

Effective Altruism and Longtermism

I hope that the posters worked

Middlebury Effective Altruism

Slides available on [go/easlides](https://go.easlides)

Fall 2022

Table of Contents

- 1 Your career is important
- 2 Effectiveness
- 3 Ethical Framework of Impact
- 4 October 27
- 5 Intro, again
- 6 Measuring lives
- 7 Longtermism

Top three career facts

- 1 You will spend more time on your career than everything except sleep.

Top three career facts

- 1 You will spend more time on your career than everything except sleep.
- 2 You will have a much bigger impact on the world in some career than others

Top three career facts

- 1 You will spend more time on your career than everything except sleep.
- 2 You will have a much bigger impact on the world in some career than others
- 3 We live in a time where there are many pressing issues, and one where one person can make a lot of difference.

80,000 hours's four career rules

- 1 Solve a pressing issue: Something with scale and understaffed

80,000 hours's four career rules

- 1 Solve a pressing issue: Something with scale and understaffed
- 2 Find a more effective solution

80,000 hours's four career rules

- 1 Solve a pressing issue: Something with scale and understaffed
- 2 Find a more effective solution
- 3 Get more leverage/influence/scale.

80,000 hours' four career rules

- 1 Solve a pressing issue: Something with scale and understaffed
- 2 Find a more effective solution
- 3 Get more leverage/influence/scale.
- 4 Find work that suits you.

Impact can be measured

- This might feel obvious, but it's worth thinking about.

Impact can be measured

- This might feel obvious, but it's worth thinking about.
- One of the key drivers behind the scientific revolution is that the physical world around us can be studied through experiments.

Impact can be measured

- This might feel obvious, but it's worth thinking about.
- One of the key drivers behind the scientific revolution is that the physical world around us can be studied through experiments.
- So it's worth re-iterating this again: Your career impact can be measured.

Day to day examples

- If you stop using plastic bags for the rest of your life, you reduce your carbon footprint by about 100kg.

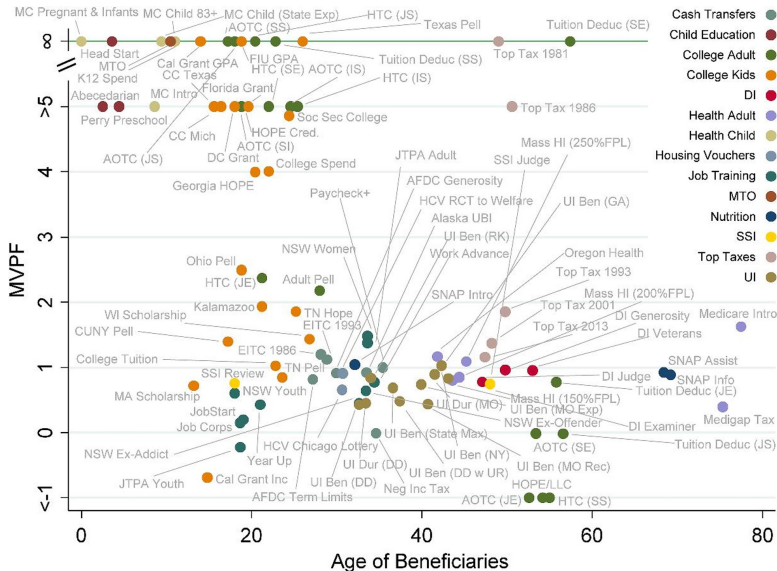
Day to day examples

- If you stop using plastic bags for the rest of your life, you reduce your carbon footprint by about 100kg.
- If you take one less round trip flight to London, you reduce your carbon footprint by 1400kg

Policy examples

- Drug Abuse Resistance Education: 10 million dollars annual budget, doesn't work at all.
- Scared Straight: Every dollar spent caused \$200 worth of damage

Policy examples: The number one graph [\[link\]](#)



Being good at your job matters [link]

Data on the dispersion of staff productivity			
Field & outcome	Source	Share of output from the top...	
		20%	1%
Output in "low" complexity jobs among applicants e.g. mail carrier	Hunter, Schmidt, & Judiesch 1990	51%	3%
Output in "medium" complexity jobs among applications e.g. cook	Hunter, Schmidt, & Judiesch 1990	58%	4%
Papers coauthored by mathematicians with at least 133 publications	Clauset et al. 2009	33%	4%
Papers written by scientist (whole career)	Sinatra et al. 2016	39%	4%
Weeks in Billboard Hot-100 (1970-2018) by musician, among artists with at least 282 weeks in these charts	Tauberg 2018	35%	5%
Box Office Gross by US top-200 movie director	Tauberg 2018	40%	7%
Citations to scientists (whole career)	Sinatra et al. 2016	51%	7%
Income (worldwide, 2005)	Anand & Segal 2014		21%
Weeks on NYT Fiction Bestseller list by author with at least 6 weeks on that list	Tauberg 2018	76%	46%
Startup founder equity by company, among Y Combinator companies	80,000 Hours 2014		>80%

Being good at your job matters

- Replacing a teacher from the bottom 5% with one from the top 5% results in an additional 1.4 million worth of earnings for the average classroom.

