AAAI 2018 Notes
New Orleans, LA

David Abel*
david_abel@brown.edu

February 2018

Contents

---

*https://cs.brown.edu/~dabel

This document contains notes I took during the events I managed to make it to at AAAI in New Orleans, LA, including sessions of the new AAAI/ACM conference on AI, Ethics & Society (AIES), IAAI, and EAAI. Please feel free to distribute it and shoot me an email at `david_abel@brown.edu` if you find any typos or other items that need correcting.

1   Friday February 2nd

I arrived to the conference in the afternoon, starting with the Workshop session on Bandits.

1.1   Tutorial: Bandits

The tutorial is led by Tor Lattimore and Csaba Szepesvari. For more on bandits, I highly recommend Alex Slivkins' new book draft on bandits[1].

Definition 1 (k-Armed Bandit): *A k-armed bandit is a learning problem setting defined by k fixed but unknown probabilistic reward functions, $R_1, \ldots, R_k$, where the learning algorithm with its environment by repeating the following two steps for n rounds:*

1. *Choose an arm $i \in [1:k]$.*

2. *Receive reward $r \sim R_i$.*

The goal is (usually) to maximize the long term sum of rewards.

There are loads of variants which we'll go over during the tutorial.
Overview:

- Bandit History. What are bandits? Why should we care?

- Finite-armed stochastic bandits. Dave: I changed sessions after this bit.

- Adversarial bandits.

- Contextual and linear bandits.

- Summary and discussion.

1.1.1   History

First some history. The first bandit algorithm was proposed by Thompson [108], where Thompson thinks about clinical trials. He didn't however use the term "bandits". The term actually came from Bush and Mosteller [15], who studied mice that solved bandit-like problems, and eventually built a proper two armed bandit and tested it with graduate students.

---

[1] `slivkins.com/work/MAB-book.pdf`

1.1.2 Why Care About Bandits?

Three reasons:

1. Loads of applications: *Clinical trials, dose discovery, advert placement, network routing, game playing, resource allocation, recommendation systems.*

2. Isolate an important component of the full reinforcement learning problem: *exploration vs. exploitation.*

3. Rich and beautiful mathematically.

1.1.3 What Makes a Bandit Problem?

As one might expect, folks have made loads of distributional assumptions. That is each R might come from a particular distributional family, such as Gaussian. More constraints on the bandit's underlying distribution usually leads to better performance.

Q: How do we tell if our problem is a bandit problem?

A: Well, three properties that isolate a bandit problem:

1. Sequentially taking **actions** of **unknown** quality.

2. The **feedback** provides information about quality of chosen action.

3. There is no **state**.

Note: there is no state in the sense that the underlying reward distributions don't change over time, as a function of something about the environment changing. If you're a Bayesian, how much information you've gathered can of course be treated as a state (that is, which round you're at of the n total rounds).

Q: How do we measure performance of a bandit algorithm?

A: Regret!

Definition 2 (Regret): *Let μ_i be the mean reward of distribution R_i. Then $\mu^* = \max_i \mu_i$. The regret after n rounds is given by:*

$$\rho_n = n\mu^* - \mathbb{E}\left[\sum_{i=1}^{m} r_i\right], \tag{1}$$

where r_i is the reward received by the agent at round i.

Note that policies for which the regret is sublinear are learning. Further, note that the goal of minimizing regret is equivalent to maximizing cumulative reward received.

Sometimes we might care about *finding the best action* as opposed to regret. We call this the simple regret:

> **Definition 3** (Simple Regret): *The simple regret is just the regret you pay on the last round:*
>
> $$\rho_n^{simple} = \mathbb{E}[\rho_n]. \tag{2}$$

The simple regret is sometimes called "pure exploration". It exists in the literature, but we wont focus on it here.

Note that the regret does *not* capture variance. The regret is just a number, but fails to capture the risk of an algorithm. There is a literature focused on risk averse bandits.

Lemma 1.1. *We can decompose the regret into the suboptimality gap, $\Delta_i = \mu^* - \mu_i$, for the i-th arm over all n rounds:*

$$\rho_n = \sum_{i=1}^{K} \Delta_i \mathbb{E}[T_i(n)], \tag{3}$$

where $T_i(n)$ is the number of times arm i is chosen over each of the n rounds.

This lemma is useful, as we can understand an algorithms regret in terms of minimizing the number of times a bad arm is chosen.

Consider the following simple algorithm which we'll call "explore-then-commit":

1. Choose each action m times.

2. Find the empirically best action:

$$\hat{\mu}_i^n = \frac{1}{n} \sum_{i=1}^{n} r_i \tag{4}$$

3. Choose $i^* = \max_i \hat{\mu}_i^n$

To analyze this algorithm, we have to understand the prboability that $\hat{\mu}_i^n$ deviates from μ^*. We can leverage our concentration inequalities, which we get from the law(s) of large numbers (chebyshev's, markov's, and so on).

A common assumption is that of *sub-gaussian*. That is:

> **Definition 4** (Subgaussian R.V.): *A random variable is σ-subgaussan if for all $\lambda \in \mathbb{R}$:*
>
> $$M_Z(\lambda) = \mathbb{E}[\exp(\lambda Z)] \le \exp(\lambda^2 \sigma^2 / 2), \tag{5}$$
>
> *where $M_Z(\lambda)$ is the moment generating function.*

Using this assumption, we can bound the deviation of the empirical mean from the true mean:

Theorem 1.2. *If $Z_1, \ldots, Z_n$ are independent and σ-subgaussian, then:*

$$\Pr\left\{\hat{\mu} \geq \sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}}\right\} \leq \delta \tag{6}$$

The proof leverages Chernoff's method.

Q: Why not use Chebyshev's? Well, the dependence on δ is much worse. We'd get:

$$\sqrt{\frac{\sigma^2}{n\delta}}, \tag{7}$$

whereas in the sub-gaussian case we get no δ in the denominator.

So how can we use this to analyse our explore-then-commit algorithm? For simplicity let's assume that $k = 2$. Also, let's assume that μ_1 is the good arm. What is the regret of our explore-then-commit?

Well:

- Let $\hat{\mu}_i$ be the aberage reward after exploring. The algorithm commits to the wrong arm if:

$$\hat{\mu}_2 \geq \hat{\mu}_1 \equiv \hat{\mu}_2 - \mu_2 + \mu_1 - \hat{\mu}_1 \geq \Delta \tag{8}$$

  Critically, we note that $\hat{\mu}_2 - \mu_2 + \mu_1 - \hat{\mu}_1$ is $\sqrt{2/m}$-subgaussian.

- So, leveraging this fact, the regret is:

$$\rho \leq m\Delta + n\Delta \exp\left(-\frac{m\Delta^2}{4}\right) \tag{9}$$

Neat! Something to think about is how the regret changes as m changes. The two above terms in the regret effectively include our exploration and exploitation terms.

Now off to the AI, Ethics and Society conference for discussions on AI & governance.

1.2 AAAI/ACM Conference on AI, Ethics & Society: AI & Governance

This session is on how to regulate AI.

1.2.1 Regulating AI: Proposal for a Global Solution [80]

The speaker is Olivia J. Erdélyi, joint work with Judy Goldsmith.

Q: Why and how can we regulate AI?

- **Why**: AI is increasingly vital in human society. Regulation is essential to ensure AI is beneficial.

- **How**: We need a *global* solution:

  - AI's global impact targets the nation states' interdependence.
  - Domestic policies not suitable for transnational problems.

  Conclusion: Let's start an International AI regulatory framework. Called: International AI organization (IAIO).

Recipe for transnational regulation:

- Respect the rules of transnational legal ordering.

- Transnational Legal Order (TLO): The collections of norms and institutions that determine practice of law across national jurisdictions.

- Crucial: no one-size-fits-all solution. Governance choices always involve trade-offs.

- Two components of the TLO:

  1. Governance design component
  2. Behavioral impact component

Preliminary conclusions for IAIO's configuration:

- Informal intergovernmental organization

- Interim objective: support/coordinate national AI policies, avoid fragmentation, alleviate tensions, avoid arms race.

- Composition:

  - Membership depending on institutional formality
  - Interdisciplinary expertise
  - Diverse stakeholders

Basically: let's set up an international, informal, regulatory body.

1.2.2 An Agile Model for the International Regulatory Body on AI [1]

The speakers are Wendell Wallach and Gary Marchant.

AI's development is *unique* relative to everything else we've seen before.

Note: "Hard Law" or traditional regulation is going to have very limited effect at regulating AI due to the unique nature of the technology. Instead, we should target "Soft Law".

Their idea: coordinating technology governance committee. Acting like an "orchestra conductor" to provide tools and subcommittees outside of any particular government. Emphasis on modularity and flexibility.

Some practical issues: how do we fund it? how do we structure it?

Their main point is that the committee should be developed organically by the participants. They propose two pilot projects: (1) AI and Robotics, (2) Synthetic Biology.

Major shift: the government *cannot* stay on top of the regulation of all of these technologies. Instead, we should encourage governments to enforce soft laws.

They propose a committee `BGI4AI.org` with the agenda of targeting bias and transparaency standards. Lots of other parties jumping into this space: UN, WEF, ICRC, EU, IEEE, ISO, OpenAI, AINow.

1.2.3   Regulating for Normal AI Accidents [69]

The speaker is Matthijs Maas. Trying to understand the failure modes of AI regulation.

Loads of issues with regulating AI. AI is complex and opaque, systems are tightly coupled and fast. AI designers, trainers, and operators have objectives beyond safety. Competitive pressures exacerbate risks (arms race, fog of war, flash of war).

A few points/criteria that summarize their work:

1. Accidents cannot be designed out: they're inevitable.

2. Claim: Regulation cannot *solely* trust either automated fail-safes or human-in-the-loop.

3. Regulation should focus on levers affecting accident risk: AI complexity, opacity, and so on.

1.2.4   (Best AIES-Co-Paper) An AI Race: Rhetoric and Risks [18]

The speakers are Stephen Cave and Sean O Heigeartaigh. This is the co-winner of the Best Paper Award (for the AIES conference).

