estimates of overconfidence.

This is especially worrisome given that Heger and Papageorge (2018) find experimentally that people who are overconfident (in that they overestimate their own ability) also tend to engage in wishful thinking, making estimates of overconfidence that ignore wishful thinking vulnerable to omitted variables bias. Indeed, they also find that if wishful thinking is ignored, overconfidence is misdiagnosed for 29% of their observations (e.g., subjects are classified as overconfident or under-confident when they are not), and the magnitude of overconfidence is estimated incorrectly for most who are not misclassified.

With these lessons in mind, we turn to a recent literature stemming from Benoît and Dubra (2011). There are several different types of overconfidence, each of which have been widely studied in both the finance and economics literatures.[2] They focus on overplacement: overestimating one's own performance relative to others (Larrick, Burson, and Soll 2007). The 'better-than-average effect' is a related concept.[3] It occurs when a majority believes their ability to be above the relevant population's mean or median (or more generally, when actors overestimate their ability quantile). For example, Svenson (1981) highlights the empirical finding that most drivers believe they are better than the median driver. This is considered a bias because of the seemingly intuitive claim that if rational drivers have common priors, it should be arithmetically impossible for new information to result in a majority of them believing they are better than the median driver.

Benoît and Dubra (2011) show that this intuition is flawed. They demonstrate that if subjects are uncertain about their type and form beliefs based on a common prior, the fact that most drivers believe they are better than the median driver can be perfectly consistent with conventional Bayesian reasoning. If there are high ability and low ability drivers, some low ability drivers will, by chance, receive positive signals and will find via Bayes' Rule that their most likely type is high ability.

Among experiments that account for this theoretical insight, the evidence on whether overconfidence exists is mixed. The typical experiment has subjects take quizzes,

[2] There are at least three psychological biases that fall under the umbrella of overconfidence (Moore and Healy, 2008): (1) overestimation of own actual performance, (2) overplacement, i.e., overestimation of own performance relative to others, e.g., Menkhoff, Schmeling, and Schmidt (2013), and (3) excessive precision in one's beliefs (too narrow confidence intervals), e.g., Van den Steen (2011) and Proeger and Moeb (2014).

[3] Logg, Haran, and Moore (2018) note that the better than average effect and overplacement are often used interchangeably in the literature, but they explain that they are related but distinct concepts.

then estimate their performance on this quiz relative to other subjects. Clark and Friesen (2009) and Moore and Healy (2008) do not find evidence consistent with overplacement. Moore and Healy (2008), for example, have subjects take quizzes. Before taking each quiz, subjects estimate the likelihood of obtaining each possible score. They do so both for their own score and for the score of another randomly selected participant. They find that the average expected value of their own score is equal to the average expected value of the randomly selected participant's score. In contrast, Eil and Rao (2011), Merkle and Weber (2011), Burks et al. (2013), and Benoît, Dubra, and Moore (2015) each find at least some evidence consistent with overplacement.

Benoît, Dubra, and Moore (2015), hereafter BDM, carefully construct theory-based tests that are each necessary conditions for the behavior they observe to be 'rationalizable,' i.e., consistent with Bayesian rationality. If any of the tests fail, they conclude that overconfidence influences behavior. These tests are based on three theorems that are corollaries of those presented in Benoît and Dubra (2011).[4] For example, their second test 'is derived from the tautology that if agents have correct beliefs then those beliefs must be correct! For instance, in a large population, at least 3/5 of agents who assign a probability of 3/5 or greater to being above average, should actually be above average.' They find that several of these tests fail, supporting the hypothesis that subjects do display true, not apparent, overconfidence.

In this study, we examine whether wishful thinking might account for seemingly overconfident behavior in these types of experiments. We conduct an experiment with a setup designed to echo the driving ability example which motivates the Benoît and Dubra

[4] BDM conduct two different experiments. In the discussion that follows, we refer to the results of their Experiment II.

(2011) and BDM framework, but does not involve the subject's ability to perform a task well. The procedure is similar to what researchers have used to demonstrate wishful thinking. Subjects observe draws of light or dark cheerios from boxes whose types are based on the mix of cheerio colors they contain, and assess the probability that the cheerio is drawn from each potential type. Participants are randomly assigned one of three boxes that correspond to their 'ability:' low, moderate or high. Their outcomes are represented by the color of the cheerio which is drawn from these boxes: a light cheerio represents a good outcome and a dark cheerio represents a bad outcome. The higher the 'ability' of the box, the more light cheerios it contains, making a good outcome more likely. After a cheerio is drawn, the subject assesses the probabilities of holding each box, i.e., the probability that they are of each 'ability' type.

However, the description of the experiment given to the subjects makes no mention of ability, and the cheerio draw involves no effort or performance on the part of the subject. This intentionally simple design allows us to perform tests similar to those conducted by BDM, but eliminates belief in one's ability to outperform others as a potential explanation for deviation from Bayesian updating. We can thus evaluate whether other motivations such as wishful thinking can explain seemingly overconfident behavior in this type of experiment.[5] A parallel treatment tests for whether pessimism about outcomes can explain underconfidence. The only difference is drawing a light cheerio is considered a bad outcome.

[5] There is a large experimental literature that investigates the extent to which people's behavior conforms to Bayes' rule for a variety of reasons other than overoptimism or wishful thinking. These works include Kahneman and Tversky (1972), Grether (1980, 1992), Ouwersloot, Nijkamp, and Reitveld (1998), Zizzo et al. (2000), Charness and Levin (2005), Charness, Karni, and Levin (2007), Holt and Smith (2009), Coutts (2019), Enke and Zimmerman (2019), Barron (2021), and Georgalos (2021).

We also replicate part of the BDM experiment, and find similar evidence of overconfidence as they do. Our BDM-style tests also find similar violations, suggesting that wishful thinking could be behind seemingly overconfident behavior among the most optimistic subjects. For subjects who assess the probability that their cheerio was drawn from the highest 'ability' box to be high, too few actually were randomly assigned this box, failing the rationality test based on BDM's Theorem 2. BDM find that this test is failed for a larger range of subject types, as do we in our replication of their Experiment II.