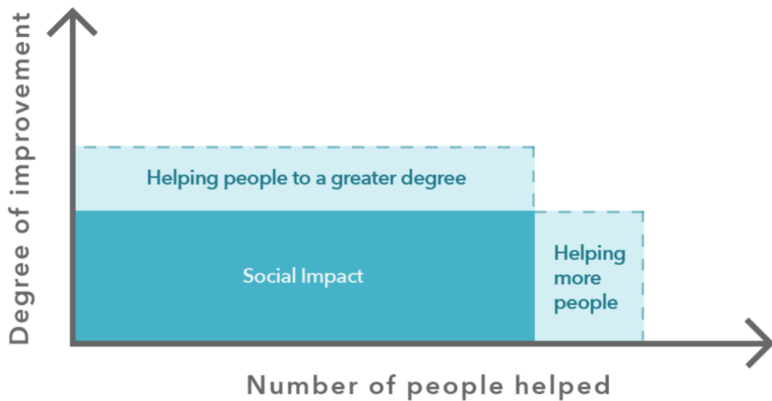
How to define social impact

- “Your social impact is given by the number of people whose lives you improve and how much you improve them, over the long term.”

How to define social impact

- “Your social impact is given by the number of people whose lives you improve and how much you improve them, over the long term.”
- “Social impact is about promoting total expected wellbeing — considered impartially, over the long term — without sacrificing anything that might be of comparable moral importance”

Two ways to have more social impact



Promoting well-being

- “Promoting” implies a positive action.
- When thinking about ethical behaviors we often think about stuff that we *should not do*.

Promoting well-being

- “Promoting” implies a positive action.
- When thinking about ethical behaviors we often think about stuff that we *should not do*.
- “Well-being” here is more complicated to define

Promoting well-being

- “Promoting” implies a positive action.
- When thinking about ethical behaviors we often think about stuff that we *should not do*.
- “Well-being” here is more complicated to define
 - Physical well-being (i.e., life expectancy)
 - Mental well-being

Promoting well-being

- “Promoting” implies a positive action.
- When thinking about ethical behaviors we often think about stuff that we *should not do*.
- “Well-being” here is more complicated to define
 - Physical well-being (i.e., life expectancy)
 - Mental well-being
 - Happiness (Are you living the best possible life for you?)
 - Fulfillment
 - Hope
 - Sense of autonomy/freedom

Promoting well-being

- “Promoting” implies a positive action.
- When thinking about ethical behaviors we often think about stuff that we *should not do*.
- “Well-being” here is more complicated to define
 - Physical well-being (i.e., life expectancy)
 - Mental well-being
 - Happiness (Are you living the best possible life for you?)
 - Fulfillment
 - Hope
 - Sense of autonomy/freedom
- How do we trade off different forms of well-being with each other?

Total expected well-being

- Total here implies that we're adding up well-being accumulated from everyone you impact

Total expected well-being

- Total here implies that we're adding up well-being accumulated from everyone you impact
- EA likes to pretend that we weigh lives equally, but that's not entirely a well-defined concept either

Total expected well-being

- Total here implies that we're adding up well-being accumulated from everyone you impact
- EA likes to pretend that we weigh lives equally, but that's not entirely a well-defined concept either
- Suppose you can save only one of two people drowning in a lake, who are of the same age. One of which, if you save, will go on and save 10 other people throughout their lifetime. The other will go on to save 100 other people throughout their lifetime.

Total expected well-being

- Total here implies that we're adding up well-being accumulated from everyone you impact
- EA likes to pretend that we weigh lives equally, but that's not entirely a well-defined concept either
- Suppose you can save only one of two people drowning in a lake, who are of the same age. One of which, if you save, will go on and save 10 other people throughout their lifetime. The other will go on to save 100 other people throughout their lifetime.
- Small diversion to talk about scope insensitivity.

