Testing the Overconfidence Trap

September 30, 2023

Motivation

Overconfidence seems to be persistent in various settings. Ultimately it leads to suboptimal choices:

- Excess entry of entrepreneurs (Camerer and Lovallo, 1999)
- Suboptimal genetic testing and savings (Oster et al. 2013)
- Workers overestimate their productivity (Hoffman and Burks, 2020)

An Example

An entrepreneur has **unknown intrinsic ability** θ and chooses a level of effort $e \geq 0$.

An overconfident entrepreneur believes θ is larger than the true value.

Effort and ability are transformed into output at an exogenous and **unknown rate** ω.

So the entrepreneur wants to maximize

$$y = (\theta + e)\omega - \frac{1}{2}e^2 + \varepsilon$$

An Example

An entrepreneur has **unknown intrinsic ability** θ and chooses a level of effort $e \geq 0$.

An overconfident entrepreneur believes θ is larger than the true value.

Effort and ability are transformed into output at an exogenous and **unknown rate** ω.

So the entrepreneur wants to maximize

$$y = (\theta + e)\omega - \frac{1}{2}e^2 + \varepsilon$$

Regardless of their own type (or their beliefs about it), they should choose $e^*.(\omega) = \omega$

Learning is Possible

This exercise is myopically repeated for $t = 0, 1, \ldots$

$$y_t = (\theta + e_t)\omega - \frac{1}{2}e_t^2 + \varepsilon_t$$

Note that both parameters are identified in this setting:

- Choosing $\hat{e}$ and $\hat{e} + 1$ allows identification of ω
- Once ω is known, θ can be backed out

How come people don't learn the true θ?

The Literature

Settings with two or more unknowns allow for different explanations of the bias:

1. Self-attribution bias with two unknowns (Coutts et al. 2022 wp):
   - Good news are attributed to high θ bad news are attributed to low ω
2. Bayesian factor test (Ba, 2022 JMP):
   - Bayesian updating on ω
   - Hypothesis testing on θ
3. Self-defeating equilibrium (Heidhues et al., 2018):
   - Never updates beliefs about θ
   - Bayesian on ω

Theory 1

Unrealistic Expectations and Misguided Learning
(Heidhues, Köszegi, and Strack, 2018)

The Setting

ω is drawn from density g_0. And the realized value is $\omega^* = E_{g_0}(\omega)$.

The entrepreneur's true ability is θ^*, they believe with certainty that it is $\hat{\theta} > \theta^*$.

At $t = 0$, the entrepreneur has the prior g_0.

They correctly choose $e_0 = \omega^*$.

The Setting

ω is drawn from density g_0. And the realized value is $\omega^* = E_{g_0}(\omega)$.

The entrepreneur's true ability is θ^*, they believe with certainty that it is $\hat{\theta} > \theta^*$.

At $t = 0$, the entrepreneur has the prior g_0.

They correctly choose $e_0 = \omega^*$.

Suppose they don't update their beliefs or their choice for a long time.

Updating the Beliefs

They observe an average output of

$$y_0 = (\theta^* + \omega^*)\omega^* - \frac{1}{2}(\omega^*)^2$$

But were expecting

$$(\hat{\theta} + \omega^*)\omega^* - \frac{1}{2}(\omega^*)^2 > y_0$$

Updating the Beliefs

They observe an average output of

$$y_0 = (\theta^* + \omega^*)\omega^* - \frac{1}{2}(\omega^*)^2$$

But were expecting

$$(\hat{\theta} + \omega^*)\omega^* - \frac{1}{2}(\omega^*)^2 > y_0$$

So they conclude that ω_1 must be such that:

$$(\hat{\theta} + \omega^*)\omega_1 - \frac{1}{2}(\omega^*)^2 = (\theta^* + \omega^*)\omega^* - \frac{1}{2}(\omega^*)^2$$

Which gives $\omega_1 = \frac{(\theta^* + \omega^*)\omega^*}{(\hat{\theta} + \omega^*)} < \omega^*$

Bayesian updating

Updating choices every period (myopically) the belief will drift even further:

A lower choice of e still gives a lower output than expected.

So ω_{t+1} must be lower than they believed in period t.

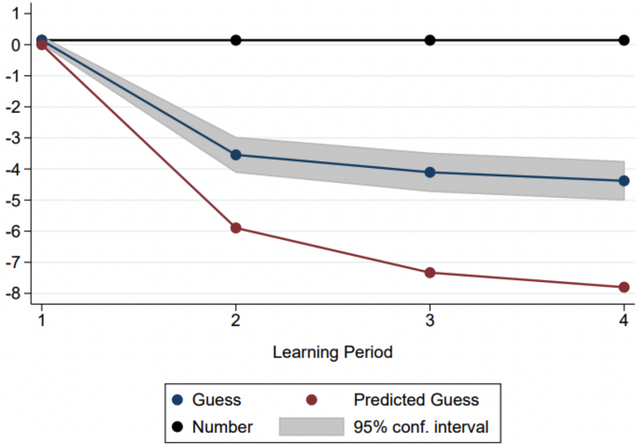
Heidhues et al. show that this process converges to a belief $\omega_\infty < \omega_1 < \omega^*$.

The result is symmetric for underconfident subjects.

In the Lab

Götte and Kozakiewicz (2022) test the predictions in a lab experiment:

Figure 2: Learning of overconfident subjects in multiple-feedback rounds.