We also present three different measures of the extent to which subjects update their beliefs about the probability that their cheerio was drawn from the highest 'ability' box. On average, subjects' estimates of the probability that their cheerio was drawn from a given box are closer to the common prior belief than Bayesian updating would suggest. These results go in the opposite direction of overconfidence, and can possibly be explained by conservatism (El-Gamal and Grether 1995, Edwards 1982), where subjects overweight prior beliefs in the face of new information.

The remainder of the paper proceeds as follows. Section 2 motivates our experimental design, and Section 3 describes this design. Section 4 presents the results. Section 5 concludes.

2. Motivation

Our experiment is motivated by and designed around the following example, which is similar to what is offered in Benoît and Dubra (2011) and BDM. There are three types of drivers: one-third are low ability, one-third are medium ability, and one-third are high ability. Low ability drivers have a 40% chance of avoiding an accident, medium ability

drivers have a 90% chance of avoiding an accident, and high ability drivers have a 100% chance of avoiding an accident.

Driving ability is represented by a box from which either a light or dark cheerio is drawn. Light cheerios represent a good outcome, e.g., avoiding an accident, while dark cheerios represent a bad outcome, e.g., having an accident. The boxes contain different proportions of dark cheerios to match the parameters of the example above. Box 1, the 'low' ability box, contains six dark cheerios and four light cheerios. Box 2, the 'medium' ability box, contains one dark cheerio and nine light cheerios. Box 3, the 'high' ability box, contains 10 light cheerios. The subjects observe a cheerio draw then evaluate the probability that the cheerio came from each box.

While the situation corresponds nicely to a canonical driving ability example often studied in the literature, the cheerio draw involves no effort or ability on the part of the subject, and no context regarding ability is given, so the event has nothing to do with the subject's ego or self-image; indeed, the subjects perform a simple updating task. Thus, any violations of Bayesian rationality must be due to a bias other than overconfidence.

Since the only relevant signal about the subject's 'ability' is the color of the cheerio that is drawn, we can directly measure how subjects update their beliefs in response to this signal and evaluate whether they do so in a manner consistent with Bayes' rule. If subjects are rational Bayesians, then, rounded to the nearest thousandth, they will calculate that

p(Box 1 | light cheerio) = 0.174

p(Box 2 | light cheerio) = 0.391

p(Box 3 | light cheerio) = 0.435.

Based on these probabilities, we perform the three tests employed by BDM in order to directly evaluate how their results might have been impacted by biases other than overconfidence.[6] BDM conduct two experiments where subjects first take logic and math quizzes. In Experiment II, following the quizzes, the subjects are asked how likely it is that they rank in the top half of quiz takers. Based on these choices, they construct several tests based on three theorems which must be passed in order for behavior to be consistent with rationality. Their Theorem 1 is as follows:

Suppose that a fraction x of the population believes that there is a probability at least q that their types are in the top $y < q$ of the population. These beliefs can be rationalized if and only if $xq \leq y$.

As BDM note, 'we can infer overconfidence if a sufficiently large fraction of people (variable x in the theorem) believe sufficiently strongly (variable q) that they rank sufficiently high (variable y).' They find that all tests for rationality based on this theorem are passed. For example, 55% of subjects report that they are at least 70% likely to be in the top half of quiz takers. While this behavior exhibits apparent overconfidence as

[6] Burks et al. (2013) rule out Bayesian rationality by asking subjects to predict in which quintile their score will fall. They show that if subjects are acting as Bayesians it must be true that of all individuals placing themselves in quintile k, the largest (i.e., modal) share of them must actually be from quintile k. They find several violations of this condition. Since we did not ask a similar question we cannot duplicate this test. The closest we can come is to use responses to measure 1, whether the subjects think Box 1 or Box 3 is most likely. By far the modal share who think Box 1 (3) is most likely actually were assigned to Box 1 (3) in the nationwide sample. A similar test for the Chicago sample is not possible since we do not have data on the boxes to which those subjects were assigned.

described by Benoît and Dubra (2011), it does not violate Theorem 1 since $qx = 0.7 \times 0.55 = 0.385$, which is less than $y = 0.5$ as required.

Their Theorem 2 is as follows:

Suppose that a fraction x of the population believes there is a probability of at least q that their types are in the top $y < q$ of the population. Let $\tilde{x}$ be the fraction of people who have those beliefs and whose actual type is in the top y of the population. This data can be rationalized if and only if $xq \leq \tilde{x}$.

Theorem 2 requires that subject beliefs about their ability and their actual ability must be consistent. They perform a number of tests for different values of x and q and find that several of these tests fail. For example, they find that 35% of subjects indicate they have a probability of at least 0.8 of being in the top half of quiz takers. By Theorem 2, at least $35 \times 0.8 = 28\%$ of the subjects should express this belief and also be in the top half, but only 20% meet both of these two conditions, so the subjects' beliefs are inconsistent with Bayesian rationality.

Finally, their Theorem 3 is as follows:

In a population of n individuals, let r_i, $i = 1, \ldots, n$, be the probability with which individual i believes his type is in the top y of the population. This data can be rationalized if and only if $\frac{1}{n}\sum_{i=1}^{n} r_i = y$.

Theorem 3 requires that the average of the likelihoods of ending in the top y be equal to y. They find that this test fails. The average probability given that the subject is in the top 50% is 0.67, which they find is significantly different from 0.5 at the 1% level. Our experimental design, which we describe in detail in the next section, allows similar tests.

We also conduct tests that examine whether the subjects' updated beliefs are consistent with the predicted probabilities derived from Bayesian updating. Such tests are not possible in overconfidence studies that examine beliefs about a particular skill. As Burks et al. (2013) note, subjects gather many private signals about their quiz-taking ability

during their life. Since these numerous signals are unobserved to the econometrician, each subject's Bayesian posterior cannot be calculated, so it is impossible to directly evaluate whether their updated beliefs are consistent with the Bayesian prediction.

