Decisions, Big & Small -- Spring 2024

Time: Tuesdays and Thursdays, 10:30-11:45
Location: WJH 105
Instructor: Tomer Ullman (tullman@fas.harvard.edu)
Student Hours: Mondays, 2:00-3:00pm, Room 190.02 (Northwest)

Sections: TBA
Teaching Fellows and sections:

Felix Sosa (fsosa@fas.harvard.edu)
Section times: TBD
Section location: TBD
Office hours: On demand

Final Exam date: TBD

Overview: Life is full of decisions, but not all decisions are made equal. Choices can be big and consequential (should I focus on my success, family, or passion), or small and everyday (going out, or staying in). This course will introduce you to the cognitive science of judging and choosing. You will learn about rational planning, the kind a perfect intelligence might carry out; Common simplifications and shortcuts that non-perfect humans use, and how these may actually be appealing approximations for any decision-making system; Regret over choices taken and not taken; Making decisions with others, Transformative decisions, the ones that change who you are as a person. As we cover these topics, we will consider how to apply the insights from the psychology of decision making to your own ordinary and extraordinary choices.

Textbook: The course does not have a specific textbook. Reading materials from textbooks, books, articles, journals, and so on will be made available online.

Objectives: The main objective is to acquaint the students with the basic tenets of different approaches and models to decision making, relevant for their future research in cognitive science and psychology, as well as in their daily life. Students should also gain an appreciation for outstanding debates regarding optimal decision making, boundedly rational decision making, and deviations from optimality.

Website: We will make use of Canvas, and it will contain readings, announcements, links, assignments, and grades.

Accessibility: Any student needing academic adjustments or accommodations is requested to present a letter from the Accessible Education Office (AEO) and speak with Collin Conwell by the second week of classes. Failure to do so may result in our being unable to respond to your needs in a timely manner. All discussions will remain confidential, although AEO may be consulted to discuss appropriate implementation.

Grading and Requirements:

Discussion Posts and Replies: 32%
Final Thing: 18%

Final Exam: 40%
Attendance and participation: 10%

Grading Scale:

A: 100-94	A-: 90-93	B+: 89-86	B: 85-80
C+: 79-76	C: 75-70	D: 69-62	
F: 61 and below			

Exams: The exam is close-book. It will consist of a mixture of mostly multiple-choice questions and some short-answer questions. The exam will be given during exam period and last approximately three hours.

Discussion posts and replies: 32% of your grade will be based on your participation in online discussions outside of class, specifically Posts and Replies. Questions based on the readings and lecture will be put on Canvas weekly, and you will have the option of posting your thoughts on the relevant question. These questions are meant to be relatively open-ended, without a strict 'correct/incorrect' response.

> *Posts*: A good Post is one that engages thoughtfully with the material of that week, by raising points not covered, weakness and strengths of different viewpoints covered, outside empirical evidence, suggestions for new studies, and so on. You should submit at least 4 Posts throughout the term of around 150-400 words each. You can submit up to 6 (grading will be based on the top 4 Posts).
>
> *Replies*: In addition to making Posts of your own, you will be asked to reply to other people's Posts. A Reply can be shorter or longer than the original Post, but should engage with it *thoughtfully* and *respectfully*. A reply that amounts to "Yeah!" or "Says you!" is *not* a good reply. As with Posts, you should submit at least 4 Replies throughout the term, but can submit up to 6 (grading will be based on the top 4 Replies).

Final Thing: Submit some *thing* that explains or addresses one topic from the class, and potentially relates it to a decision you yourself need to make or have made. This can be anything! Past submissions have included animation shorts, graphic novel excerpts, a final paper, a theater piece, dance, novel experimental data, podcast episodes, anything. You can address open questions or pull in additional readings. You are encouraged to think empirically and creatively. Please discuss your proposed Thing with the instructor or a TA two weeks or so before the end of classes to make sure you are on the right track.

Policies

Academic integrity: This course adheres to the university's standards regarding academic integrity. Suspected cheating or plagiarism will be referred to the Honor Council of Harvard College, as is required by the university. Students are responsible for knowing what constitutes plagiarism; please refer to the Harvard Guide to Using Sources for a detailed description of the different types of plagiarism.

Lectures and attendance: You are expected to attend the lectures. Lectures will have associated handouts on Canvas, but this is not the same as attending the lecture and participating in exercises and discussions. Students who cannot attend the class should write to the TF or instructor to accommodate.

Attention splitting and note taking: Classes will be taking place in person. The use of laptops and cellphones is *strongly* discouraged, unless the instructors explicitly mention their use for a class exercise. This policy is for the benefit of the students.