Scope insensitivity

- “Once upon a time, three groups of subjects were asked how much they would pay to save 2,000 / 20,000 / 200,000 migrating birds from drowning in uncovered oil ponds. The groups respectively answered \$80, \$78, and \$88”

Scope insensitivity

- “Once upon a time, three groups of subjects were asked how much they would pay to save 2,000 / 20,000 / 200,000 migrating birds from drowning in uncovered oil ponds. The groups respectively answered \$80, \$78, and \$88”
- We do a very bad job at conceptualizing large numbers.

Expected value

- $E[X] = \int_{\Omega} xp(x)dx$

Expected value

- $E[X] = \int_{\Omega} xp(x)dx$
- Fancy average

Expected value

- $E[X] = \int_{\Omega} xp(x)dx$
- Fancy average
- If I have a large number of college students who have an expected lifetime earnings of \$5 million dollars each. The idea is that, after they die, I could average their actual lifetime earnings, it should come down to about \$5 million each.

Expected value

- $E[X] = \int_{\Omega} xp(x)dx$
- Fancy average
- If I have a large number of college students who have an expected lifetime earnings of \$5 million dollars each. The idea is that, after they die, I could average their actual lifetime earnings, it should come down to about \$5 million each.
- Suppose that I'm a good teacher, then maybe I could increase their expected earnings to \$6 million dollars.

Expected value with discrete choices

- In the case with lifetime earnings, expected impact makes a lot of sense because there is a continuum of possibilities.

Expected value with discrete choices

- In the case with lifetime earnings, expected impact makes a lot of sense because there is a continuum of possibilities.
- Now think about this scenario: You have a group of 20 people sitting in a room.

Expected value with discrete choices

- In the case with lifetime earnings, expected impact makes a lot of sense because there is a continuum of possibilities.
- Now think about this scenario: You have a group of 20 people sitting in a room.
- We are going to flip a coin. If it comes up as heads, everyone dies. If it comes up as tails, no one dies.

Expected value with discrete choices

- In the case with lifetime earnings, expected impact makes a lot of sense because there is a continuum of possibilities.
- Now think about this scenario: You have a group of 20 people sitting in a room.
- We are going to flip a coin. If it comes up as heads, everyone dies. If it comes up as tails, no one dies.
- Suppose you worked really hard and changed the rules. Now you have to flip two coins, and everyone dies if and only if the two coins both come up as heads.

Expected value with discrete choices

- In the case with lifetime earnings, expected impact makes a lot of sense because there is a continuum of possibilities.
- Now think about this scenario: You have a group of 20 people sitting in a room.
- We are going to flip a coin. If it comes up as heads, everyone dies. If it comes up as tails, no one dies.
- Suppose you worked really hard and changed the rules. Now you have to flip two coins, and everyone dies if and only if the two coins both come up as heads.
- Before, ten people are expected to die. Now, only five people are expected to die.

Expected value with discrete choices

- In the case with lifetime earnings, expected impact makes a lot of sense because there is a continuum of possibilities.
- Now think about this scenario: You have a group of 20 people sitting in a room.
- We are going to flip a coin. If it comes up as heads, everyone dies. If it comes up as tails, no one dies.
- Suppose you worked really hard and changed the rules. Now you have to flip two coins, and everyone dies if and only if the two coins both come up as heads.
- Before, ten people are expected to die. Now, only five people are expected to die.
- Did I save five lives?

Follow up on scope insensitivity from Dickert et al. (2015)

- “If I look at the mass I will never act. If I look at one, I will.” - Mother Teresa
- Emotions like sympathy and compassion are stronger when it is easy to visually attend to or mentally imagine the victims.
- Genevsky et al. (2013) finds that when they are shown a photo of the victim as opposed to a silhouettes the reward centers of their brains are activated.
- In fact, telling participants that people tend to “react more strongly to specific people who have problems than to statistics about people with problems” lower their donations in the case where a specific individual (Small et al., 2007).

The original Boyle et al. (1994)

“In 1989, for example, about 2000/20,000/200,000 migratory waterfowl died in these holding ponds. This was less than 1 percent of the 8.5 million migratory waterfowl in the Central Flyway. The affected migratory waterfowl include: mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes.”

- The paper doesn't actually give the questionnaire

The original Boyle et al. (1994)

“In 1989, for example, about 2000/20,000/200,000 migratory waterfowl died in these holding ponds. This was less than 1 percent of the 8.5 million migratory waterfowl in the Central Flyway. The affected migratory waterfowl include: mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes.”