Achieving global leadership in AI is starting to be a priority for both nations and companies. We're already seeing a narrative around an AI arms race with respect to technological superiority and autonomous weapons. These risks include:

1. Risks of rhetoric alone:

    (a) Hinders necessary broad debate and consultation
    (b) Can spark an actual race

2. Risks of rhetoric *and* a race

    (a) Corner-cutting on safety
    (b) Increases risk of real conflict

3. Risks of a race being won

    (a) Concentration of power

    (b) Risks of abuse and conflict.

Q: What can we meaningfully do about this?

A: Let's look at alternative narratives: AI for global good, cooperation on safety-critical systems, responsible development for building public trust. That is: let's democratize the power and benefits of AI. Then, it matters less who gets there first.

Finally: What should we in the research community do?

A: There might be an actual information hazard associated with talking about the power of AI. If we distribute the knowledge to use AI. Thus: the role of the research community is to promote positive narratives directly. Lobby policymakers for positive narratives.

1.2.5 Q & A: AI and Governance

They're now holding a quick Q&A with the speakers.

Q: Are there aspects of AI regulation that call for domain specific oversight?

- Wendell – Comprehensive oversight can anticipate developments across sub-disciplines. Sure, we need domain specific regulation, too, but because we're dealing with a technology where breakthroughs have such far reaching impact. So, we need both comphrensive and domain specific regulation.

- Sean – I totally agree!

- Gary – Yeah, there are legitimately too many areas to cover.

Q: Regulation is important. But: lots of questions we need to answer like, why is AI any different from other tech?

- Gary – The only thing worse than too quick of regulation is too slow of regulation.

- Stephen – People are looking at the landscape of existing regulation and see what of it can already address AI.

Q: If we're in soft law, you can't enforce certain aspects, how can we motivate people to cooperate with the regulation?

- Gary – The flaw in soft law is that it's hard to enforce. So, this is a major challenge for us to focus on. We're working on it. For instance, we can change the scientific narrative like folks have done with nanotech and stem cells. Already happening in liability insurance. Trying to build a set of tools to deal with the problem.

- Olivia – Enforcability is an illusion. People will form a co-allition against whatever enforcements exist. We're looking for consensus, which is why we're all focusing on reaching consensus.

Q: More of a comment. There are different orgs whose focus are on regulation. I'm concerned about this diffusing regulation. Suppose you're outside academia and want to find an AI expert.

- Matthijs – One disadavantage of globalized regulation early, is that all of our eggs are in one basket. If we mess it up, we lose legitimacy and poisons the well for future endeavors.

- Olivia – Constant dialogue and information exchange between various stakeholders is key. So, let's encourage everyone to chat. We can learn from the financial regulatory conversations a bit, but the risks are even more dramatic here.

Q: The first slide by Olivia said: "regulation is necessary to ensure AI is beneficial". That's really been a theme. In the Gary/Wendell talk, it's difficult for regulatory agencies to keep up with innovation. The Q: is it really the case that regulation is necessary now? Or is it more that some idealized form of regulation is necessary sometime in the future? How much might the regulatory framework harm the growth of beneficial AI?

- Wendell – Well, we really need a good faith coordinating body like the UN that exposes the rest of the world to AI. Lots of countries have no idea what to do. The stealth aspect that's going on in AI can be damaging. Whether it's regulation or standards we have lots of folks talking about transparency or bias. But we don't have standards for what kinds of systems can industry or even individual actors deploy without it being of serious concerns, and what kinds of systems can they *not* deploy. These standards we need now. It's already on the table today.

- Stephen – Evidence driven policy "would be a wonderful idea" (referred to Ghandi talking about the western world), but sadly it doesn't happen all that often. There's a mix with AI that it's both similar to previous technological revolutions but also new.

....

2 Saturday February 3rd

Today I'll mostly be spending time at the AIES conference, there are loads of interesting talks lined up on AI and law, automation, and philosophy.

Dave: I also want to note that AIES is doing something super unique; they have a wonderful artist (Michelle S Royal from RIDG[2]) in the room each session drawing phrases and imagery that captures the sentiments of the talks and discussion. They're all going onto large poster boards that are up outside next to the posters being displayed. Really nice idea! Here are a few:

Figure 1: AIES hand drawn posters summarizing events of the session, done by Michelle S Royal from RIDG.

2.1 Richard B. Freeman on The Great AI Jobs Scare

Overview of the talk:

1. Futurist vs. Traditionalist Views of Robolution

2. Specter of AI/Robots Changing Comparative Advantage and Raising Inequality

3. Assessing/Resolving Debate: Needed evidence

4. A policy, a Policy, My kingdom for a Policy

Following his talk, there will be some back and forth with Jason Furman, the session chair.

2.1.1 Futurist vs. Traditionalist Views

In the 1960's there was great fear of automation. Herb Simon concluded his analysis of the 1960s automation scare, declaring that: "The bogeyman of automation consumes worrying capacity that should be saved for real problems... The world's problem in *in this generation and the next* are problems of scarcity, not of intolerable abundance."

---

[2]http://ridg.com/
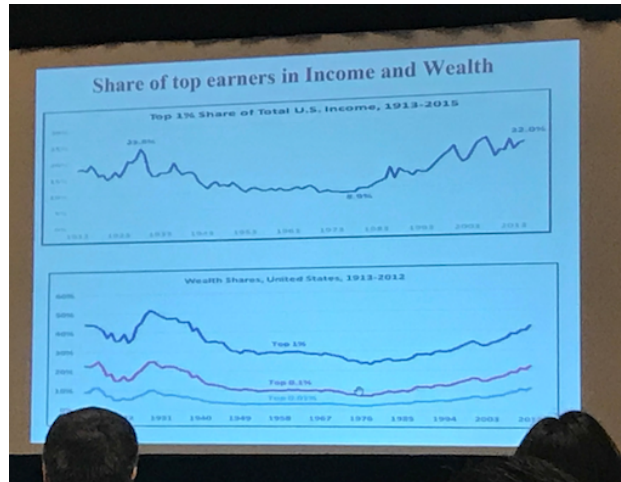
Figure 2: Rising Wealth Inequality

Through the 1970's, wages rising with productivity, stabilizing labor's share of income, pay inequality stable/falling, pension funds increasing.

But: Today's economy is different. Wages stagnate even as productivity grows: labor's share of income has fallen while top executives pay is linked to capital income. Pay inequality has risen. Capital/wealth inequality is at 1920s levels. AI competes with humans at cognitively demanding jobs.

Q: Will the AI technologies of the "4th" industrial revolution exarcebate the trend in inequality or create intolerable abundance for all?

Some answers:

- **Futurist**: Project massive job losses. AI experts predict that machines will have *absolute* advantages over humans in many jobs in around 50 years and attain full automation of labor in 125 years.

- **Traditionalists**: see the modest growth of productivity and tight job market as a sign that the new technologies are unlikely to disrupt the economy and remind us that all previous automation scares fizzled out

Comparative advantage suggests that rather than destroying jobs, the new technologies will alter the division of labor between humans and machines and the distribution of income from work. Updating Simon: **the world's problems when AI/robots do the work and earn the income will be problems of distribution**.

2.1.2 Specter of AI

Let's think about robot lawyers (robot and AI are conflated here).

Article from Rory Cellan-Jones: "The Robot Lawyers are here - and they're Winning". Basically: some AI's out-predicted some lawyers on the outcome of a particular case.

Also discussed robot baristas. But claimed: there's a big difference between an experimental bot doing it once and getting rid of baristas.

His Take: When we reach the place where robots *do* takeover, what do we do? The concern: "those who own the robots **rule the world**".

Traditionalist Response: You see AI robots in the headlines, but not in the productivity or job statistics! Same with computers. Productivity growth in the 2010s is lower than in the last five decades. E/Pop is high, unemployment is low. Rising inequality began before AI as a result of measured factors: fall of unions, trade immigration. Dave: Wasn't clear if it was "fall of trade, fall of immigration", or "trade, immigration, and fall of unions" (my guess is the former)

But: It's actually really *hard* to measure productivity. The nature of productivity changes. If workers are now working more hours and taking longer commutes, it's different from walking into a building getting clocked and walking out.

Q: Why should this time be different?

- Past fears that automation destroys jobs fizzled. FDR blamed the Great Depression joblessness on failure to "employ the surplus of our labor which the efficiency of our industrial processes has created". US Commission on Automation, Rifkin's End of Work (1995).

- Past periods of change produced rapid productivity.

- However technology changes, expansionary macro-policies

2.1.3   AI Progress Evidence

Time-line of AI Winning Games:

- 1997: Deep Blue [17] beats Kasparov at Chess

- 2011: Watson wins Jeopardy

- 2016: AlphaGo [99] beat Lee SeDol at Go

- 2016: Libratus [14] beats top poker champions

- 2016: DeepStack [77] wins no limit texas hold'em poker tournament

- 2017: AlphaZero [100] learns chess and beats chess programs in a weekend, triumphs over Go, Chess, and Shogi. Dave: Other folks have discussed qualifications to these claims, namely, that the Chess program stockfish was a slightly modified version (it didn't include it's library of starting moves, among other constraints). Still an extremely impressive feat, of course.
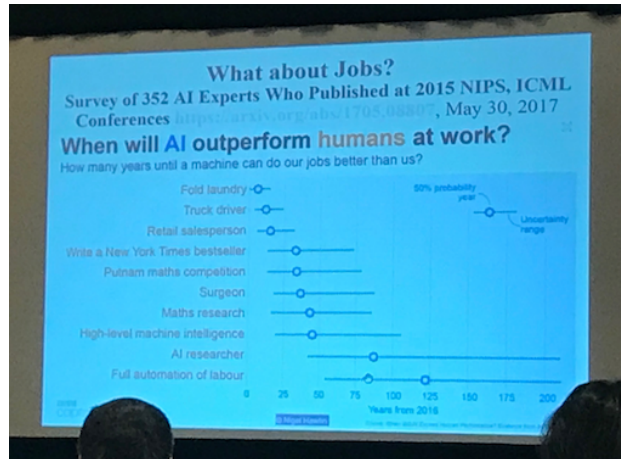
Figure 3: Expert Predictions of AI Dominance from Grace et al. [39].

So: surely *something* is going on. We can't just dismiss these trends, even if it doesn't yet show up in the economic statistics.