Email Policy and scheduling meetings: Professor Ullman is happy to meet with any student for any reason, and you are encouraged to come to student hours. When asking to meet at an alternate time outside student hours, please include several alternative times and a description of the reason for meeting. Questions having to do with the syllabus or assignments are best shared with the class, and so it is better to ask these during or right after class.

Use of AI, such as ChatGPT: Your words and thoughts in the discussion posts and replies, as well as in the Final Thing, are expected to be your own. The best way to think about the use of ChatGPT in this class is that it is similar to paying someone in a different country to do your work for you: imagine how it would go if it was found out that you paid someone to write and submit your work for you, that is how it will be treated. That said, you are welcome to use any online tool you like (including ChatGPT) to check your understanding of the material, with the usual caveats about any online source.

Schedule, Readings

(Readings are subject to change, especially as the class progresses)

Prologue

Tuesday Jan 23st: STOP! The first decision; Optimal stopping.

- *"Algorithms to Live By", Chapter 1, Christian and Griffiths.*

ACT I: The Decision-Making Machine

Thursday Jan 25th: **How Should a Balance Beam Decide?** Costs and Rewards; Decisions as Weighing of Options under Uncertainty; Signal Detection Theory

- A decision-making theory of visual detection, Tanner and Swets 1954
- Macmillan, N. A. (2002). Signal detection theory. Stevens' handbook of experimental psychology, 4, 43-90 [parts 1-3, but don't get caught in the details]

Tuesday Jan 30th: How Should a Balance Beam Decide (cont)? Costs and Rewards; Decisions as Weighing of Options under Uncertainty; Signal Detection Theory

- Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). Detection theory: A user's guide. Routledge.
- Wixted, J. T. (2020). The forgotten history of signal detection theory. Journal of experimental psychology: learning, memory, and cognition, 46(2), 201.
- "I had a funny feeling in my gut", interview with Stanislav Petrov

Thursday Feb 1st: **How Should a Robot Decide?** The Fundamental Equation of AI; Expected Utility; Simple Choices

- Artificial Intelligence, a Modern Approach, 3rd edition, Chapter 16: up until 16.3.2)
- Skim Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk.)

- skim the wiki page https://en.wikipedia.org/wiki/Von_Neumann–Morgenstern_utility_theorem)

Tuesday Feb 6th: **Whose Goals are These Anyway?** The Construction of Goals and the Value Alignment Problem

- Superintelligence: Paths, Dangers, Strategies (2014), Bostrom
- Human Compatible: AI and the problem of Control (2019), Russell
- https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/
- Bonus, play: https://www.decisionproblem.com/paperclips/index2.html

Thursday Feb 8th : The Ant and the Wind; Waiting for the World to Decide; the Accumulation of Evidence; Drift Diffusion and Sequential Sampling

- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. Trends in neurosciences, 27(3), 161-168.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. Trends in cognitive sciences, 20(4), 260-281.

Tuesday Feb 13th: **The Ant and the Wind;** Waiting for the World to Decide; the Accumulation of Evidence; Drift Diffusion and Sequential Sampling

- The Drunkard's Walk, Mlodinow + previous readings in more depth

INTERMISSION

Thursday Feb 15th: Transition/game/simulation

ACT II: The Machine Breaks Down

Tuesday Feb 20th: The Battle-cry of Behavioral Economics: Prospect Theory

- "Thinking, Fast and Slow", Prospect Theory, Kahneman
- "Prospect Theory: An Analysis of Decision under Risk" (Kahneman and Tversky, 1979)
- "Misbehaving", Thaler

Thursday Feb 22nd: Silly Utility: More Prospect theory and the Endowment effect

- "Thinking, Fast and Slow", Endowment Effect, Kahneman
- "Misbehaving", Endowment Thaler

Tuesday Feb 27th: **Would you Bet Your Life on it?** Probability weighting, Framing

- "Thinking, Fast and Slow", Fourfold Pattern and Rare events, Kahneman
- The probability weighting function, Prelec, 1998

Thursday Feb 29th: **Nudge, Nudge, Wink Wink**: Nudges and their Discontents

- "Nudge" by Thaler & Sunstein
- Sunstein, C. R. (2017). Nudges that fail. Behavioural public policy, 1(1), 4-25.
- Loewenstein, G., & Chater, N. (2017). Putting nudges in perspective. Behavioural Public Policy, 1(1), 26-53.
- Sunstein, C. R. (2015). The ethics of nudging.