3. Experimental Design

We recruited participants from two sources. The first sample is recruited from the student body at the University of Chicago through advertisements distributed to an experimental list host. This sample participated in the experiment face-to-face with an experimenter. To provide external validity, the second sample was recruited online from a nationwide database of homeowners via Mechanical Turk (https://www.mturk.com). In this sample, the average age of participants is 31.53, 82.90% are Caucasian, 45.98% have at least a college degree and 98.85% have at least a high school diploma, 79.48% earn $60,000 or less per year, and 31.03% are married. We had a total of 395 participants in the University of Chicago sample, 10 of whom were excluded from the analysis,[7] and 348 participants in the nationwide sample. Average earnings were $7.44 for the University of Chicago sample and $6.85 for the nationwide sample, and the experiment took roughly 15 minutes to complete in both the face-to-face and online versions.

The face-to-face version of the experiment proceeded as follows. When participants arrived at the lab, the experimenter showed them a tin with 3 boxes inside. The contents of each box were then revealed to participants. Box 1 (which corresponds to low ability in the driving example described above) had 6 dark cheerios and 4 light cheerios;

[7] One participant did not answer the rationality check question correctly. Five participants gave probabilities that did not add up to 1. One participant said Box 1 was equally likely as Box 3. Three participants indicated in an exit survey that they were confused by one or more of the questions. Results remain effectively the same without such removals.

Box 2 (which corresponds to medium ability) had 1 dark cheerio and 9 light cheerios; and Box 3 (which corresponds to high ability) had 10 light cheerios and 0 dark cheerios.[8]

After proper inspection, the experimenter privately, and randomly, drew a box from the tin. Thus, as is required by the Benoît and Dubra (2011) setting, subjects' types are uncertain, and they should have common priors since they are all aware that each box is equally likely to be selected. The experimenter then randomly drew a cheerio from that box, wrote the color of the cheerio on top of the instruction sheet, returned the cheerio to the chosen box, and set the box aside, out of sight of the participant.

The experimenter then instructed the participant that there are four parts to the experiment (see Appendix A for verbatim instructions). The participant was advised to read and answer carefully as their payment will depend on their answers. In each part, the participant had the opportunity to earn $3, and therefore a total of $12. If nothing was earned in these four parts, the participant was given $1.

We execute two treatments. Our main condition is the overconfidence treatment, in which 232 of the 395 subjects in the Chicago sample and 175 of the 348 subjects in the nationwide sample were randomly selected to participate. In this treatment, a draw of a light cheerio represents a good outcome. Since most participants will randomly draw a light cheerio (expected value of 23 out of 30, or 77%), most should rationally conclude, for example, that the chance that they randomly were given Box 3 is higher than Box 1. Hence, in line with Bayesian updating, a majority of the people rationally should believe that their box is better than the median box.

[8] Again, to be clear, no context regarding ability was provided.

The second is an underconfidence treatment, which is a mirror of the overconfidence treatment. Everything in this experimental protocol is identical except drawing a light cheerio is now incented to be a bad outcome and drawing a dark cheerio is a good outcome. In this case, only a minority of individuals will receive a positive signal, so more than half the participants rationally should believe that their box is inferior to the median box.

Part 1 in the experiment represents a rationality check, as we inquired whether the participant understood the earnings process. Subjects in the overconfidence (underconfidence) treatment are told that if the experimenter drew a light (dark) cheerio, then they will earn $3 for this part of the study. They are then asked how much money they earned from this part of the study. All participants but one answered this question correctly, suggesting that they understood the rules and the outcome of the random draw.

In Part 2, we asked the participant whether she believes the box the experimenter drew from the tin was more likely to have been Box 1 or Box 3. Subjects earn $3 if they select the box that is actually more likely to have been drawn from the tin, conditional on the color of the cheerio that was drawn, as calculated using Bayes' rule. We made the choice dichotomous to focus on the relative likelihood of the two events.

Part 3 asked participants to state the exact probabilities of each box given the cheerio draw—earning $3 if the participants' stated probability is within 5% of the true probability, for each of the three boxes.

Finally, in Part 4, the participant was informed that the experimenter was to choose another cheerio and asked participants if they would like the draw to be from the same box or from one of the other two boxes. Of course, a draw of a light cheerio before Step 1

should induce participants to stick with their box in the overconfidence treatment, while a draw of a dark cheerio before Step 1 should induce participants to stick with their box in the underconfidence treatment. Subjects earn $3 if a light (dark) cheerio is then drawn in the overconfidence (underconfidence) treatment. After each part was completed, the participant was then paid their earnings from each part privately as promised in the instructions.

The online version of the experiment recreated the face-to-face version as closely as possible. Upon arrival on the experiment website, participants first listened to a message that reports potential earnings and ended by asking them to type in a number that is read at the end of the message. The experiment then proceeded in exactly the same fashion as the face-to-face version, but the presentation of boxes and cheerio drawings were shown in animated videos.[9] The order of the steps, earnings structure, and instructions are identical to their face-to-face counterparts.

In addition, we later replicated both BDM's Experiment II and the overconfidence treatment of our Cheerio experiment at the University of Chicago in a within-subject design with 80 subjects. In the Experiment II replication, we followed BDM's procedures as closely as possible; verbatim instructions are presented in Appendix B. This experiment has subjects take a 20 question test. Subjects first take a five question sample quiz to see what the questions will be like. They are then told that the median score achieved by previous test takers was 18, and state how likely they think it is that they will rank in the top half of test takers. Finally, they take the test.

4. Results

[9] The experiment website, design and videos can be viewed at http://www.ibere.org/surveys/ch001/.

4.1 Overconfidence: BDM-Style Tests

We begin with BDM's tests of consistency with Bayesian rationality based on their Theorems 1 through 3. We conduct these tests using the data from our replication of BDM's Experiment II, as well as the data from the overconfidence treatment of our Cheerio experiment. Table 1 presents the results.

Panel A presents the results from our BDM replication, while Panel B presents the results from our companion Cheerio experiment.[10] First, consider the tests based on Theorem 1. These tests find behavior consistent with overconfidence if a sufficiently large fraction of subjects believe sufficiently strongly that their score is in the top half for the BDM quiz experiment, or that they were assigned Box 3, the 'high ability' box that contains 10 light cheerios, in our ability-free companion experiment.