Results

Biased updating in the same direction as the theoretical predictions.

The bias is not as large as the theory predicts.

They argue that some agents do update their belief in the ability parameter.

How are people updating their beliefs on two parameters?

The Research Question

We know people fall somewhere between Bayesian Learning and the very extreme Heidhues et al. setting.

- Do the existing alternative theories explain the behavior well?
  - Is there heterogeneity in how people update their beliefs? (Bayesian, self-attribution, hypothesis testing, self-defeating)
- Can we design an intervention that will help people learn their true type?

Theory 2

Robust Misspecified Models and Paradigm Shifts
(Ba, 2022 JMP)

The Setting

Same as Theory 1

Now the entrepreneur entertains an alternative level of ability θ' (assume $\theta' = \theta^*$).

Instead of updating $P[\theta]$ every period, they perform a Bayesian hypothesis test:

Adopt model θ' at time t iff

$$\frac{\ell_t(\theta')}{\ell_t(\theta)} > \alpha \geq 1$$

Where

$$\ell_t(\theta) := \sum_{\omega} g_0(\omega) \prod_{\tau=0}^{t-1} \pi^{\theta}(y_\tau | a_\tau, \omega)$$

Results

The self-defeating trap is difficult to escape when the entrepreneur is overconfident.

Underconfident agents escape the trap at a higher rate than overconfident and learn the true parameter.

Explains why underconfidence is less prevalent in the data.

Overconfident agents can escape the trap if their prior is not too "tight" around a self confirming equilibrium.

Theory 3: Motivated Beliefs

Optimal Expectations
(Brunnermeier and Parker, 2005)

What to blame? Self-serving attribution bias with multi-dimensional uncertainty
(Coutts et al., 2022 wp)

The Setting

Assume e is fixed and $\theta \in \{\underline{\theta}, \bar{\theta}\}$

Let the probability of success be increasing in θ and ω and denote it by by

$$q(\theta, \omega) := Pr[success|\theta, \omega]$$

After a success the agent updates their belief about θ with distortion γ_θ:

$$p_{t+1}(\bar{\theta}|success) = \frac{\gamma_\theta \pi_t(\bar{\theta}) \sum_\omega g_t(\omega) q_t(\bar{\theta}, \omega)}{\gamma_\theta \pi_t(\bar{\theta}) \sum_\omega g_t(\omega) q_t(\bar{\theta}, \omega) + \pi_t(\underline{\theta}) \sum_\omega g_t(\omega) q_t(\underline{\theta}, \omega)}$$

Predictions and Evidence (Coutts et al. 2022)

Prediction: Even unbiased agents will overweight θ_H after a success and end up being biased.

Experiment: Evidence of biased updating when θ is ego-relevant.

The framework does not allow direct comparisons with the other two theoretical predictions.

It rationalizes the formation of overconfident beliefs.

A framework where we can compare

Theories 1 and 2 draw conclusions within the same setting already.

Can we bring the self-serving attribution bias into that setting?

1. A larger α in Theory 2 is consistent with self-serving beliefs when the update is done by evaluating the likelihood ratio.
2. We can write out the fully Bayesian update within the setting and introduce the bias as in Theory 3.
3. If some subject starts out with the correct belief θ and ends up being overconfident, it is evidence of a self-serving bias.

A Theory-Inspired intervention

The parameters in this setting are identified. Allowing for a trade-off between exploitation and experimentation is sufficient for learning.

There are two ways in which experimentation can lead to correct learning:

1. Experimentation on e:
   - Götte and Kozakiewicz show that some agents do learn (are they experimenting?)
   - Would overconfident/underconfident subjects choose to experiment if the option is made explicitly available to them?

2. Allowing experimentation over ω (switching environments):
   - Hesterman and Le Yaouanq (2021) show that in a model with fixed e and the possibility to experiment in different environments, overconfident subjects learn θ faster than unbiased/underconfident subjects

Testing in the lab

Part 1: Test

- Get the ego-relevant parameter
- Elicit beliefs

Testing in the lab

Part 1: Test

- Get the ego-relevant parameter
- Elicit beliefs

Part 2: Investment game (N rounds)

- Treatment is paid according to their own score as θ
- Control is paid according to an ego-neutral parameter

1. Choose an amount e to invest
2. Returns are $\theta\omega + e(\omega - \frac{1}{2}e) + \varepsilon_t$
3. Belief elicitation

Testing in the lab

Part 1: Test

- Get the ego-relevant parameter
- Elicit beliefs

Part 2: Investment game (N rounds)

- Treatment is paid according to their own score as θ
- Control is paid according to an ego-neutral parameter

1. Choose an amount e to invest
2. Returns are $\theta\omega + e(\omega - \frac{1}{2}e) + \varepsilon_t$
3. Belief elicitation

Part 3: Explicitly allow experimentation

- Draw a random e for the next n rounds
- Draw a new return parameter ω

What needs to be done

1. Define self-serving attribution bias in the framework of theories 1 and 2 (??)

2. Get predictions in a simplified environment (finite parameter sets).

3. Parameters should be such that there are differences among models:
   - Subjects update in Theory 2 but fall into self-defeating eq. in Theory 1
   - Self-serving beliefs make the default model more sticky in Theory 2
   - Self serving bias over weights high θ relative to the Bayesian
   - Under self serving bias, unbiased subjects become biased. And overconfident subjects become more confident after a success

The end

Thank you!