- https://bppblog.com/2017/06/02/much-ado-about-nudging/
- https://www.nytimes.com/2010/05/16/magazine/16Sunstein-t.html

Tuesday March 5th **Decisions for My Future Self:** The Effect of Time on Decisions; Hyperbolic Discounting

- Breakdown of Will, Chapter 3 (Ainslee)
- Urminsky, O. (2017). The role of psychological connectedness to the future self in decisions over time. Current Directions in Psychological Science, 26(1), 34-39.

Students choose at least one of the following:

- Schelling: Self Command in Practice, in Policy, and in a Theory of Rational Choice
- Ersner-Hershfield, H., Wimmer, G. E., & Knutson, B. (2009). Saving for the future self: Neural measures of future self-continuity predict temporal discounting. Social cognitive and affective neuroscience, 4(1), 85-92.
- Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. Journal of economic literature, 40(2), 351-401

Thursday March 7th: **The Dead Hand of the Past**: Sunk cost; Counterfactuals; Action and inaction

- “Keeping Score" from Thinking Fast and Slow
- "Sunk cost" from Misbehaving

Tuesday March 12th and Thursday March 14th: Spring Break (No class, No Regrets)

Tuesday March 19th: **I Will have Been Regretting it**: Anticipatory Regret; Action and inaction

- "How things might be different”, Chapter 3 from “The Rational Imagination” by Byrne
- The Simulation Heuristic (Kahneman and Tversky, 1982)

Bonus (Bonus)

- Counterfactual Thought (Byrne, 2016)
- Counterfactual Thinking (Roese, 1997)
- Inaction effect and regret (Zeelenberg, 2002)
- The Psychology of Doing Nothing (Anderson, 2003)
- Regret and Elation following Action and Inaction (Landman, 1987)
- "I Lost on Who Wants to Be a Millionaire and it Almost Destroyed me": http://www.slate.com/articles/arts/culturebox/2015/02/who_wants_to_be_a_millionaire_i_lost_on_the_show_and_it_almost_destroyed.html

Thursday March 21st: **What if…?**: Experienced Regret and Counterfactuals in Decision Making

- (See previous readings)

Intermission

Tuesday March 26th: The Value of (Statistical) Life, Taboo Tradeoffs

- Choice and Consequence, Schelling:
  Chapter 5, "The life you save could be your own"
  Chapter 6, "strategic relationships in dying"

- "The value of Life", Schelling, 1991
- Thaler and Rosen, 1976,
- "What's bad is easy: Taboo values, affect, and cognition" (2007)

ACT III: The Machine Reflects

Thursday March 28th: Odysseus and the Ropes: Commitment devices and meta-preferences

- The Odyssey, "The bit with the Sirens", Wilson translation,
- "Carrots and Sticks", Ian Ayres, and the related https://www.stickk.com/
- - Rogers, T., Milkman, K. L., & Volpp, K. G. (2014). Commitment Devices: Using Initiatives to Change Behavior. Journal of the American Medical Association (JAMA), 311(20), 2065-2066.
- Kavka, G. S. (1983). The toxin puzzle. Analysis, 43(1), 33-36,
- Frankfurt, H. G. (1988). Freedom of the Will and the Concept of a Person. In What is a person? (pp. 127-144). Humana Press.
- Hirschman, A. O. (1984). Against parsimony: Three easy ways of complicating some categories of economic discourse. Bulletin of the American Academy of arts and Sciences, 37(8), 11-28.

Tuesday, April 2nd: "Time for some Game Theory": The Decisions of Others

- Artificial Intelligence, a Modern Approach – section on Game Theory and Mechanism Design
- Camerer, C. F., Ho, T. H., & Chong, J. K. (2015). A psychological approach to strategic thinking in games. Current Opinion in Behavioral Sciences, 3, 157-162.
- Camerer, C. F., & Fehr, E. (2006). When does" economic man" dominate social behavior?. science, 311(5757), 47-52.
- Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). Behavioural game theory: thinking, learning and teaching. In Advances in understanding strategic behaviour (pp. 120-180). Palgrave Macmillan, London.
- https://www.nytimes.com/interactive/2015/08/13/upshot/are-you-smarter-than-other-new-york-times-readers.html

Thursday, April 4th: Everyone is Stupid but Me: deviation from 'optimal' game theory:

- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. Cognition, 113(3), 329-349.
- Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nature Human Behaviour, 1(4), 1-10.
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. Trends in cognitive sciences, 20(8), 589-604.

Tuesday, April 9th: **FREEDOM**! What to Do when you can Do Anything; Action Set Selection; Thought Suppression

- Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in cognitive sciences*, *23*(12), 1026-1040.
- Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of personality and social psychology*, *53*(1), 5
- Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649-4654.
- Sartre, J. P., & Mairet, P. (1960). *Existentialism and humanism* ← specifically the bit with "the case of a pupil of mine…"
- Sartre, J. P. (2001). *Being and nothingness: An essay in phenomenological ontology*. Citadel Press. ← specifically, the account of The Waiter as it relates to bad faith.