- The paper doesn't actually give the questionnaire
- "This is outrageous"

The original Boyle et al. (1994)

“In 1989, for example, about 2000/20,000/200,000 migratory waterfowl died in these holding ponds. This was less than 1 percent of the 8.5 million migratory waterfowl in the Central Flyway. The affected migratory waterfowl include: mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes.”

- The paper doesn't actually give the questionnaire
- I think the idea is that they are covering ponds of oil that birds would drown in.

Results table

TABLE II
Descriptive Statistics for Experimental Treatments^a

	2000 birds	20,000 birds	200,000 birds
Mean	\$80	\$78	\$88
Standard deviation	11	8	10
Median	25	25	25
Mode	0	100	100
Range	0–1550	0–1000	0–1000
Shapiro–Wilk statistic ^b	0.43	0.60	0.54
<i>N</i>	288	286	281

^aOutliers removed from data.

^bThis test statistic indicates that these distributions are not normal.

Follow up on happiness and money

Figure: I saw this pic on a CNBC article and I had to include it.



Jebb et al. (2018)

- Gallup world poll using 1.7 million respondents worldwide
- In North America, life evaluation maxed out at \$105,000, positive affect maxed out at \$65,000, and negative affect on \$95,000.

Clingingsmith (2016)

a

Effective Altruism and Longtermism

I hope that the posters worked

Middlebury Effective Altruism

Slides available on [go/easlides](https://go.easlides)

Fall 2022

Basic overview of Effective Altruism

- Why don't we go around and each say what Effective Altruism means to us?

Basic overview of Effective Altruism

- Why don't we go around and each say what Effective Altruism means to us?
- Here is perhaps the more canonical introduction:

Basic overview of Effective Altruism

- Why don't we go around and each say what Effective Altruism means to us?
- Here is perhaps the more canonical introduction:
- We all want to help people

Basic overview of Effective Altruism

- Why don't we go around and each say what Effective Altruism means to us?
- Here is perhaps the more canonical introduction:
- We all want to help people
- We have the same obligation to help a person whether they are drowning in front of us or dying from Malaria thousands of miles away

Basic overview of Effective Altruism

- Why don't we go around and each say what Effective Altruism means to us?
- Here is perhaps the more canonical introduction:
- We all want to help people
- We have the same obligation to help a person whether they are drowning in front of us or dying from Malaria thousands of miles away
- It is better to help more people than less

Basic overview of Effective Altruism

- Why don't we go around and each say what Effective Altruism means to us?
- Here is perhaps the more canonical introduction:
- We all want to help people
- We have the same obligation to help a person whether they are drowning in front of us or dying from Malaria thousands of miles away
- It is better to help more people than less
- We have limited resources, so it's important to do as much as we can with what we have

Cause areas

- Effective Altruists (EAs) have certain areas they like to focus on

Cause areas

- Effective Altruists (EAs) have certain areas they like to focus on
- Global Health and Development: We love Malaria Nets and Vitamin A tablets

Cause areas

- Effective Altruists (EAs) have certain areas they like to focus on
- Global Health and Development: We love Malaria Nets and Vitamin A tablets
- Extinction risk reduction: AI Alignment and Biorisks

Cause areas

- Effective Altruists (EAs) have certain areas they like to focus on
- Global Health and Development: We love Malaria Nets and Vitamin A tablets
- Extinction risk reduction: AI Alignment and Biorisks
- Farm Animal/Wild Animal welfare

Cause areas

- Effective Altruists (EAs) have certain areas they like to focus on
- Global Health and Development: We love Malaria Nets and Vitamin A tablets
- Extinction risk reduction: AI Alignment and Biorisks
- Farm Animal/Wild Animal welfare
- If you disagree about a certain area or approach to solving a problem (e.g., RCTs v. growth), that's still effective altruism! EA is about an approach to life, not a specific cause area.

Cause areas

- Effective Altruists (EAs) have certain areas they like to focus on
- Global Health and Development: We love Malaria Nets and Vitamin A tablets
- Extinction risk reduction: AI Alignment and Biorisks
- Farm Animal/Wild Animal welfare
- If you disagree about a certain area or approach to solving a problem (e.g., RCTs v. growth), that's still effective altruism! EA is about an approach to life, not a specific cause area.
- We mostly focusing on helping everything think about having an impactful career.