Really nice paper surveying AI experts projecting when AI will achieve human level performance on various tasks by Grace et al. [39]. He showed some highlights:

So: machines have gained comparative advantage in physical, repetitive work, as John Henry learned in 1970s.

Q: Have machines gained comparative advantage in cognitive tasks? Quote from Eleni Vasilaki from Sheffield University: "AI fails in tasks that are surprisingly easy for us, from running, walking, and kicking a ball.

2.1.4 Assessing the Debate

Two conflicting pieces of evidence:

1. The economic statistics don't show an impact from AI. So we shouldn't worry!

2. AI is taking over in lots of areas, from physical repetitive tasks, to the cognitive tasks.

Quote from "America's greatest philosopher" Dave: ?: "What, me, worry? About work! You are MAD. I have funner things to do. If this time is different, it's the baby's problem, not mine."

Three Laws of Roboeconomics:

1. AI + improved computers + robots $\rightarrow$ better machine substitutes, higher elasticity of human substitution.

2. Tech changes reduces cost of robot substitutes for human labor, leading to a bound in wages.

3. Effect on income depends on who owns robot. So, again: **who owns the robots rules the world**.

Point: Technological feasibility does *not* mean quick adoption. Costs must be lower and organizations must adjust.

Federation of Robots released a report on robots and automation, available here.[3]. Other folks have done similar things.

Point: We really need a database of AI/robot advances by occupation and skill. They're putting this together.

2.1.5 Policies

This automation wave is *inevitable*. It's purely a matter of what year it takes place. There's nothing in the data to suggest that income inequality will reverse. Who owns the robots will make the money.

Some ideas:

- Bill Gates: a robot tax! *Concern*: Probably don't want to tax a robotic public servant, like a police officer.

- A global capital tax (global meaning all countries)

- Social democrats: Universal Basic Income.

- IR traditionalists: laws to strengthen unions, minimum wages

- Populist demagogues: attack trade, immigrants

- Elon Musk: put computer implants in our brains to raise human productivity.

- Give workers ownership of their firms.

  Concern: Not enough. Need ownership stake beyond the employer. Take a look at Peter Drucker's Pension Fund Socialism [27].

Q: Why do we think differently about our policies for climate change and policies for AI and automation?

Conclusion: Final answer to Herb Simon. When we have intolerable abundance, we have other problems.

2.1.6 Rebuttal from Jason

The Traditionalist Argument:

- Robots/AI will replace *skills* or individual tasks not jobs.

- We'll be richer from robo-production so we'll buy more stuff.

- People will compete with robots so wages will adjust.

---

[3]https://ifr.org/ifr-press-releases/news/world-robotics-report-2016

- In the 1920s someone could have said 90% of jobs will be replaced by automation, and they *would have been right.* However, new jobs were created.

- Recent data suggests were seeing the exact opposite of the AI automation concerns.

What Jason is worried about: robots will affect wages by introducing new competition. If the process happens too quickly, the workforce participation could decrease rapidly. It's a high class problem we don't yet have, we'd love to see more productivity. If the pie is bigger, there are more ways to solve the problem.

This will all result in more inequality and less participation, not a lack of jobs. So, we should emphasize focus on making work more rewarding and increase participation.

2.1.7 Q & A with Richard and Jason

Q: Another possibility: longer term we democratize the technology so everyone has access to AI and 3d printers. In this world, everyone can make what they need for themselves. If there's a job they don't want to do, they can just have a robot do it.

- Jason – I'm thinking of the next 50 years.

- Richard – Sure, it's intolerable abundance.

  Reponse from Question Asker: Don't call it intolerable. Also, if it is 50 years away, what should we do now.

  Someone else chimed in: At ICRA, someone asked the same question. The last job an AI will take is a hairdresser! Lots of dexterity, requires trust. Follow up: why would robots get wages?

- Richard – Everyone will be rich, but they don't need the money due to abundance. So they can do whatever they want.

Dave: They again decided to allow several questions to be asked before answering, so things will be a bit out of order.

Q: Implicit in these arguments is that the goal is to get the work done at the lowest cost possible. But, isn't it rather that the goal is to maintain a good society? I would argue that we can create that 90% of new jobs, but we can't rely on the current mechanisms on creating jobs because our goal is wrong.

- No one answered this question. I loved this question, though (from Ben Kiupers from U. Michigan).

Q: Expert systems in health care like IBM's Dr. Watson – will these machines produce new knowledge? If not, couldn't there be a pandemic if no people do anything?

Dave: I can answer this one: AI can produce new knowledge, so no worries on that one! (Richard said the same.)

2.2 AIES Session on Ethical Issues and Models

Now some spotlight talks from submitted papers.

2.2.1 Jill Watson Doesn't Care if You're Pregnant [13]

The speaker is Bobbie Eicher, joint work with Lalith Polepeddi and Ashok Goel.

Jill Watson was a chatbot TA made at Georgia Tech for a grad level AI class.

The students didn't actually know that Jill was an AI. They've continued running Jill under different name and obscuring which of the TAs is the AI.

Some observations:

1. It's important to consider the ways that an AI may be biased by the demographics of the people it is basing its data on.

2. Deception is *sometimes acceptable*, such as to ensure that interactions are as natural as possible or provide an improved experience, but must be handled with care (like not telling the students which TA is Jill).

3. The ethical implications of a decision may vary greatly depending on the population of users an agent needs to serve.

4. AI designers must take care to be aware of their competing ethical obligations, such as theirs when they act as both educators as researchers.

2.2.2 Ethical Challenges in Data-Driven Dialogue Systems [83]

The speaker is Peter Henderson, joint work with Koustuv Sinha, Nicolas Angelard-Gontier, Nan Rosemary Ke, Genevieve Fried, Ryan Lowe and Joelle Pineau.

Goal: What are the problems of current dialogue systems? What are the risk factors? How can we improve them?

A dialogue system is a chatbot, virtual personal assistant, natural language interface for an y system. We should care about them because they're prevalent in the world.

Q: So, what are the ethical challenges?

A: Bias, privacy, safety, adversarial examples, reproducibility.

Bias: prejudice for or against a person/group/idea/thing particularly expressed in an unfair way.

Problem: End-to-end dialogue systems are trained on data sets that include biases, even if the biases are subtle. They address this by using a word-vector debiasing method.

Safety: Lots of critical application areas (diabets diagnosis, mental health). Goal should be to provide:

- Performance guarantees, stability

- Proper objective specification

- Model interpretability

Reproducibility: Ability of researchers to duplicate results of a prior study.

2.2.3 The Dark Side of Ethical Robots [109]

The speaker is Dieter Vanderelst, joint work with Alan Winfield.

First: robots can be dangerous. Most promising applications are cases where they have lots of interactions with humans, and are relied on to be semi-autonomous. Once they interact with people, though, they need to be designed ethically and behave ethically.

They propose a thought experiment where two people interact and play the "ball under the three cups game" with a robot watching. The ethical dilemma is whether the robot should intervene and let the human know that the person made a mistake, otherwise they should do nothing.

Then, they change the game to a competitive variant: the robot plays the game, too, but is faster than the person (usually).

Point: Requires increased knowledge about the human to delineate desirable from non-desirable behavior. But: this knowledge can be abused, too. In the competitive robot game, they robot uses its knowledge that the human trusts the robot's suggestion to mislead the human player so the robot can win the game.

Conclusion: Engineering is only part of the solution. Equally important are societal responsibility, government regulatory frameworks.

2.2.4 A Computational Model of Commonsense Moral Decisions [87]

The speaker is Richard Kim, joint work with Max Kleiman-Weiner, Andres Abeliuk, Edmond Awad, Sohan Dsouza, Josh Tenenbaum and Iyad Rahwan.

Focus: how do we design an AI that makes decisions about moral dilemmas? Further, how do humans make moral decisions and what can we learn here?

One thought: well deep learning has done (almost) everything else! So why not moral decision making?

Lots of issues with deep learning and moral dilemmas: lack of data, lack of transparency.
Their contribution:

1. Propose a cognitive model of utilitarian moral decision making

2. Address poverty of stimulus

3. Interpretable and predictable model

Principles underlying their model:

- Abstraction is central to our ability to make value judgments.

- The mind is utilitarian.

- Group norms exist and matter.

Q: How do these three principles play out when making moral decisions? The moral machines dataset [85] offers some data on this front.[4]

Advocate for moral decisions as a utility calculus, with an underlying hierarchy of moral principles. They do some analysis of the moral machines data set, showcasing moral biases of different cultures.

They propose a particular hierarchical Bayesian model of moral decision making. The goal of the model is *validation*. That is: the model is seeking to capture human cognition and decision making correctly, not predict labels/decisions. Dave: I think? It's not entirely clear what the goal of the model is, actually. For more check out the paper! [56].

Future work: tackle models beyond utilitarian moral theory, data beyond the trolley problem.

2.2.5   Q & A

Q: When will Jill Watson be available for me to use in my class?

- Bobbie – Not yet!

Q: To the guys that talked about the limits of ethical robots: What do you think about governance?

- Dieter – Super important, but I'm not an expert in that area so I can't really say.

Q: Ben K. – one of Asimov's laws was that robots shouldn't deceive. You (the Jill Watson folks) mentioned that sometimes deception is important. So what should we do here?

- Bobbie – That makes sense. Our case was particular: we had the students meet up and chat about their interactions with the TAs. We thought it enhanced the students' experience. Could go either way.

Q: Right now lots of chat bots are trained off of social media. Why not use a data set from something else like kids fantasy stories?

- Bobbie – What we're doing in a class is almost like that, because it's a set of things you'd be willing to say to your professor.

---

[4]Check it out here: http://moralmachine.mit.edu/

Q: For Richard. There's another way to think about moral dilemmas, like from analytical ethics. They establish ethical axioms and derive ethical decisions. What are your thoughts on a hybrid analytic inference method with a more Bayesian approach?

- Richard – Lots of opportunities to combine symbolic AI and machine learning.

2.3 AIES Session on Transparency and Social Good

Next session on transparency.

2.3.1 Non-Greedy RL Agents with Diminishing Reward Shaping [31]

Speaker is Shou-De-Lin (I think, the room was loud while he introduced himself), joint with Fan-Yun Shan, joint with Yen-Yu Chang, and Yueh-Hua Wa.