In the original BDM Experiment II, their tests cannot rule out the possibility that the Bayesian model describes subject behavior for all levels of q that are able to examine. As they explain, 'For instance, 35% of subjects indicate they have a probability of at least 0.8, or 0.79 when we allow for rounding, of ending in the top half. From Theorem 1, up to 62% of subjects could rationally make such an indication, so the data pass this test.' We find similar results in our replication. As shown in Panel A, the Theorem 1-based tests are passed for all levels of q we are able to consider.[11]

[10] The results of the tests based on Theorem 1 and Theorem 3 that we present in Panel B of Table 1 are those performed using data from the Chicago sample. Tests performed using data from the nationwide sample yield the same conclusions. The tests based on Theorem 2 require knowledge about from which box a subject's cheerio was actually drawn. We have these data only for the nationwide sample, so those results are based on those data.

[11] The levels of q that are possible to consider are limited by the predicted probabilities that are chosen by the subjects.

Similarly, as shown in Panel B, our ability-free experiment reveals no violations of Bayesian rationality based on Theorem 1. For all levels of q that we examine, the tests cannot rule out the possibility that the Bayesian model describes subject behavior. For example, 48.3% of subjects believe the probability that they were assigned Box 3 is at least 0.4. By Theorem 1, the product of these (0.483×0.4 equals 0.193) should be no more than 0.333, so this is consistent with Bayesian updating.

Second, we consider tests based on Theorem 2, which require that enough subjects actually are of high ability to justify their beliefs. Recalling the variable definitions in Section 2, BDM show that the probability that the data cannot be explained by Bayesian rationality is $P(w \leq n\tilde{x})$, where w is a random variable with binomial distribution B(nx, q) and n is the total number of subjects. They illustrate how these tests works by citing the same example as above, where 35% of subjects indicate they have a probability of at least 0.79 (allowing for rounding) of ending in the top half: 'Of

Table 1. Tests based on Benoît, Dubra and Moore (2015), overconfidence treatment

Panel A: Benoît, Dubra and Moore (2015) Experiment II Replication (Quiz experiment)

Based on:			Theorem 1		Theorem 2		Theorem 3
				Pass if			
y	q	x	qx	$qx < y$	$\tilde{x}$	$P(w \leq n\tilde{x})$	Avg. stated P(in top half)
0.5	0.40	0.88	0.35	Pass	0.50	0.005***	0.56***
0.5	0.42	0.75	0.32	Pass	0.46	0.002***	(0.17)
0.5	0.44	0.74	0.32	Pass	0.45	0.012**	
0.5	0.48	0.73	0.35	Pass	0.44	0.066*	
0.5	0.50	0.71	0.36	Pass	0.44	0.111	
0.5	0.52	0.54	0.28	Pass	0.34	0.172	
0.5	0.54	0.53	0.28	Pass	0.35	0.216	
0.5	0.56	0.49	0.27	Pass	0.30	0.523	
0.5	0.60	0.46	0.28	Pass	0.29	0.868	
0.5	0.64	0.34	0.22	Pass	0.21	1.000	
0.5	0.65	0.31	0.20	Pass	0.20	1.000	
0.5	0.66	0.30	0.20	Pass	0.19	0.830	
0.5	0.68	0.26	0.18	Pass	0.18	1.000	
0.5	0.70	0.23	0.16	Pass	0.14	0.443	
0.5	0.76	0.13	0.10	Pass	0.08	0.266	
0.5	0.80	0.11	0.09	Pass	0.06	0.086*	
0.5	0.86	0.04	0.04	Pass	0.03	0.097*	
0.5	0.90	0.04	0.03	Pass	0.01	0.028**	
0.5	0.99	0.03	0.03	Pass	0.01	0.020**	

Panel B: Cheerio experiment

Based on:			Theorem 1		Theorem 2		Theorem 3		
				Pass if			Avg. stated P(box i), i = 1,2,3		
y	q	x	qx	$qx < y$	$\tilde{x}$	$P(w \leq n\tilde{x})$	Box 1	Box 2	Box 3
0.333	0.10	0.80	0.08	Pass	0.309	1.00	Chicago:		
0.333	0.20	0.80	0.16	Pass	0.303	1.00	0.33	0.32	0.33
0.333	0.30	0.79	0.24	Pass	0.291	0.99	(0.24)	(0.11)	(0.19)
0.333	0.33	0.79	0.26	Pass	0.286	0.93			
0.333	0.34	0.55	0.19	Pass	0.274	0.87	Nationwide:		
0.333	0.35	0.53	0.18	Pass	0.200	0.58	0.35	0.31**	0.34
0.333	0.40	0.48	0.19	Pass	0.188	0.27	(0.24)	(0.13)	(0.21)
0.333	0.44	0.27	0.12	Pass	0.125	0.06*			
0.333	0.45	0.18	0.08	Pass	0.097	0.09*			
0.333	0.50	0.14	0.07	Pass	0.063	0.003***			
0.333	0.60	0.04	0.02	Pass	0.011	0.003***			

Notes: *, **, and *** indicate statistically significantly different from 0.5 (Panel A) or 1/3 (Panel B) at the 10%, 5% and 1% levels, respectively.

Panel A: x is the percentage with belief greater than q that they will place in the top half; qx

is the fraction of x who actually place in the top half. $\tilde{x}$ is the percentage of subjects who estimate the probability that their quiz score is in the top half to be at least q and actually did score in the top half. $P(w \leq n\tilde{x})$ is the probability the data come from a rational, Bayesian model where w is a binomial (nx, q) and n is the total number of subjects.
Panel B: x is the percentage with belief greater than q that their cheerio draw came from Box 3; qx is the fraction of x whose draw actually came from Box 3. $\tilde{x}$ is the percentage of subjects who estimate the probability that their draw came from Box 3 to be at least q and actually did have their draw come from Box 3. $P(w \leq n\tilde{x})$ is the probability the data come from a rational, Bayesian model where w is a binomial (nx, q) and n is the total number of subjects.
Theorem 3: Standard deviations in parentheses. For the BDM Experiment II replications in Panel A, each test examines the null hypothesis that the average stated probability that the subject's quiz score is in the top half is equal to 0.5. For the Cheerio experiment in Panel B, each test examines the null hypothesis that the stated probability that the cheerio was drawn from each box is equal to 1/3.

these subjects, 58% are actually in the top half. These data do not pass the test, as there is less than a 1% chance that a sample as apparently overconfident as this, or more, will arise from a rational population—$P(w \leq 74\times0.58) < 0.01$, for $w \sim B(74\times0.35, 0.79)$.' They reject the null hypothesis for several values of q in the middle of the range of those they consider, but cannot reject it for the lowest and highest values of q.