Where does EA spend its money? [link]

Cause Area	\$ millions per year in 2019	%
Global health	185	44%
Farm animal welfare	55	13%
Biosecurity	41	10%
Potential risks from AI	40	10%
Near-term U.S. policy	32	8%
Effective altruism / rationality / cause prioritisation	26	6%
Scientific research	22	5%
Other global catastrophic risk (inc. climate tail risks)	11	3%
Other long term	1.8	0%
Other near-term work (near-term climate change, mental health)	2	0%
Total	416	100%

What do EAs do?

For those who answered '5' for engagement, the breakdown was:

5-engaged EAs cause currently working in (normalised)	↑ % ↓
AI	18
Movement building	15
Rationality	12
Other near term (Near-term climate change, mental health)	12
Cause prioritisation	10
Other GCRs	10
Animal welfare	10
Global poverty	6
Biosecurity	4
Other	4
Total	101

Figure: Note that these are people who are extremely engaged with the movement

How do EAs think about measuring their outcome [link]



EA Forum



Tim Hua



Global health and development

Measuring Good Better

by MichaelPlant, GiveWell, Jason Schukraft, Matt_Lerner,
Innovations for Poverty Action

18 min read 14th Oct 2022 18 comments ...

209

Measuring and comparing value

GiveWell

Open Philanthropy

Happier Lives Institute

Founders Pledge

Innovations for Poverty Action

Moral weight

Wellbeing

Adjusted life year

Subjective wellbeing

Global health and development

Cause prioritization

Curated

+ Add Topic

At EA Global: San Francisco 2022, the following organisations held a joint session to discuss their different approaches to measuring 'good':



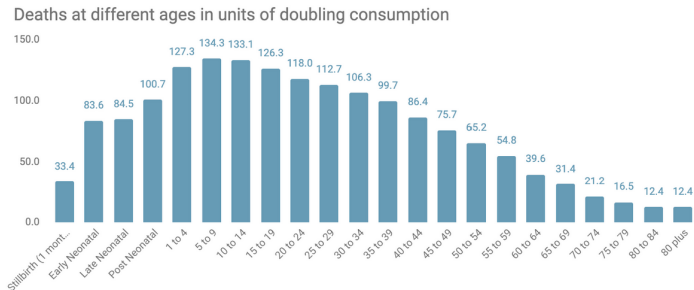
Give Well's moral weights

This is 60% from donor surveys, 30% from a survey of people living in extreme poverty in Kenya and Ghana, and 10% staff opinion.

Value of doubling consumption for one person for one year	1.0
Value of averting one year of life lived with disease/disability (YLD)	2.3
Value of averting one stillbirth (1 month before birth)	33.4
Value of preventing one 5-and-over death from malaria	83.1
Value of averting one neonatal death from syphilis	84.0
Value of preventing one under-5 death from malaria	116.9
Value of preventing one under-5 death from vitamin A deficiency	118.4

More from GiveWell

Figure 1: GiveWell's moral values of deaths at different ages (in units of doubling consumption)



Longtermism

- While Radical Empathy focuses on helping people near and far, Longtermism says that we need to extend our empathy to people in the future, even if they don't exist yet

Longtermism

- While Radical Empathy focuses on helping people near and far, Longtermism says that we need to extend our empathy to people in the future, even if they don't exist yet
- “To see how intuitive this is, suppose that, while hiking, I drop a glass bottle on the trail and it shatters. And suppose that if I don't clean it up, later a child will cut herself badly on the shards. In deciding whether to clean it up, does it matter when the child will cut herself? Should I care whether it's a week, or a decade, or a century from now? No. Harm is harm, whenever it occurs.”

Formal definition

Longtermism is the view that:

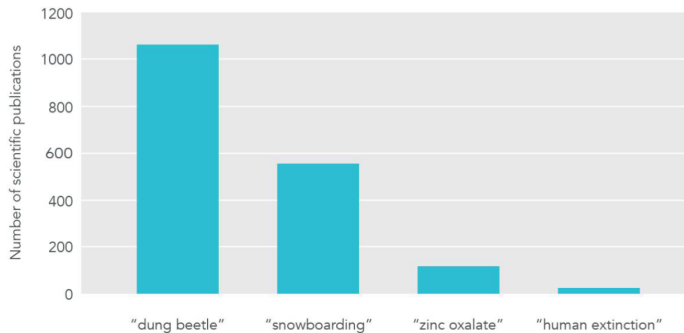
- Those who live at future times matter just as much, morally, as those who live today
- Society currently privileges those who live today above those who will live in the future
- We should take action to rectify that, and help ensure the long-run future goes well.