Goal: Is it possible to prevent "starving" RL agents by avoiding greedy decisions?

Key concept: diminishing reward shaping. So, for some true underlying reward function $\mathcal{R}$: $\mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$, we augment by a temporally modified potential-shaping function, ϕ:

$$\mathcal{R}_t(s, a, s') = \phi_t(s, a, s') + \mathcal{R}(s, a, s') \tag{10}$$

Here $\phi_t(s, a, s')$ depends on the counts, $c(s, a, s')$, denoting the number of times a is executed in a and landed in s'. They tried a few types of discounting, including a step function and an exponentially decaying shaping function.

Benefits:

- Non-greediness

- Non-homogenous equality

- No global observability required. Prior work from Lerer and Peysakhovich [62] requires the historic information of reward received over time of all agents.

- Cost-effective

- Universal: applicable to most RL models.

Experiments: (1) Gathering game with multiple agents, and (2) Hunter-prey. All grid games.

They use a DQN on these environments. Dave: Why? Isn't it tabular?. They find the DQN does better than any of their shaping methods. Ah: the results from both agents side by side, with shaping, the agents do similarly well, whereas with the DQN, one agent tends to dominate. So: the shaping can enforce equality among competing agents. Dave: I *think* that's the main selling point on the non-greedy shaping piece.

2.3.2 (Co-Best Paper of AIES) Transparency and Explanation in Deep RL [84]

The speaker is Katia Sycara, joint work with Rahul Iyer, Yuezhang Li, Huao Li, Michael Lewis, Ramitha Sundar.

Goal: want to understand what's going on in Deep RL models. That is: get the model to explain itself, generate visual or natural language explanations for its decisions.

Example: Ms. Pac Man, why did the agent choose to go right? Some explanations would be: avoiding ghosts, eat pellets, get the cherry.

Q: How can we get these explanations visible for humans?

Their method: Object-DRL. Incorporate object characteristics into the image.

Then, instead of doing pixel saliency, *a la* Cheng et al. [19], they can do object level saliency:

- Rank the objects in a state s based on their influence on $Q(s, a)$.

- Non-trivial to design the derivative of $Q(s, a)$ with respect to discrete objects.

- They apply a sensitivity-analysis style approach.

Intuition: Look at relevance of an object on decisions. To do this, they mask over the object and compute $Q(s_{mask}, a)$. That is:

$$w = Q(s, a) - Q(s_{mask}, a). \tag{11}$$

So, if this difference is large, then that object was really important for achieving a high $Q(s, a)$. Using this difference, they generate visuals of important objects.

Conducted an experiment with people in the loop to determine if the visuals were actually helpful.

2.4 AIES Session on Ethics and Policy

I missed the first two talks as I was out to lunch. The next two sessions are on AI & Philosophy, with focuses on Ethics/Policy in the first and Machine Ethics in the second.

2.4.1 Regulating Autonomous Vehicles: A Policy Proposal [68]

The speaker is David Danks, joint with Alex John London.

Focus: what's the right way to regulate autonomous vehicles?
Current regulator systems based on Performance Standards, acceptable tolerances of performance in specifiable contexts.

Problem with autonomous car regulation: We don't necessarily know the relevant causal factors that contribute to transfer and generalization. We don't know what counts of success. What does driving safely even mean?

22

One domain with the same issues: drugs and medical interventions. We have very similar challenges. So, let's conjure up a regulatory body inspired by medical interventions.

Drug regulatory structure: (1) Early-phase testing, (2) Transitional testing, (3) Confirmatory testing. This structure presents opportunity for moving forward at an appropriate rate *and* moving backward and revising when necessary. Some pros and cons:

- **Pros**: Increased trust and adoption rates.

- **Cons**: More invasive, requires dangerous public testing.

2.4.2 AI Policy as Entangled Super Wicked Problems [41]

The speaker is Ross Gruetzemacher.

"Wicked Problems" formalized by Rittel and Webber [88] via 10 characteristics, such as "solutions to wicked problems are not true or false, but good or bad" and, "every wicked problem is unique".

Since 1973, Head [45] and Camillus [16] update this notion with further clarity.

In 2011, we get super wicked problems, featuring global climate change [63]. Four properties:

1. Time is running out

2. No central authority

3. Those seeking to end the problem are creating it

4. Hyperbolic discounting

The only agreed upon super wicked problem is climate, but some suggestions that nuclear warfare is also a super wicked problem.

In this work, the author looks at AI strategy as entangled super wicked problems. Specifically, three classes:

1. Class 1: AGI strategy problem (beyond super wicked). Entangled with the others.

2. Class 2: Narrow AI strategy problems (super wicked):

   - Militarization of AI
   - Economic effects of AI & automation

3. Class 3: Ancillary AI problems (wicked):

   - Automation of healthcare systems
   - Algorithmic accountability

Goal: is there anything in existing literature in how to address previous wicked or super wicked problems that we can draw on to guide strategy for AI?

2.4.3 When Do People Want AI To Make Decisions? [72]

The speaker is Max Kramer, joint work with Jana Schaich Borg, Vincent Conitzer and Walter Sinnott-Armstrong.

Main question: should AI development follow public demand? If so, what decisions *should* AI be making?

They did a study of exactly this latter question, asking for explanation as to *why.*

2.5 AIES Session on Machine Ethics

Now onto machine ethics!

2.5.1 Value Alignment, Fair Play, and the Rights of Service Robots [29]

The speaker is Daniel Estrada.

Q: Turing Test – Can we build a machine that exhibits human like behavior? Later, we get something more like a Moral Turing Test.

New agenda: AI Safety, friendly AI, value alignment. Core issue of this literature: can we guarantee behavioral constraints on (superintelligent) AI?

Now: Turing's Test isn't great, while value alignment is really popular. Main shift doesn't really care about the nature of intelligence, but instead focuses on models of divergent AI or data.

1947 "Fair Play for Machines" Lecture from Turing: "It might be argued that there is a fundamental contradiction in the diea of a machine with intelligence... The machines must be allowed to have contact with human beings in order that it may adapt itself to their standards." This is exactly Turing forecasting about value alignment.

But: it doesn't just say something about what machines need to do (get data about human behavior/preferences), but also that humans need to focus on giving machines good data/preferences.

Q: What makes a good test?

- Accuracy of results.

- No administration.

- No unwarranted biases.

In light of these qualities, two (meta)interpretations of Turing's test:

- **Benchmark**: It's just a benchmark.

- **Fairplay**: Motivated by the fair play lecture; we should think about Turing's Test as advocacy for fair play tests where the benchmark isn't particularly important, but that the human and machine shouldn't be stacked differently.

Q: Does a robot ban violate robot rights? San Francisco just issued a ban on robots on sidewalks. We've also seen plenty of cases of "bullying" robots, where a Boston Dynamics pushes a robot in the stomach with a hockey stick.

2.5.2 Grounding the Moral Status of Intelligent Machines [92]

The speaker is Michael Scheessele.

Scenario 1: An android is drowning in Lake Michigan. Do you help it?

Scenario 2: you're standing on a street corner in your neighborhood. An apparently "intelligent" android is crossing the street. The android is struck in the crosswalk by a self-driving car. Do you help the android?

Two assumptions:

1. An intelligence machine can have more than just **instrumental value**.

2. Moral status comes in degrees.

Moral Status Pyramid (MSP):

- (Lowest level) NM: Existence

- MS: A good of its own.

- SF: Can experience pain/pleasure.

- (Highest Level) F: Entities with moral agency.

Q: Where should we place intelligent machines?

One Answer: How about SF? He says no. This is for persons (which is philosophically questionable ground, as he acknowledges).

A better answer: MS. Machines will have goals and self interests, so MS is the right level of moral status to obtain.

Conclusion: Our moral obligations to current (and future?) intelligent machines will fall short.

2.5.3 Impacts on Trust of Healthcare AI [61]

The speaker is Emily Larosa, joint work with David Danks.

Goal: Investigate at AI's impact doctor-patient trust.

Three kinds of trust: (1) Behavioral, (2) Role based, (3) Understanding.

Three policy recommendations:

1. Doctors using AI systems and results must have training

2. AI use necessitates educated consent from patient/care giving

3. Until AI is "standard of care", must provide a human alternative.

2.5.4   Toward Non-Intuition Based Machine Ethics [48]

The speaker is Tae Wan Kim, joint work with John Hooker.

Problem: Intuition-based model is a major theory in machine ethics, but it's bad.

Here's an example: they call out Anderson/Anderson for advocating for this theory (and showed an email thread of them talking about).

Intuitionism! A & A: "We used ethicists' intuitions to tell us the degree of satisfaction/violation of the assumed duties.

Q: Why use ethicists' intuition?

A: Because common sense says so!

For moral decision making, let's use mathematical common sense/intuition. But, this is a violation of the is/ought naturalistic fallacy. Example: bribery (people do it $\rightarrow$ you should).

A & A attempt to jump the Is/Ought gap with ethicists moral intuition.

Some points on intuitionism:

1. Daniel Dennet: Intuition is great, but just an intuition.

2. Moral philosophers' intuitions are not significantly different from those of ordinary people.

3. Human error is leading cause of accidents. Likewise, human moral intuitions are not reliable, so we need non-intuition based machine ethics.

Their contribution:

- Deontological ethics for machines.

- Deriving ethical principles based on *formal properties*.

- Let's used quantified modal logic.

The Generalization Test: Suppose a robot attends CMU. Its taking an exam and wants to cheat. First checks to see if it is ethically permissible or not.

Suppose he finds: Yes, it is permissible! What might have been the premises?

1. It is permissible for anyone who wants to cheat in this exam to get good jobs later.

2. I want to cheat to get good jobs later.

A few other tests (from Kant, I think?):

- The Autonomy Test
- The Utility Test

Dave: The speaker is hilarious!

2.5.5   Q & A

Q: People who worry about value alignment worry about accidentally destroying humanity. Fact is: we have machines now that are misbehaving like LLCs and governments. Why is it that we're not worried but these "machines" that exist today that are misbehaving already?