In contrast, in our replication of this experiment with University of Chicago students, we find that Bayesian rationality is rejected on the basis of these tests for the lowest and highest values of q rather than those in the middle. For example, as shown in Panel A of Table 1, 88% of subjects think the probability that their quiz score will be in the top half is at least 0.4, but only 57% of these subjects actually score in the top half. The hypothesis that a sample as apparently overconfident as this, or more, will arise from a rational population is rejected at the 1% level. Similar hypotheses are rejected for values of q ranging from 0.42 to 0.48, and from 0.8 to 0.99 at various levels of significance.

In the context of our ability-free Cheerio experiment, given how many subjects believe their draw came from Box 3 with a certain probability q, there must be enough subjects whose type is actually Box 3 to justify those beliefs. The results, presented in Panel B of Table 1, show that for the most optimistic subjects who assess the probability that their cheerio was drawn from Box 3 to be higher than the Bayesian posterior following a draw of a light cheerio (0.435), not enough subjects actually had their cheerio drawn from Box 3. The Bayesian model is rejected at the 10% level of significance when q is set to 0.44 or 0.45 and at the 1% level when q is set to 0.5 or 0.6. Like the most optimistic subjects in our replication of BDM's Experiment II, these subjects display behavior consistent with overconfidence; were the event in question quiz-taking ability, the literature would interpret this as evidence that subjects are overconfident. However, overconfidence cannot be what motivates their choices since the subjects are merely updating their beliefs about an event that has nothing to do with their ego or ability. Accordingly, wishful thinking may explain some of the overconfidence identified by BDM among the most optimistic subjects in their tests based on Theorem 2.

Tests based on Theorem 3 reveal a different story. As noted in section 2, Theorem 3 requires that the average of the likelihoods of ending in the top y be equal to y. BDM find that this test fails since this average likelihood is 67%, well above 50%. As shown in Panel A of Table 1, we find similar (but less severe) overconfidence in our replication: the average likelihood is 56%, which is significantly different from 50 at the 1% level. However, as presented in Panel B of Table 1, similar tests from our experiment that removes ability largely cannot rule out Bayesian rationality. Only the average probability for Box 2 in the nationwide sample is significantly different from 1/3, but that average and

all of the others are all very close to it. There is no instance like BDM find where the average probability of an 'ability level' is far above what it is supposed to be. Wishful thinking cannot explain the overconfidence identified by BDM via the test based on Theorem 3.[12]

4.2 Overconfidence: Additional Measures

Next, we examine whether our direct measures of how subjects update beliefs are consistent with Bayesian updating in the overconfidence treatment. We have three measures of how subject beliefs evolve in response to the cheerio draw: 1) whether they chose Box 3 in Part 2, 2) whether they gave a higher probability to Box 3 than to Box 1 in Part 3,[13] and 3) whether they chose to keep their box in Part 4. Recall that all three measures are incentivized. While the first and second measures might be hampered by a language problem (participants may not understand 'probability' and 'likely' the way we intended), the third measure is not subject to this concern.

Table 2 presents summary statistics for each of our measures. While the subjects exhibit apparent overconfidence since the majority of our participants believed they were

[12] We conducted the replications of BDM's Experiment II and the overconfidence treatment of our ability-free Cheerio experiment in response to a referee who noted that the differences between our findings and those of BDM might be due not to differences between overconfidence and wishful thinking, but rather to differences in some other aspect of how we implemented our designs. Our original experiment at the University of Chicago included 315 subjects, 152 of whom were randomized into the overconfidence treatment. Combined with the 80 subjects who participated in our replication, we have 232 subjects from the University of Chicago who participated in the overconfidence treatment. In our analysis presented in Tables 1 and 2, we use this combined sample. The results from the replication are quantitatively and qualitatively similar to the results from our original experiment. Results from each separate sample are available upon request. Since we are able to replicate the qualitative results of BDM's Experiment II and our own overconfidence condition with the same set of subjects, this assuages concerns that differences between the two are driven by implementation differences instead of differences between overconfidence and wishful thinking.

[13] Subjects estimate the posterior probability of all three boxes. We first compare their reported posteriors for Box 3 and Box 1 to maintain comparability with the other two measures, but will later look at the reported posterior probabilities for all three boxes.

above the median, they *underestimate* the probability that the cheerio was drawn from the better box, Box 3, according to all three measures. It also presents tests to see if these differences from the Bayesian predictions are statistically significant.

Consider first measure 1 and measure 3, which are presented in panels A and C, respectively. Each measures the degree to which the subjects think the cheerio was more likely to have come from Box 3 (the 'high ability' box) than Box 1 (the 'low ability' box). In each case, we perform a test of proportions where the null hypothesis is that the proportion who think Box 3 is more likely is equal to the Bayesian prediction (0.797 for the Chicago sample and 0.749 for the nationwide sample).

For measure 1, fewer subjects think their draw was more likely to have come from Box 3 than would be predicted if all subjects updated beliefs using Bayes' Rule. For example, in the Chicago sample, only 74.1% of the subjects think that their box is more likely to have been Box 3 than Box 1, which is less than the 79.7% who would have if subjects were perfect Bayesians. However, only the measure from the Chicago sample is statistically significantly different from the Bayesian prediction.

For measure 3, however, the proportion of subjects who choose to take their next draw from a different box is well below what Bayesian rationality would predict, and the differences are statistically significant at the 1% level in both samples. For example, only 66.8% of the subjects choose to have their next draw come from the same box in the Chicago sample, far below the 79.7% that Bayesian rationality predicts.