Our future is vast

See title link

One of the most depressing graphs

If we look at scientific attention instead, we see a similar picture of neglect (though, some of the individual risks receive significant attention, such as climate change):



Existential risks are quite concerning

- It's one am and I've been reading about them for the last hour

Existential risks are quite concerning

- It's one am and I've been reading about them for the last hour
- Summary: Not pog

Tim's take: We don't really need Longtermism

- Everything that's bad when you believe in longtermism is bad even if you don't believe in it

Tim's take: We don't really need Longtermism

- Everything that's bad when you believe in longtermism is bad even if you don't believe in it
- You don't need 500IQ to realize that reducing extinction risk is a good thing.

Tim's take: We don't really need Longtermism

- Everything that's bad when you believe in longtermism is bad even if you don't believe in it
- You don't need 500IQ to realize that reducing extinction risk is a good thing.
- Personally, whether or not I should be a Longtermist is one of those questions that, even if I knew the answer to, wouldn't really change my actions much

Tim's take: We don't really need Longtermism

- Everything that's bad when you believe in longtermism is bad even if you don't believe in it
- You don't need 500IQ to realize that reducing extinction risk is a good thing.
- Personally, whether or not I should be a Longtermist is one of those questions that, even if I knew the answer to, wouldn't really change my actions much
- So I sort of ignore it.

Random variables are not reality

- Whenever we say that X has some percent chance of happening, we are treating X as a random variable.

Random variables are not reality

- Whenever we say that X has some percent chance of happening, we are treating X as a random variable.
- When we're talking about the outcome of a dice roll, modeling it as a random variable makes a lot of sense.

Random variables are not reality

- Whenever we say that X has some percent chance of happening, we are treating X as a random variable.
- When we're talking about the outcome of a dice roll, modeling it as a random variable makes a lot of sense.
- RVs are actually very fancy math objects, whose formal definition requires that you take multivariable calc, linear algebra, and then measure theory to understand because it's honestly a deranged object.

Random variables are not reality

- Whenever we say that X has some percent chance of happening, we are treating X as a random variable.
- When we're talking about the outcome of a dice roll, modeling it as a random variable makes a lot of sense.
- RVs are actually very fancy math objects, whose formal definition requires that you take multivariable calc, linear algebra, and then measure theory to understand because it's honestly a deranged object.
- I don't think it makes sense to treat extinction risks as random variables and use expected values.

Taken from Boaz [link]

- The probabilities are too small to reason about.
- I cannot justify sacrificing current living humans for abstract probabilities.
- We cannot improve what we cannot measure.
- Long-term risk mitigation can only succeed through short-term progress.
- Summary: think of the long term, but act and measure in the short term.

Go forth and discuss

- When it comes to the future, what do you worry about the most?
- What is most holding you back? What are your biggest bottlenecks?
- What is the biggest mistake you've made in the last 6 months that you feel comfortable sharing? Have you successfully changed the default path of the world such that such a mistake is very unlikely to happen again?
- Suppose you leave this conversation with some concrete action points that are obviously good ideas that will majorly improve your life. What are they? Make them concrete and give deadlines.
- Are you procrastinating on anything?
- What is the best and worst thing in your life right now?
- What important truth do very few people agree with you on?
- What is the best compliment you have ever received?
- Do you trust people more if they're socially awkward?
- Do you get overly affected by social validation?
- What is the most important social pressure that you feel you had to constantly resist?
-

Resources

Bibilography I

- Boyle, K. J., W. H. Desvousges, F. Johnson, R. W. Dunford, and S. P. Hudson (1994). An investigation of part-whole biases in contingent-valuation studies. *Journal of Environmental Economics and Management* 27(1), 64–83.
- Clingingsmith, D. (2016). Negative emotions, income, and welfare: Causal estimates from the psid. *Journal of Economic Behavior and Organization* 130, 1–19.
- Dickert, S., D. Västfjäll, J. Kleber, and P. Slovic (2015). Scope insensitivity: The limits of intuitive valuation of human lives in public policy. *Journal of Applied Research in Memory and Cognition* 4(3), 248–255. Modeling and Aiding Intuition in Organizational Decision Making.
- Genevsky, A., D. Västfjäll, P. Slovic, and B. Knutson (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *Journal of Neuroscience* 33(43), 17188–17196.
- Jebb, A. T., L. Tay, E. Diener, and S. Oishi (2018). Happiness, income satiation and turning points around the world. *Nature Human Behaviour* 2(1), 33–38.
- Small, D. A., G. Loewenstein, and P. Slovic (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes* 102(2), 143–153.