- Tae – I agree. I teach MBAs! I don't trust corporations. I don't trust governments.
- Daniel – It's hard to figure out what to do about a government or a corporation. A robot is a good way to simplify the story.

Q: Similar to previous one. Where would you place a corporation on the moral standing pyramid?

- Michael – It's a legal person, not an ethical person, so it doesn't fit into the pyramid at all.

Q: Why deontology? It's your intuition that deontology works, but not mine. So you've punted.

- Tae – Intuitionism is broken. Philosophical and logical argument says X is bad. If I have moral intuitions, they're just mathematical intuitions.

Q: If we have intelligent agents, what if they can be backed up and restored? How does that change their moral status? If something is backed up, what can we take away from it?

- Michael – I think it has to have personhood for it to be punished or have moral status. With a backup, it's not clear how that comes into play.

Comment: Richard Kim from MIT Media Lab – we didn't make moral machine to prescribe a particular code of morality, or use human intuition. Instead we're just trying to study it.

- Tae – Are we cool now?

  Richard: Yeah, we're cool.

Q: Kant is nice, but can we actually compute something like the categorical imperative (over possible world space) or the generalizability test?

- Tae – The best model we can use is an expert system, rather than Kant's system. It's less about how can we, and more about how should we. That's the question we have to figure out together.

Q: There are loads of forms of the categorical imperative. How do you even begin to unpack modal logic on concepts that a computer can't understand?

- Tae – we agree with Kant in that the laws are different. Ultimately we are on board with deontology. Further we don't use traditional modal logic. We've come up with our own modal lexicon and deontology, which we think works. We're not saying the theory is philosophically perfect.

- Huw (the chair) – That's a good place to stop!

2.6 Oxford Style Debate! Has Machine Learning Replaced the Need for Logic

Tonight they're hosting the first "Oxford Style" Debate. The topic is: "Have Advances in Machine Learning Replaced the need for Logic in AI?".

Yes: Tom Dietterich and Bart Selman will be arguing in favor.

No: Gary Marcus and Fracesca Rossi will be arguing against.

The MC is Kevin Leyton-Brown.

Focus: what should our research priorities be? That is, think about the question *forward facing* in terms of what research we should do. They've also said: **the debaters should forcefully take on the side they represent, even if it overrides what they actually think.**

2.6.1 Opening Statements

Bart (for) – What brought logic into AI? Let's go back to John Mccarthy. Mccarthy thought of AI as a large knowledge based system and an inference system. Drawn to logic because of t's simplicity and power (ZFC, Peano arithmetic). In the 80s: lots of knowledge is uncertain, so logic isn't the full story (see: Pearl [82]). To Bart: the difficulty with Mccarthy's view is that building this (logical) knowledge base is extremely difficult, perhaps impossible. Conversely, Deep Learning presents a means by which ths can happen. Think about AlphaZero and chess. When you actually play a game, the game revolves around exceptions to the rules, so the construction of these general rules is infeasible. Very high level of play can come from forming a "knowledge base" via self play with a deep net, *not* logic.

Gary (against) – How many folks saw the headline that "AIs can read better than humans?" Most interesting reading you do is about connecting things across the text. General abstract principles that you can reason about are critical for putting together solutions. Consider a yarn dispenser. It's an upside down bowl with a stick on top with yarn around it. Even if you've never seen on you now understand what it is. That's logic. You don't need to see millions of images of a yarn dispenser. Problems like distinguishing a poodle from a labrador are not why advocate for logic.

Tom (for) – Two things that make logic logic. (1) The use of boolean connectives. (2) The use of logical inference. Both of these things are not going to scale, while deep learning will scale. Lets think about a contract. Excitement about blockchain as a contract. But it's failed, because you

need to specify the contract in logic. And specifying a self-enforcing contract in logic is extremely difficult. Most concepts are *not* decomposable into necessary and sufficient conditions. What we need are uncertainty and partial concepts. Toward (2): we have things like modus ponens. Example: if you rob a bank, then you will have money. So, suppose I rob a bank. Am I rich? Trouble with logic: it loses the content. Reasoning is too contextual for modus ponens to actually make sense. Refinement of the antecedent destroys modus ponens. So what does deep learning offer? Suppose we want to do classical kind of reasoning like taxonomy (all ducks ar birds). In deep learning we have generative models for learning ducks, and we certainly have discriminative methods for classifying ducks. Well, we can just use these two to determine if all ducks our generator makes are classified as birds. So to can the representation and complexity of search and planning be done by deep learning. Gary made the point that we use too much data, well right now maybe, but we'll get better.

Francesa (against) – Logic shouldn't be used for everything, as it can't capture the nuances of every day life. But, we can leverage the benefits of both sides. Both capabilities are needed. Nature decided they're both needed for us. AIs will be interacting with us, so they'll need abstract symbols, they'll need to explain themselves, and they'll need to be interpretable. To do these things logic must be a part of the puzzle. Further, deep learning requires too much data, while logic does not. So, bringing the two together, combined, we get the benefits of both.

2.6.2 Debate

Format: the debaters ask each other questions. If it lulls, the MC will ask.

Q: What would evidence that logic is or is not needed in AI look like?

- Tom – One thing would not count is the claim that we need logic as an intermediate language from humans to machines. Mentioned the paper "AI meets Natural Stupidity" McDermott [73]. If you could show that the *onlt* way to do certain functions requires logic that might do it.

- Gary – When you find a solution to a certain type of problem, you'll find a core in that solution that looks an awful lot like the core of logic (variables, bindings, and so on).

- Bart – Excited about AlphaZero (Dave: Originally said "DeepBlue" :) Bart and the audience got a kick out of that). Knowledge is in the network. Suppose you say: for the next three moves, you can't move a pawn. A human could do this and adapt. DeepBlue could take this into account. A deep net could *not* take this into account.

- Tom – When we look at deep networks and give explanation, we're just coming up with a story. We're not doing neuroscience. We find some very approximate explanation for certain situations. The method is fundamentally ineffective, though. So: Gary, why do you think it's not possible to explain this in terms of variables and so on?

- Gary – I'm saying in some models the rules of the variables are present. In many cases there's a transparent mapping to these sorts of things. Some cases that are explicit cases that have logical machinery. Sure, there might be others that aren't like this.

- Tom – typical logical move to say "If there exists even one case" then *au contraire.*

- Francesa – Logic is definitely needed for some tasks. If we find some tasks where abstractions and symbols are needed, or logic is needed to yield interpretable and explainable AI, then we're done. That's the evidence we're looking for.

Q: Some of the discussion has hinted that Machine Learning and Logic are at odds, or perhaps logic is something we can sprinkle in. So: what do you think is essential to ML and Logic? Can they cooperate?

- Gary – There can be displacement. AlphaZero could mostly be done without rules, so perhaps this is a case where rules have been displaced (moved from logic/planning/search to learning). Mentioned the "past tense case" as a motivating example. Some domains you can get rid of the logical stuff, but some you can't. The argument is: there are places where logic is necessary.

- Bart – When I think of logic, what i think of are what properties it can add that we don't get from deep nets. The compositional nature of logic is not found in deep nets. That is: what capabilities does the language give you? What might happen: deep learning folks might find that logic and compositionality are crucial. The approaches are complementary.

- Gary – I like your example of chess. Consider AlphaZero playing on a bigger board with no retraining. A human could do it, AlphaZero could not. You don't get the idea that "rooks can slide all the way across the board."

- Bart – But it knows extremely well, and with certain kinds of generalization it does get it (everything on that board, for instance). It's just not a generalization across tasks of that kind.

- Francesa – Reluctant to take one specific task and say for sure you will need logic. In general, though, I don't see how you can get rid of abstractions and inference. These systems don't need to interact with other people and act in the real world, so we don't need these kinds of things.

- Tom – But we mostly reason with symbols on paper. Why did we invent logic? Because it's too hard to do in our head. Seduced the founders of AI to thinking we can do everything that way. Instead, a good model of human inference is lots of context and quick snappy decisions.

- Gary – I think peoples' representation of the world is a variable binding. Moving a rook along a row is a variable binding. You raise an interesting question though: why can't we do formal logic that well? Well: the binding is expensive. We have lots of makeshifts around it.

2.6.3   Closing Statements

Summary by Kevin: AI might use logic, but it will need paper to figure it out. Closing statements are supposed to mention something they like about the other side.

Francesa – I like Kevin's summary! There will be tasks/domains where deep learning will not solve the whole task. For that reason, logic will still needed. Deep learning showed us great potential

when there's lots of data available.

Tom – Let's talk causality. ML people need to think more about Causality, and think more about Judea Pearl's work. Causal transportability is super powerful (we can think about the center of a blackhole). Causality is also story telling. We posit that there are variables that we manipulate, but they're fiction. Still a powerful fiction. More general and transferable than current ML. So: where should we put our research investment? To build systems we can be confident will work and we can interact with, we need ways of constraining and testing them with background knowledge.

Gary – I love Bart's point about AlphaZero and knowledge. Problems, though: (1) MCTS is already logic/rules. (2) It only works on a limited set of things, not the open world. That's the fundamental problem with deep learning: they work on closed world problems. Other kinds of problems might need the logic.

Bart – As has been pointed out the generalizability of human cognition is amazing. Symbols can be useful for that. Deep Learning work, though exciting, and surprising, has not paid attention to the limitations of their systems. That's a challenge for us: bring out these issues. How *can't* it generalize? Dave: The "Frame Problem" of deep learning, sort of. Might be that after this we need symbols after all.

....

3 Sunday February 4th

I arrived just in time for Tom Dietterich to give the Blue Sky Awards:

1. First Prize: Engineering Pro-Sociality with Autonomous Agents by Ana Paiva, Fernando Santos and Francisco Satnos.

2. Second Prize: Learning Fast and Slow: Levels of Learning in General Autonomous Agents by John Laird and Shiwali Mohan.

3. Third Prize: Imagination Machines: A New Challenge for Artificial Intelligence by Shridhar Mohadevan.

Next up: the presidential address.

3.1 The Presidential Address: Challenges of Human-Aware AI Systems

The president of AAAI is Subbarao Kambhampati.