Table 2. Summary statistics and results, overconfidence treatment

A. Measure 1: Which box do you think is more likely?			
Percent who think Box 3 (the better box) is more likely			
	Chicago sample	Nationwide sample	
Observed	0.741**	0.709	
	(0.439)	(0.456)	
	[0.029]	[0.034]	
Bayesian prediction	*0.797*	*0.749*	
B. Measure 2: What do you think the probability is of each box?			
Following a light draw:			
	Box 1	Box 2	Box 3
Chicago sample	0.228***	0.345***	0.415**
	(0.092)	(0.081)	(0.107)
	[0.007]	[0.006]	[0.008]
Nationwide sample	0.241***	0.324***	0.436
	(0.120)	(0.101)	(0.117)
	[0.010]	[0.009]	[0.010]
Bayesian posterior	*0.174*	*0.391*	*0.435*
Following a dark draw:			
	Box 1	Box 2	Box 3
Chicago sample	0.744***	0.238***	0.019*
	(0.187)	(0.162)	(0.073)
	[0.027]	[0.024]	[0.011]
Nationwide sample	0.680***	0.286***	0.035
	(0.209)	(0.179)	(0.137)
	[0.032]	[0.027]	[0.021]
Bayesian posterior	*0.857*	*0.143*	*0*
C. Measure 3: For the next draw, keep current box or draw from a new one?			
Percent who choose to keep box:			
	Chicago sample	Nationwide sample	
Observed	0.668***	0.634***	
	(0.472)	(0.483)	
	[0.031]	[0.036]	
Bayesian prediction	*0.797*	*0.749*	

Notes: Standard deviations in parentheses; standard errors in brackets.
*, ** and *** indicate statistically significantly different from the Bayesian prediction at the 10%, 5% and 1% levels, respectively.

Measure 2 asks subjects to report their predicted probabilities for each of the three boxes. The results are presented in Panel B of Table 2. First, consider participants who randomly drew a light cheerio. When a light cheerio is drawn, the Bayesian posteriors are

as reported above in Section 2. Relative to these probabilities, our subjects overestimate the probability that the draw came from Box 1, underestimate the probability that the draw came from Box 2, and correctly estimate the probability that the draw came from Box 3. The Box 1 and Box 2 estimates are both closer to the common prior of 1/3 for each box than Bayes' Rule would predict. For example, the subjects stated an average posterior probability for Box 1 of 0.228 in the Chicago sample and 0.241 in the nationwide sample. Each is statistically significantly higher than the true posterior probability 0.174 at the 1% level using a two-tailed *t*-test. The average probability stated for the 'high ability' box, Box 3, is also underestimated in the Chicago sample: 0.415, which is below the actual posterior probability of 0.435; the difference is significant at the 5% level. However, in the nationwide sample, the average stated Box 3 probability is 0.436, almost identical to the actual posterior probability.

Similarly, for those who drew a dark cheerio, stated posterior beliefs for the low and medium boxes are closer to the prior than Bayes' Rule would predict. The average posterior assigned to Box 1 is 0.744 in the Chicago sample and 0.680 in the nationwide sample. Each is statistically significantly different from the Bayesian posterior of 0.857 at the 1% level.

Finally, the average posterior given for Box 3 following a dark cheerio draw is 0.019 in the Chicago sample and 0.035 in the nationwide sample, slightly above the true posterior of 0. This result is driven by a small minority of subjects; the large majority behaved in accordance with Bayes' rule. Only four participants overestimated this probability in each sample whereas 43(40) participants gave the correct probability of 0 in the Chicago (nationwide) sample. Moreover, this result is influenced by the fact that it is

impossible to understate the probability of being in Box 3 since the true posterior probability of having drawn from Box 3 is 0.

These results are consistent with conservatism, under which people need more evidence to change their priors than implied by Bayes rule (El-Gamal and Grether, 1995). A closer look at the data reveals they may be driven by subjects who do not update their beliefs at all following the cheerio draw. 23% of the subjects in the Chicago sample and 17% of the subjects in the nationwide sample report updated beliefs of 0.33 to 0.34 for all three boxes. It is possible that these subjects are truly that conservative in their reactions to the cheerio draw, but it is also possible that they did not fully understand the task.

4.3 Underconfidence

We now turn to the results from the underconfidence condition, where the only difference is that a dark cheerio is now the good outcome and a light cheerio is now the bad outcome. Table 3 presents the results of our BDM-style tests using the data from the nationwide sample. Unlike the overconfidence condition, there is no evidence that any of our subjects exhibit underconfidence according to any of the three families of tests. For Theorems 1 and 2, there are no statistically significant differences from what we would expect if subjects conducted Bayesian reasoning. For Theorem 3, the average estimated probability of holding Box 3 by subjects in the Chicago sample is slightly lower than 0.333; the difference is significant at the 10% level.

Table 4 presents each of our three measures of underconfidence. For each measure, subjects under-adjust their beliefs relative to the Bayesian prediction. For measure 1, fewer subjects think it is more likely that their cheerio was drawn from the

Table 3. Tests based on Benoît, Dubra and Moore (2015), underconfidence treatment

Panel A. Tests based on Theorem 1 and Theorem 2

Based on:		Theorem 1:			Theorem 2:	
y	q	x	qx	Pass if $qx < y$	$\tilde{x}$ (5)	$P(w \leq n\tilde{x})$ (6)
0.333	0.05	0.740	0.037	Pass	0.324	1.00
0.333	0.20	0.734	0.147	Pass	0.324	1.00
0.333	0.30	0.717	0.215	Pass	0.318	1.00
0.333	0.33	0.705	0.233	Pass	0.318	1.00
0.333	0.34	0.682	0.232	Pass	0.318	1.00
0.333	0.35	0.486	0.170	Pass	0.237	1.00
0.333	0.40	0.451	0.180	Pass	0.225	0.97
0.333	0.43	0.364	0.157	Pass	0.185	0.92
0.333	0.44	0.329	0.145	Pass	0.162	0.82
0.333	0.45	0.277	0.125	Pass	0.139	0.80
0.333	0.50	0.214	0.107	Pass	0.098	0.37
0.333	0.60	0.052	0.031	Pass	0.023	0.27

Panel B. Test based on Theorem 3: Average probabilities for each box

Chicago sample			Nationwide sample		
p(Box 1)	p(Box 2)	p(Box 3)	p(Box 1)	p(Box 2)	p(Box 3)
0.367	0.329	0.304*	0.361	0.325	0.315
(0.259)	(0.123)	(0.209)	(0.228)	(0.125)	(0.211)

Notes: *, **, and *** indicate statistically significantly different from 1/3 at the 10%, 5% and 1% levels, respectively.