Thursday, April 11th: **The Story of Your Life**: Narrative self and Experiencing self

- "Thinking, Fast and Slow", Part V: "Life as a story", Kahneman
- Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. The quarterly journal of economics, 112(2), 375-406.
- Diener, E., Wirtz, D., & Oishi, S. (2001). End effects of rated life quality: The James Dean effect. Psychological science, 12(2), 124-128.

Bonus

- McAdams, D. P. (2001). The psychology of life stories. Review of general psychology, 5(2), 100-122.
- Ullmann-Margalit, E. (1992). Final Ends and Meaningful Lives. Iyyun: The Jerusalem Philosophical Quarterly

Tuesday, April 16th: **Guest Lecture, Prof LA Paul**: Big Decisions, Hard decisions, Transformative experience

Thursday, April 18th: **Decisions, Big and Small**: Transformative Decisions, Picking & Choosing

- Paul, L. A. (2015). What you can't expect when you're expecting. Res Philosophica, 92(2), 149-170.
- Ullmann-Margalit, E. (2006). Big decisions: opting, converting, drifting. Royal Institute of Philosophy Supplements, 58, 157-172.
- Ullmann-Margalit, E., & Morgenbesser, S. (1977). Picking and choosing. Social research, 757-785.
- S.D. Levitt, Heads or Tails: The Impact of a Coin Toss on Major Life Decisions and Subsequent Happiness, National Bureau of Economic Research, 2016, No. w22487.
- McCoy, J. P., & Ullman, T. (2019). Transformative Decisions and Their Discontents. Rivista internazionale di Filosofia e Psicologia, 10(3), 339-345.
- New Yorker piece: https://www.newyorker.com/magazine/2019/01/21/the-art-of-decision-making

Tuesday April 23rd: Closing Points, Presentations/Reflections by Students

Addendum 1: Read/Watch List (Partial)

Some of the references from class for stuff that's not just Decision Making books, contains both 'high brow' and 'low brow'.

1. Lucretius, "De Rerum Natura" ("on the nature of things"). Referenced in DDM discussion. A 2000 year old poem meditating on the physical and mental nature of reality starting from the assumption that everything is atoms and void.

   Consider in particular Lucretius' description of dust motes as a 'model' for how reality works, describing more or less Brownian motion 2 millenia ahead of its time:

   There's a model, you should realise,
   A paradigm of this that's dancing right before your eyes -
   For look well when you let the sun peep in a shuttered room
   Pouring forth the brilliance of its beams into the gloom,
   And you'll see myriads of motes all moving in many ways
   Throughout the void and intermingling in the golden rays
   As if in everlasting struggle, battling in troops,
   Ceaselessly separating and regathering in groups.
   From this you can imagine all the motions that take place
   Among the atoms that are tossed about in empty space.
   For to a certain extent, it is possible for us to trace
   Greater things from trivial examples, and discern
   In them the train of knowledge. Another reason you should turn
   Your attention to the motes that drift and tumble in the light:
   Such turmoil means that there are secret motions, out of sight,
   That lie concealed in matter. For you'll see the motes careen
   Off course, and then bounce back again, by means of blows unseen,
   Drifting now in this direction, now that, on every side.
   You may be sure this starts with atoms; they are what provide
   The base of this unrest. For atoms are moving on their own,
   Then small formations of them, nearest them in scale, are thrown
   Into agitation by unseen atomic blows,
   And these strike slightly larger clusters, and on and on it goes -
   A movement that begins on the atomic level, by slight
   Degrees ascends until it is perceptible to our sight,
   So that we can behold the dust-motes dancing in the sun,
   Although the blows that move them can't be seen by anyone.

1. Flight of the Conchords: New Zealand comedic folk duo, videos of song available widely online, also turned into a 2-season show.
2. Rick and Morty ("you pass butter" robot), popular animated series, I would recommend not going near the dedicated fan base though (true for most things).
3. R.U.R (Rossum's Universal Robots), 100-year-old play, first to feature mention of 'robot'. Other works of fiction on created beings somewhat similar to us are too numerous to count, but consider: Battlestar Galactica, Frankenstein, Bladerunner.
4. Count of Monte Cristo -- Referenced with regards to 'reference point'. the audiobook version is about 56 hours, and worth it.
5. The Good Place -- Referenced in 'party games'. TV show with philosophy consultant in residence