Some quotes/quips on AI:

- AI is the new electricity (Ng)

- AI is bigger than Fire and Electricity (Pichai)

- AI is GOD (Levandowski)

- "I hope the smart tooth brush's values are aligned with my teeth". (Subbarao)

- AI is bigger threat than North Korea... AI will start the third world war (Musk)

- AI is highly like to end human civilization (Musk)

- AI could be the worst even in human history (Hawking)

- (Fake) News Article: "AI helps old lady cross street."

What Subbarao wants to talk about: Why is this last item fake news? Why isn't human-aware AI (HAAI) and human-in-the-loop AI there, yet?

Four pieces to the talk:

1. Why isn't HAAI all over the place already?

2. Why we should pursue it (Hint: broad scope and promise of AI)

3. Research challenges in HAAI

4. Long term issues: Trust, ethical dilemmas.
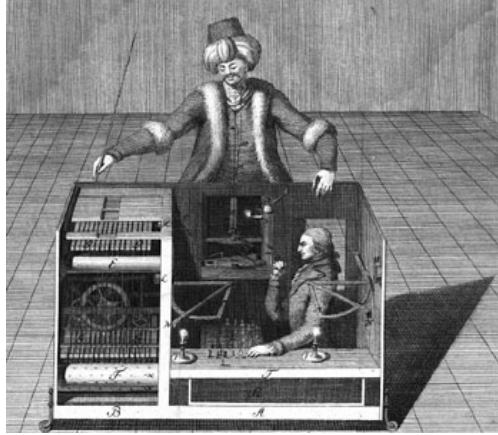
Figure 4: The original Mechanical Turk

3.1.1 Why isn't HAAI Everywhere? Why pursue it?

AI's curious ambivalence to humans: our systems seem happiest far away from humans (Curiosity on Mars) or in an adversarial stance to humans (DeepBlue, AlphaGo).

What happened to co-existence? McCarthy's advice taker? Janet Kolodner's house wife?

Part of the worry: HAAI is cheating. Putting a human-in-the-loop *dilutes* the AI problem. Like the original Turk, seen in Figure 4.
Types of intelligence (in capacity: Humans move top down, AIs move bottom up):

- Perceptual and manipulation intelligent

- Emotional intelligence

- Social intelligence

- Cognitive/reasoning tasks

Point: HAAI expands the reach and scope of the AI enterprise. It's needed everywhere, as in intelligent tutoring systems and social robotics (both already exist), but also digital assistants, hospital assistants, factory floor teamwork, and so on.

In some sense a return to Prof. Barbara Grosz's 1994 AAAI presidential address on Collaborative Intelligence [40].

3.1.2 Research Challenges in HAAI

Posed a particular architecture of an intelligent agent with and without a human-in-the-loop:
Point of AI (from a planning person): what is the next action in the world I should take?

Offered a nice summary of three axes of AI planning techniques, and later, a spectrum of domain models:
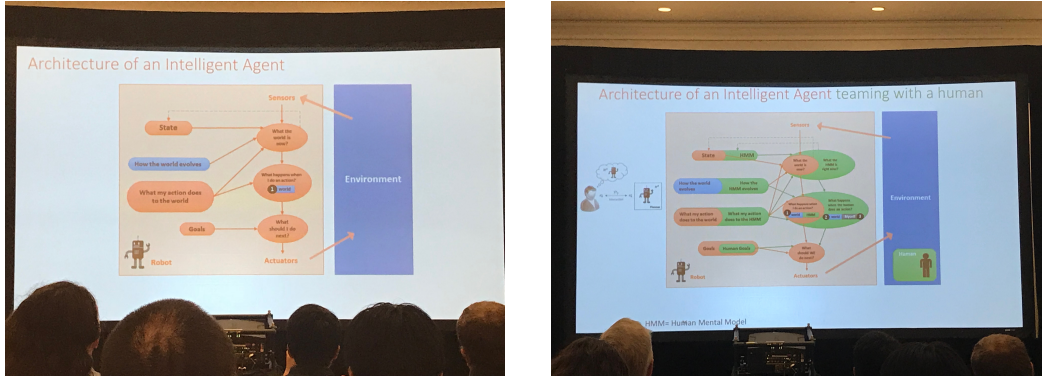
Figure 5: Subbarao's architecture of an intelligent agent (left) and with a human-in-the-loop (right): the main point is that adding a human to the problem changes *every* aspect of the architecture.

In HAAI, this question becomes: what should the next action I take, for the human(s) I'm interacting with?

Primary Agenda for HAAI: Modeling and reasoning with human mental models. Two parts: Modeling and managing (1) Human's mental state, (2) Human's model of the AI system.

Q: Do we really know what (sort of assistance) humans want?

Teamwork with robots requires modeling the human: need a theory of mind that enables intention recognition and intention projection.

Necessary consequence: the human and the robot may have different models of the same tasks. Thus, we get so called "inexplicable plans" where the agents disagree about optimal behavior. Two options:

1. *Explicable Planning*: sacrifice optimality in own model to be explicable to human
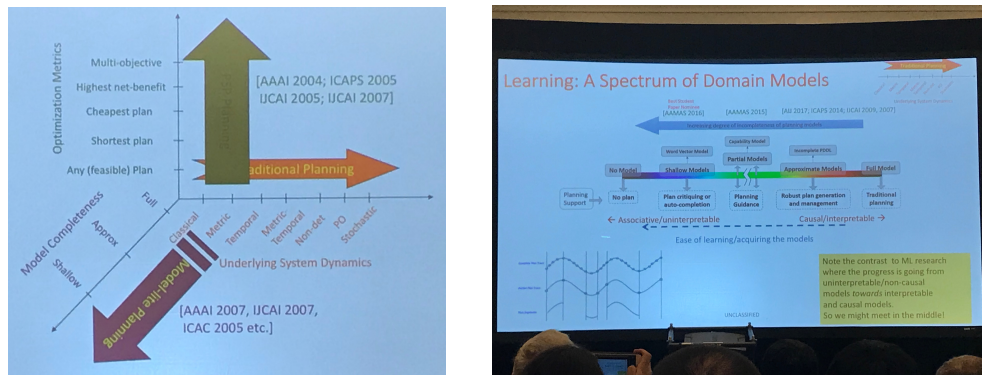
Figure 6: The axes of planning problems (left) and the spectrum of domain models (right).

34

2. *Plan Explanations*: resolve perceived suboptimality by revealing relevant model differences.

Advocates heavily for model conciliation as a route to robust explainable AI, hews close with psychological literature, such as Lombrozo [67]. Model conciliation is a means of identifying and clarifying differences between the human model of the task and the robot model.

Proposes some kinds of explanations in the context of the model reconciliation view. Suggestive of a trade off between explicability and explanation.

3.1.3 Ethical Quandries for HAAI

Mental modeling allowed them to cooperate, compete, or sabotage each other. Lying is possible only because we can model others' mental states.

HAAI systems with mental modeling bring additional ethical quandries and challenges:

- Automated negotiating agents that misrepresent their intentions to gain material advantage.

- Your personal assistant tells you a white lie to get you to eat healthy.

- Humans example closure tendencies are more pronounced for emotional/social intelligence aspects.

- "If only it weren't for the people, the goddamned people, always getting tangled up in the machinery, if it weren't for them, Earth would be an engineers paradise." From the Player Piano by Kurt Vonnegut.

The Fundamental Questions Facing Our Age:

1. Origin of the Universe.

2. Origin of Life.

3. Nature of Intelligence.

It's pretty exciting we all get to tackle one of these!

3.2 Planning and Scheduling

Next up I went to the planning session.

3.2.1 Semi-Black Box: Rapid Development of Planning Based Solutions [75]

The speaker is Michael Katz, joint with Dany Moshkovich and Erez Karpas.

Planning problems are described either (1) symbolically, as in PDDL [74] or STRIPS, (2) A black box approach.

The black box approach describes a problem via an initial state, a successor function, and a goal test.

Q: How do they compare?

A: Symbolic approach allows us to automatically derive heuristics, sometimes lacks interpretability. Conversely, black box is simple, no unnecessary parameters/predicates.

Planning people prefer the symbolic approach, but "Arguably, all others prefer the black-box approach".

This work: introduces a semi black box approach to planning:

- Combines the flexibility of the black-box approach with the ability to automatically derive heuristics.

- Their implementation: a Java framework called "object oriented planning". Allows for incorporation of heuristics.

3.2.2 State-Dependent Action Costs and Conditional Effects [89]

The speaker is Robert Mattuller.

Problem: cost-effect mismatch can lead to an uninformative heuristic. Costs and effects shouldn't be separate. Their approach: handle them in combination.

This is hard, though! Exponentially many combinations. Instead, they propose a new representation for joint cost-effect called decision diagram, based on an edge-valued multi-valued decision diagram (EVMDD) [22].

They come up with a clever method for constructing the joint EVMDD that has some nice properties:

- Worst case: the size of the combined cost-effect EVMDD is a product of factor sizes.

- Best case: the size is the max of factor sizes.

- Efficient computation of relaxed semantics, can compute heuristics.

3.3 Reinforcement Learning

Now for some RL! Chaired by Matthew Taylor.

3.3.1 Efficient Bounds for Inverse Reinforcement Learning [24]

The speaker is Daniel Brown, joint work with Scott Niekum.

The works fall under the umbrella of Learning from Demonstration (LfD).

Goal: How can we bound performance for LfD. In particular, want to reason about the correctness, generalizability and safety of a policy.

36

Bounding policy loss:

$$V_R^{\pi^*} - V_R^{\pi}. \tag{12}$$

That is: how close is the performance of π to π^*? Here, though, in LfD, we don't know R. Instead, we get D, a set of demonstrations, and an evaluation policy, π_{eval}. Want to find an ε such that:

$$V_R^{\pi^*} - V_R^{\pi_{eval}} \leq \varepsilon. \tag{13}$$

Approach: Bayesian Inverse RL: recover a reward function given D.:

$$\Pr(R \mid D) \propto \Pr(D \mid R)\Pr(R). \tag{14}$$

Use Markov Chain Monte Carlo (MCMC) to sample from the above posterior. Assume demonstrations follow a softmax policy:

$$\Pr(D \mid R) = \prod_{(s,a) \in D} \frac{e^{Q^*(s,a)}}{\sum_{a'} e^{Q^*(s,a')}}. \tag{15}$$

How can we bound the polcy loss? Risk sensitive performance bound. α-value at risk (α-quantile.