Panel A: x is the percentage with belief greater than q that their cheerio draw came from Box 3; qx is the fraction of x whose draw actually came from Box 3. $\tilde{x}$ is the percentage of subjects who estimate the probability that their draw came from Box 3 to be at least q and actually did have their draw come from Box 3. $P(w \leq n\tilde{x})$ is the probability the data come from a rational, Bayesian model where w is a binomial (nx, q) and n is the total number of subjects.

Panel B: Standard deviations in parentheses. Each test examines the null hypothesis that the probability is equal to 1/3.

worse box, Box 3, than from the best box, Box 1. In the nationwide sample, for example, 64.7% think that Box 3 is more likely; 74.0% would think so if subjects updated their beliefs using Bayes rule. For measure 3 from both samples, more subjects elect to keep their box than Bayes rule would predict.

For measure 2, in both samples following a draw of a 'bad' light cheerio, relative to the Bayesian prediction, subjects overestimate the probability that their cheerio came from the 'good' box, Box 1, and underestimate the probability that their cheerio came from the 'bad' box, Box 3. These averages go in the opposite direction of overconfidence. The averages are what one would expect of underconfident subjects following a draw of a 'good' dark cheerio, however. Subjects underestimate the probability that their cheerio was drawn from the good box, and overestimate the probability that their cheerio came from the bad box.[14] All of these adjustments are closer to the prior than Bayes rule predicts.

In sum, there is little evidence that pessimism about outcomes can explain seemingly underconfident behavior in experiments that involve the performance of subjects. When subject performance is removed as a channel, no patterns consistent with underconfidence remain. The results of our BDM-style tests cannot reject the null hypothesis that the observed behavior is consistent with Bayesian reasoning, and our three measures of belief updating are better explained by a tendency to under-adjust beliefs after new information is acquired. However, as in the overconfidence treatment, these results are partially driven by subjects who do not update their beliefs at all after the Cheerio draw: 17% in the Chicago sample and 21% in the nationwide sample.

[14] However, much like in the overconfidence treatment, the average reported 'bad box' probability is driven by a small minority of participants. Five (four) participants overestimated this probability, while 38(41) gave exactly the correct value of 0 in the Chicago (nationwide) sample. Also, as before, this could be due to the fact that it is impossible to understate the probability of being in Box 3.

Table 4. Summary statistics and results, underconfidence treatment

A. Measure 1: Which box do you think is more likely?			
Percent who think Box 3 (the worse box) is more likely			
	Chicago sample	Nationwide sample	
Observed	0.680	0.647***	
	(0.468)	(0.479)	
	[0.038]	[0.036]	
Bayesian prediction	*0.719*	*0.740*	
B. Measure 2: What do you think the probability is of each box?			
Following a light draw:			
	Box 1	Box 2	Box 3
Chicago sample	0.231***	0.358***	0.410**
	(0.111)	(0.083)	(0.124)
	[0.011]	[0.008]	[0.012]
Nationwide sample	0.259***	0.324***	0.417
	(0.128)	(0.100)	(0.131)
	[0.011]	[0.009]	[0.012]
Bayesian posterior	*0.174*	*0.391*	*0.435*
Following a dark draw:			
	Box 1	Box 2	Box 3
Chicago sample	0.714***	0.256***	0.031*
	(0.199)	(0.169)	(0.109)
	[0.030]	[0.026]	[0.017]
Nationwide sample	0.650***	0.327***	0.024*
	(0.202)	(0.181)	(0.085)
	[0.030]	[0.027]	[0.013]
Bayesian posterior	*0.857*	*0.143*	*0*
C. Measure 3: For the next draw, keep current box or draw from a new one?			
Percent who choose to keep box:			
	Chicago sample	Nationwide sample	
Observed	0.392***	0.479***	
	(0.490)	(0.501)	
	[0.039]	[0.038]	
Bayesian prediction	*0.281*	*0.291*	

Notes: Standard deviations in parentheses; standard errors in brackets.
*, ** and *** indicates statistically significantly different from the Bayesian prediction at the 10%, 5% and 1% levels, respectively.

5. Conclusion

The recent literature on the better-than-average effect presents the results of experiments where overconfidence could potentially influence behavior and assume that

choices not in accordance with Bayesian rationality can be attributed to overconfidence. However, the literature notes that overoptimism can also be caused by wishful thinking: optimism about outcomes rather than optimism about ability or performance. Failing to account for such biases could lead to distorted estimates of overconfidence.

We examine whether wishful thinking could also explain seemingly overconfident behavior by conducting an experiment which is as close as possible to those used in the overconfidence literature but has the performance channel removed, leaving only other biases as potential reasons why subjects might arrive at beliefs inconsistent with Bayesian rationality. The setup is designed to map carefully to the canonical driving ability example studied in the literature, but the event that subjects evaluate – a draw of either a light or dark cheerio from one of three randomly selected boxes – involves no effort or ability, and the context does not mention ability in any way. By removing the psychological drivers of overconfidence such as the subject's ego or self-image, we can conclude that any beliefs that are inconsistent with Bayes' Rule are arrived at due to some bias other than overconfidence. Should such biases play a role in this experiment, we should worry that they could be confounds in similar experiments where overconfidence can be a potential driver of behavior.