6. Face/Off -- referenced in Robot Heist. terrible movie from the 90's, still enjoyable
7. War Games (movie) -- Referenced in SDT simulation. Captures cold war paranoia as well as AI paranoia, a lot of it seems quaint now
8. The Odyssey -- Referenced in stopping problems, new translation by Emily Wilson in particular is good.
9. The Bourne Ultimatum -- Referenced in Ultimatum Game, doesn't actually have anything to do with it. The entire Bourne series is a totally fine waste of time. Also referenced in a really good recent paper by Cushman that is related to decision making, 'Rationalization is Rational'
10. Paperclip game -- referenced in class on value alignment. A complete waste of time, and that's part of the point. Hijacks evolutionary reward signals.
11. The Strange Case of Dr Jekyll and Mr Hyde -- relevant for thinking about temporal discounting, and in general the 'battle for the self' later thought of as an internal civil war or marketplace (e.g. by Ainslee or Schelling). Also just classic piece of Victorian writing.
12. Back to the Future - mentioned as a throwaway in temporal discounting. Surprised to hear how many haven't seen it! Fun 80's movie.
13. "Pitch meetings" -- a throw-away, Ryan George is the guy going "whoops! Whoopsie!" when we talked about probability weighting. Fun youtube channel.
14. Big Lebowski - a throw-away, mentioned when talking about 'sunk cost' (she lost her toe!). Cult classic.
15. Outbreak -- classic 90's movie about an Ebola outbreak, not really relevant for class though does involve some high-stakes decisions during a pandemic.
16. 10 things I hate about you -- classic 90's film
17. Taming of the shrew -- classic-er Shakespeare play
18. St Augustine and the Pears: References in multiple selves. Book II of the "Confessions" by St Augustine describes an incident in which he and his friends as teenagers stole pears. It's an amazing moment of self-reflection.
19. Allie Brosh, Hyperbole and a Half (also, a followup book: Solutions and other Problems) -- referenced in multiple selves and commitment devices, interesting comics and personal anecdotes.
20. Dr Strangelove -- referenced in commitment devices and game theory, old movie spoofing various Cold War paranoias
21. Death's End -- book 3 of "the three body problem", recent hit of sci-fi, referenced in commitment devices and game theory.
22. "On Bullshit" short book by Harry Frankfurt, the philosopher who wrote on meta-preferences. Not directly related to the class but relevant for everyday life, very readable. The gist of the argument is that bullshit is in some sense worse than lying. The liar knows what the truth is, the bullshitter doesn't care one way or another.
23. The Princess Bride -- classic book/movie (where 'the battle of wits' is taken from), probably hits a bit different when you're 9-12 as opposed to 20-30 years old, but still, very good.

Addendum 2: "Serious" bookshelf (partial)

Some books for edification mentioned in this class or relevant for it:

- Algorithms to Live By -- Great book on CS applications to daily life, particularly the explore-exploit chapter relates to questions that came up in class but which we will not have time to explore.
- Thinking Fast and Slow -- We'll be reading a bunch of this, but not all. It's all worth a read.
- Nudge -- All about 'choice architecture' and 'liberal paternalism'. There's an updated and final version now.
- Transformative Experience -- Related more to the end of the class and 'big decisions', more philosophy than psychology.
- Human Compatible -- Recent book on building AI that doesn't kill us all.

- The Alignment Problem -- Even *more* recent book about building AI that doesn't kill us all, also somewhat of a recent-history perspective of the efforts underway to do this.
- The Drunkard's Walk -- related to DDM, but applied to the world itself rather than to internal psychology. Does touch on psychology in a roundabout way -- the fact that we have a hard time understanding the fact that the world works this way, and so see it as over-determined rather than taking into account the role of luck.
- Carrots and Sticks - Commitments and contracts, pop sci.
- How to Decide - pop sci, a book-length treatment of how to intuitively apply maximizing expected utility in daily life, includes tips, tricks, and hacks, and the focus on optimizing the decision making process itself rather than the outcome.
- Misbehaving - Fun book by Thaler, popsci treatment of his academic ideas.
- Choices, Values, and Frames -- Academic book, classic papers.
- Being and Nothingness -- relates to big decisions and action set construction, philosophical, heavy going, not directly on JDM as usually understood.
- Breakdown of Will -- Classic treatment of future self vs current self and hyperbolic discounting, academic.
- Decision Making (Janis) -- classic study of DM, especially relevant for group-think, and why making decisions is agonizing.
- Choice and Consequence -- Everyone should have more Schelling in their lives. More game-theory/philosophy than empirical study, well done.
- The Construction of Preference -- academic book, classic papers on more recent DM topics.
- Normal Rationality -- less known book, collection of papers/essays.