Experimented with a gridworld with colors denoting features and a driving simulator.

Assumption (for experiments) – linear combination of features:

$$R(s) = w^T \phi(s) \tag{16}$$

Given this assumption, can get a worst-case feature count bound:

$$WFCB(\pi_{eval}, D) = ||\hat{\mu}^* - \mu(\pi_{eval})||_\infty \tag{17}$$

This is a pretty loose bound relative to their actual performance. Also use their performance bound to do policy ranking.

3.3.2 Belief Reward Shaping in RL [71]

The speaker is Ofir Maron, joint work with Benjamin Rosman.

Challenge for RL: reward signal is sparse. Learning would be faster if we rewarded subgoals.

Q: Why not just include subgoal rewards to shape the agent's behavior?

A: Might actually change the problem if you shape incorrectly. But: potential-based shaping from Ng et al. [78] let's us shape without changing the underlying policy:

$$\mathcal{R}_\phi(s, a, s') = \mathcal{R}(s, a, s') + \phi(s, a, s'). \tag{18}$$

Problem with potential shaping: restricts the expressiveness of a potential shaping function.

Their work: belief reward shaping. For a transition (s, a, s') the true environment reward distribution $\mathcal{R}(s, a, s')$ is unknown. We suppose $\mathcal{R}(s, a, s') \sim N(0, 1)$ (that is, it's from a normal distribution).

Set priors on (s, a, s') in a way that guarantees consistent policies without the effect of potential-based frameworks.

In deterministic cases:

$$\hat{r} = \frac{\lambda}{\lambda + n} r_{prior} + \frac{n}{\lambda + n} r. \tag{19}$$

Experimented with backgammon and TD(λ), results look good.

3.3.3 Learning with Options that Terminate Off-Policy [43]

The speaker is Anna Harutyunan, joint with Peter Vrancx, Pierre Luc-Bacon, Doina Precup, Anne Nowe.
Focus: temporal abstractions, formalized by options:

> **Definition 5** (Option [105]): *An option is a triple $(\mathcal{I}, \beta, \pi)$ where:*
>
> - $\mathcal{I} : \mathcal{S} \mapsto \{0, 1\}$ *is an initiation condition.*
>
> - $\beta : \mathcal{S} \mapsto \Pr([0 : 1])$ *is a termination condition*
>
> - $\pi : \mathcal{S} \mapsto \Pr(\mathcal{A})$ *is a policy.*

Focus on the terminal condition, β.
Problem: how do we benefit from the efficiency of long options without suffering from their potentially non-ideal quality (namely, get stuck for too long).

Proposal: terminate off policy.

Off-policy learning: ability to disambiguate behaviors from targets. That is, its behavior policy is different from its target policy.

One idea:

$$\delta_t^{\pi} = R_{t+1} + \gamma \mathbb{E}_{a \sim \pi_{target}} Q(s_{t+1}, a) - Q(s_t, \pi_{behavior}(s_t)). \tag{20}$$

Intuition for the key technical idea:
Instead of having off-policy-ness be implicit in the terminations β, use off-policy learning to impose arbitrary off-policy ness.

So: distinguish between target termination β and behavior termination ζ. Correct for this piece by updating the learning rules for learning with options. Similar flavor to $Q(\sigma)$ from De Asis et al. [25].

Q: Can we run continuing options but learn about terminating ones?

A: Experiments show yes!

The next set are all shorter flash talks.

3.3.4 PAC RL with an Imperfect Model [52]

The speaker is Nan Jiang.

Algorithms are too sample intensive.

Idea: train an algorithm in a simulator and transfer it to the real world. If we have a good enough simulator, we should expect to reduce the sample complexity of learning in the real world.

Surprise: if the simulator differs from the real world by 1 state, than the simulator is "information-theoretically useless."

By looking into this result, can illuminate the conditions that avoid pathological cases. Result: algorithms that achieve sample complexity polynomial in the number of error states in the simulation.

3.3.5 Alternating Optimisation and Quadrature for Robust Control [81]

Goal: learn a policy for controlling a simulator.

Their method: in simulator, you can wisely choose environment parameters to find combinations of policies and simulations to evaluate that lead to more robust policies with fewer samples.

3.3.6 An Experimental Study of Advice [11]

The speaker is Bruno Zanutinni, joint work with Benavent, Caen, and France.

Goal: learn a policy from observations/demonstrations.

Example: national lottery in jackpot land. Pay 1 to play, 0.25 chance to win 100, .5 to win 10. Another one .25 chance to win 16,.
They study: more informative advice demonstrations of the form: " you should play, motivated by games like the first lottery." and look at the impact of such advice on RL.

3.3.7 Change Detection Framework for Piecewise-Stationary MABs [65]

The speaker is Fang Liu, joint with Joohyun Lee and Ness Shroff.

Propose the Non-stationary Multi-Armed Bandit, and introduce two algorithms including CUSUM-UCB that achieves the best known regret bound. Evaluate on Yahoo click-through rates and corroborate the usefulness of algorithm.

3.4 Outstanding Educator Award: Todd Neller on Playful AI Education

The award is given to Todd W. Neller from Gettysburg college.

Encourage students to get to class *early*. Gives them a small incentive – a magic trick – for arriving to class early.

Intro:

- Teachers teach bets when sharing from the core of their enjoyment of the material. For example: those with enthusiasm for graphics should use graphical examples.

- Game playing is one of his favorite hobbies, so he uses games in teaching.

- "Play is our brain's favorite way of learning." - Diane Ackerman

This talk:

- What makes a good game for teaching?

- What are non-game-tree search examples of game use in AI teaching?

- What are some future opportunities in AI education?

Point: Good games have simple rules. (Where good here means effective for teaching AI). Example: breakthrough! (a game from 2001 by Dan Droyken). Conversely, need "fun" depth: Tic-Tac-Toe has simple rules but isn't very fun once you see to its shallow depth.

So: want games simple rules that lead to fun depth. Best games reward players' learning through improvement.

Physical sports are suggestive of robotic competitions that students will get excited about.

"Case" example: Clue, modified. In the modified variant he throws out the board, and players just guess *suspect-room-weapon* triples.
For constraint satisfaction, uses some clue like reasoning.[5]
How about first-order logic? Too many concepts! Like: sentences, operators, literals, truth assignments, satisfiability, models, validity, tautologies, soundness, completeness, ... Too many needed concepts. Instead, let's just go with predicate logic.

Example: the dice game of pig. Rules:

- The first player to reach 100 points wins.

- On each turn, a player rolls a die as many times as desired until either the player holds and scores the sum of the rolls, or rolls a 1 and scores nothing.

---

[5]Details here: http://modelai.gettysburg.edu/2011/clue/

Really high ratio of fun-to-lines-of-code. Can teach probabilities, control structures, monte carlo methods, and so on.

For AI education can use Pig for:

- *Reinforcement Learning*: implement Value Iteration, compare VI solution to TD-Learning, Q-Learning, On/Off-Policy Monte Carlo.

- *Supervised Learning*: regression to fit collected human roll/hold data

- *Unsupervised Learning*: examination of clusters of suboptimal human plays.

Games throughout AI:

- *Bayes*: Gin rummy, estimate an opponents hands.

- *Computer Vision*: Set, robot soccer.

- *NLP*: Computer play of interactive fiction.

Point: We can gamify almost everything in AI.

Activities that bring people together in cultures universally: Working, Eating, Playing Games.

Q: What is a game in its most general sense?

A: After some research, his consensus game definition: *a volunthary activity where players pursue goals according to an agreed-upon set of rules*. This is so general, we can gamify anything! Don't need to be competetive.

Games are not the only means of bringing joy to AI education.

He asked the audience: what things do you in the audience find joy in? And how do you bring it into your AI classrooms?

A: Puzzles! Especially for constraint satisfaction problems.

A: There's a nice app for logic and inference.

A: Having students try to represent their own knowledge on paper for learning knowledge.

A: Working directly with a client.

Future opportunities: computer aided design, game design, trainer. All great opportunities so stimulate AI education!

But, some challenges: computer-aided cheating, game AI assignment plagiarism.

Conclusion: teachers teach bets when they enjoy what they share. Teach to your strengths and ethusiasms.

3.5 Computational Sustainability

Now for some CompSust!

3.5.1 Load Scheduling of Simple Temporal Networks [60]

The speaker is T. K. Satish Kumar, joint work with Zhi Wang, Anoop Kumar, Craig Rogers, and Craig Knoblock.

Temporal constraint solving is really critical for lots of applications, like activity management, scheduling, and space exploration.
Two models of resource consumption: each process P_i (A) consumes electricity at the rate of w_i watts, or (B) demands its entire energy requirement: $W_i = w_i duration(P_i)$.

Solving model B is easier, which they'll use to help solve model A.

Main modeling method is a temporal network. Edges between nodes represents a temporal constraint, with nodes defining processes. Solving the temporal constraint network is effectively a combinatorial problem, which they do some nice graph theory to solve.
The next talks are all shorter flashtalks of the posters:

3.5.2 Traffic Optimization for a Mixture of Compliant Agents [96]

The speaker is Guni Sharon.

Problem: traffic congestion! Leads to hundreds of billions of dollars of unneeded expenditure.

Their goal: redirect drivers to improve traffic flow.

This paper: Uses a stackelberg game to influence a subset of agents in order to achieve desirable behavior.

3.5.3 Kernel Cross-Correlator [110]

Work by Chen Wang, Le Zhang, Lihua Xie, and Jungsong Yuan.

Traditional kernelized correlation filter (KCF) [46].

Limited to circulant training data and non-weighted kernel.

They propose KCC to break these limitations. Retain the low complexity but relax restrictions.

3.5.4 Poisson Gamma Model for Membership in Dynamic Networks [98]

Goal: Design generative models to infer overlapping group structure for *dynamic relational data.*

Run some experiments using this new model on some dynamic networks, seems to work really well.

3.5.5 Dependence Learning with Context via Variational Auto-Encoder [106]

Multi-entity interaction: lots of things interacting in an environment. For example, a leopard and its prey are of course related, but wont often seem the colocated.