We find some evidence that wishful thinking might impact the findings of studies such as BDM or Burks et al. (2013), but only among the most optimistic subjects. Based on the tests derived from BDM's Theorem 2, we find that too many subjects believe the probability that they have the highest 'ability' box is 0.44 or higher. BDM find that subjects are overly optimistic for a larger range of probabilities, as do we in a replication of their Experiment II. We also find no evidence that beliefs over outcomes can explain excessive

pessimism about one's own ability. Conservatism is perhaps more of a worry; if unaccounted for, overconfidence and underconfidence might be underestimated.

References

Barron, Kai (2021). Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*, 24(1), 31-58.

Benoît, Jean-Pierre, and Juan Dubra (2011). Apparent Overconfidence. *Econometrica*, 79(5), 1591-1625.

Benoît, Jean-Pierre, Juan Dubra, and Don A. Moore (2015). Does the Better-than-Average Effect Show that People are Overconfident? Two Experiments. *Journal of the European Economic Association*, 13(2), 293-329.

Burks, Steven V., Jeffrey P. Carpenter, Lorenz Gotte, and Aldo Rustichini (2013). Overconfidence and Social Signalling. *Review of Economic Studies*, 80(3), 949-983.

Charness, Gary, Edi Karni, and Dan Levin (2007). Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty*, 35(2), 129-148.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven (2018). Self-Confidence and Strategic Behavior. *Experimental Economics*, 21(1), 72-98.

Clark, Jeremy, and Lana Friesen (2009). Overconfidence in Forecasts of Own Performance: An Experimental Study. *The Economic Journal*, 119(534), 229-251.

Claussen, Carl A., Egil Matsen, Øistein Roisland, and Ragnar Torvik (2012). Overconfidence, Monetary Policy Committees and Chairman Dominance. *Journal of Economic Behavior & Organization*, 81(2), 699-711.

Coutts, Alexander (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2), 369-395.

Edwards, Ward (1982), Conservatism in Human Information Processing, In *Judgment Under Uncertainty: Heuristic and Biases*, eds. Daniel Kahneman, Paul Slovic, and Amos Tversky, Cambridge, U.K.: Cambridge University Press, 359-369.

Eil, David, and Justin M. Rao (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics*, 3(2), 114-138.

El-Gamal, Mahmoud A., and David M. Grether (1995). Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association*, 90(432), 1137-1145.

Enke, Benjamin and Florian Zimmermann (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1), 313-332.

Ertac, Seda (2011). Does Self-Relevance Affect Information Processing? Experimental Evidence on the Response to Performance and Non-Performance Feedback, *Journal of Economic Behavior & Organization,* 80(3), 532-545.

Georgalos, Konstantinos (2021). Dynamic decision making under ambiguity: An experimental investigation. *Games and Economic Behavior*, 127, 28-46.

Grether, David M. (1980). Bayes Rule as a Descriptive Model: The Representativeness Heuristic. *Quarterly Journal of Economics*, 95(3), 537-557.

Grether, David M. (1992). Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1), 31-57.

Grossman, Zachary, and David Owens (2012), An Unlucky Feeling: Overconfidence and Noisy Feedback, *Journal of Economic Behavior & Organization*, 84(2), 510-524.

Grubb, Michael D. (2009). Selling to Overconfident Consumers. *American Economic Review*, 99(5), 1770-1807.

Heger, Stephanie A., and Nicholas W. Papageorge (2018). We Should *Totally* Open a Restaurant: How Optimism and Overconfidence Affect Beliefs. *Journal of Economic Psychology*, 67, 177-190.

Herz, Holger, Daniel Schunk, and Christian Zehnder (2014). How do Judgmental Overconfidence and Overoptimism Shape Innovative Activity? *Games and Economic Behavior*, 83(C), 1-23.

Holt, Charles A. and Angela M. Smith (2009). An update on Bayesian updating. *Journal of Economic Behavior & Organization*, 69(2), 125-134.

Kahneman, Daniel and Amos Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3), 430-454.

Koszegi, Botond (2006). Ego Utility, Overconfidence, and Task Choice, *Journal of the European Economic Association*, 4(4), 673-707.

Larrick, Richard P., Katherine A. Burson, and Jack B. Soll (2007). Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence (and When it Does Not). *Organizational Behavior and Human Decision Processes*, 102(1), 76-94.

Logg, Jennifer M., Uriel Haran, and Don A. Moore (2018). Is Overconfidence a Motivated Bias? Experimental Evidence. *Journal of Experimental Psychology: General*, 147(10), 1445-1465.

Malmendier, Ulrike, and Geoffrey Tate (2005). CEO Overconfidence and Corporate Investment. *Journal of Finance*, 60(6), 2661-2700.

Menkhoff, Lukas, Maik Schmeling, and Ulrich Schmidt (2013). Overconfidence, Experience, and Professionalism: An Experimental Study. *Journal of Economic Behavior and Organization*, 86(C), 92-101.

Merkle, Christoph, and Martin Weber (2011). True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence. *Organizational Behavior and Human Decision Processes*, 116(2), 262–271.

Moore, Don A., and Paul J. Healy (2008). The Trouble with Overconfidence. *Psychological Review*, 115(2), 502-517.

Ouwersloot, Hans, Peter Nijkamp, and Piet Rietveld (1998). Errors in probability updating behaviour: Measurement and impact analysis. *Journal of Economic Psychology*, 19(5), 535-563.

Proeger, Till, and Lukas Moeb (2014). Overconfidence as a Social Bias: Experimental Evidence. *Economics Letters*, 122(2), 203-207.

Svenson, Ola (1981). Are We All Less Risky and More Skillful than our Fellow Drivers? *Acta Psychologica*, 47(2), 143-148.

Van den Steen, Eric (2011). Overconfidence by Bayesian-Rational Agents. *Management Science*, 57(5), 884-896.

Vosgerau, Joachim (2010). How Prevalent is Wishful Thinking? Misattribution of Arousal Causes Optimism and Pessimism in Subjective Probabilities. *Journal of Experimental Psychology: General*, 139(1), 32-48.

Zizzo, Daniel. J., Stephanie Stolarz-Fantino, Julie Wen, and Edmund Fantino (2000). A violation of the monotonicity axiom: Experimental evidence on the conjunction fallacy. *Journal of Economic Behavior & Organization*, 41(3), 263-276.