They propose a framework based on conditional variational auto-encoders that estimates these interaction relations given rich information about the environment. They use the framework to analyze inter-species relations in the Amazon River Basin. Also produce visuals of these relations.

3.5.6 Catching Captain Jack [118]

Problem: A pirate attack occurred in 2015, with roughly 100 similar incidents occurring between 2007-2017.

Goal: intervene and prevent these attacks with patrol ships.

Challenges: huge number of merchant ships passing through a vast area, and there's high space and time dependence on the attacks.

Their work: Stackelberg model of the Oil-Siphoning Problem along with an algorithm to compute the equilibrium, with an accompanying compact representation that facilitates effective decisions.

3.6 More Reinforcement Learning

Now for some RL!

3.6.1 Phase-Parametric Policies for RL in Cyclic Environments [95]

The speaker is Arjun Sharma joint with Kris Kitani.

Phase in RL: consider a drone flying in a grid world with wind that changes direction. The wind's direction is one of four directions. Thus, there is a *phase* in the MDP: what direction is the wind currently blowing?

Focus: look at problems where phase varies during an episode.

One idea: append phase to state. Problem, though: high dimensional state, low dimensional phase.

Another idea: learn separate estimators for each phase. Problem, though: sample inefficient, no shared learning.

Contributions:

- Look at problems where phase structure is cyclic
- 3 problems: phase varying (i) reward, (ii) dynamics, (iii) states.

- Paraetermize weights of a network on phase.

- Create phase-parametric Q-functions/policies. This can be used with standard algorithms.

Phase Deep Reccurnt Q Network: add control weights of layer j:

$$\beta^j = \{\beta^{j,0}, \ldots, \beta^{j,k-1}. \tag{21}$$

Weight of network at each time step t:

$$W_{p_t}^j = \theta(p_t, \beta^j). \tag{22}$$

What's the θ function? Well: they use cubci Catmull-Rom spline with four ontrol points since it provides "C continuity", and its easy to make a closed loop.

Phase DDPG. Use phase-parameterized MLP or RNN as actor: $a = \mu(s; p, \beta_a)$ and critic:

$$Q(s, \mu(s; p, \beta_a)). \tag{23}$$

They compare their phase-aware methods to vanilla DQN, DRQN, DDPG without phases, and these three with the phase appended to the state.

Run some experimentys in (cyclic reward) grid worlds and (simulated) drone flying, and locomotion (in Mujocu) where their phase-aware methods dominate. Code[6].

3.6.2 Action Branching Architectures for Deep RL [107]

The speaker is Arash Tavakoli, joint work with Fabio Pardo and Peter Kormushev.

Goal: scale discrete-action algorithms to high action space problems.

Major disadvantage of the DQN: direct application to domains with high-dimensional discrete or continuons action spaces is considered intractable.

Combinatorial increase of the number of actions with increasing action dimensions.

New architecture: shared-representation across N dimensions of action branches of the network. Achieves linear increase vs combinational increase in output space.

Key insight: for solving problems in high dimensional action spaces, it is possible to optimize for each action.

Shared network module can help coordinate the action branches and stabilize training.

Ablation study: compare the shared-rep for all action branches to independent representations per action branch, and find dominance by the shared-rep method.

---

[6] github.com/sharma-arjun/phase-dqn

Branching dueling Q-network: a proof of concept action branching agent.

Experiment in a large action space robotic control task, comparing against Dueling DQN with prioritized replay, and find that as the degrees of freedom in the arm increases, their branching approach does extremely well. Also did not some ablation studies again removing the share-rep in some Mujocu tasks, verifying hypothesis that the non-shared rep deteriorates quickly. Lastly, achieved competitive performance on the humanoid control task in Mujocu. Code[7].

3.6.3 $Q(\sigma)$: Multi-Step RL – A Unifying Algorithm [26]

The speaker is Kristopher De Asis, joint work with Fernando Hernandez -garcia, Zacharias Holland, and Rich Sutton.

Return:
$$G_t \triangleq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots \tag{24}$$

Sarsa:
$$\hat{G}_t = R_{t+1} + \gamma G(s_{t+1}, a_{t+1}) \tag{25}$$

Expected Sarsa:
$$\hat{G}_t = R_{t+1} + \mathbb{E}[\gamma G(s_{t+1}, a_{t+1})] \tag{26}$$

Q-Learning:
$$\hat{G}_t = R_{t+1} + \gamma \max Q(s_{t+1}, a) \tag{27}$$

1-step sarsa:
$$\hat{G}_t = R_{t+1} + \gamma Q(s, a) \tag{28}$$

2-step sarsa:
$$\hat{G}_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 Q(s, a) \tag{29}$$

And so on...

Unification of multi-step SARSA. Introduces sampling parameter σ. With a constant value, it's a weighted average between tree backups and regular SARSA.

Allows for interpolation in bias-variance tradeoff, if σ is dynamically adjusted, can enforce a desirable tradeoff. Experimented with prediction in a chain task and a cliff variant of mountain car.

Some intuition: varying σ sort of varies the breadth of the tree backup, contrasted with $TD - \lambda$ where λ varies the depth.

Final two are spotlights of posters.

---

[7] github.com/atavakol/action-branching-agents

3.6.4 An Optimal Online Method of Selecting Source Policies for RL [64]

The speaker is Siyuan Li, joint work with Chongjie Zhang.

Challenge: what and how do we transfer in RL? State of the art lacks theoretical guarantees.

Their approach: optimal method of selecting source policies, formulated as a bandit problem. That is, each policy is an arm. Combine guided exploration with random.

3.6.5 When Waiting is Not an Option [42]

Speaker is Jean Harb, joint with Pierre, Martin Klissarov, and Doina Precup.

Working with Options.

Previously introduced the Option-Critic [8].

New objective for learning options formulated in the bounded rationality framework [101].

Combined this new objective with A2C, called it A2OC. Experimented on Atari.

4 Monday February 5th

First up is the invited talk from Charles Isbell.

4.1 Charles Isbell (and Michael Littman) on Interactive ML

Focus: Interactive Reinforcement Learning. Why bother bringing human-interaction into the mix: Very large search space $> 10^{\text{big}}$. Too many trajectories. Traditional ML says get more data! Interactive ML says: **get a small amount of good data**.

Challenges:

- Representation/model gap between agent/human.

- Effective interaction between agent/human, hard to ensure good agent performance.

- Humans are item noisy, suboptimal, non-stationary.

Summary of interactive ML:
Assumption of a lot of this work: the data from humans aren't great. The noise/suboptimality of the human advice causes issues.

But remember our goal: get *good* data from humans!

Charles' Problem: Why bother taking information from people and use it to accelerate learning, when instead approaches that use lots of data has actually worked! I want some help from the audience. Can we figure out: is interactive ML going to work? Or is traditional ML just going to

Figure 7: A survey of interactive ML approaches.

win out? Can I get some help from the audience?

Michael Littman hopped on stage – (Dave: yay!) some back and forth: "We've never met!" :)

Michael: there *is* a point to interactive ML. Opening arguments in *favor* of interactive ML:

- People are the final arbiter of what the actual goal is.
- People know what's best for themselves.

Charles: so people know what's best for themselves? Allow me to retort:
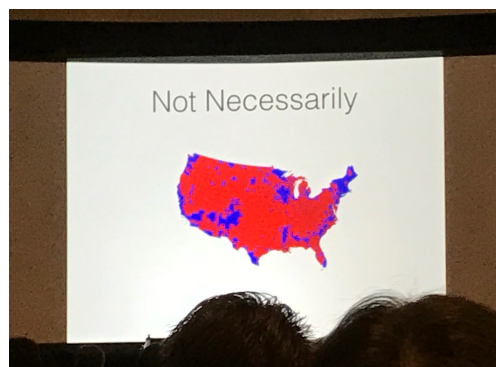
Figure 8: Charles' retort

Michael: let's bring it back to ML. Conjunctions in Trigger-Action programming. Let's think about the Internet of Things, specifically, If-This-Then-That recipes[8].

---

[8] https://ifttt.com/

Michael's work set out to show that people can build more complex recipes than just IFTT. Namely, people can make complex recipes involving conjunctions!

Charles: but someone else wrote a paper a year later that showed people don't know what they mean, logically [50].

For example: Consider the phrase, "If it is snowing, then turn the thermostat to 75 degrees F". Folks were asked: when will the thermostat be set to 75 degree and when will it turn off? Most folks didn't get this right.

Michael: Let me use an example from *your* research. Problem decomposition from humans in Pac-Man [103]. Asked people which high level actions would help make the game easier, exampled include "go to pac dot, go to energizer, eat ghost, avoid ghost". These map really well to options. So: you can't disagree with yourself!

Charles: I agree, that is great work! But, there's another example in the same space we should mention. Consider Radiation World: sparse reward grid world, effectively search and rescue with pellets of radiation lying around that mes with the robots navigation system. Ask humans to do the same thing: high level actions "go to human, avoid radiation". But! What people tell you, you cannot take literally. People specify incorrect constraints such that in some cases you have to ignore it. System believes too strongly what people tell them.

Michael: So, if you make systems that take all of this literally, we end up having to ignore them. So how about instead we think about Human RL feedback.

Charles: studies indicate people are *different* –

1. Feedback on actions [58].

2. No reward is context dependent [66].

3. Explicit rewards decrease over time [51].

Michael: We knew these shortcomings. Instead let's do this: plug the human feedback into the advantage function of an actor-critic, *instead* of into the reward function. Supports: (1) Diminishing returns, (2) Differential feedback (multi-scale), (3) Policy shaping. So, maybe I've done my job. Now you can go back to your regular talk.

New talk title: "How to Learn From Humans: In Spite of the Fact that they Don't Know What They Are Doing."

Charles: It's close, but how about: "How to Learn From Humans With Methods That Know What They Are Doing Even If people don't".

What have we learned:

1. It's *not* smart vs dumb. People are *different*. It's a different form of feedback from traditional ML labels or RL feedback (from a fixed reward function).

2. We cannot outsource our understanding of humans. We need to do the psych/cog sci to understand how humans work. The humans *are the loop*, they're not in the loop.

Michael: so my work here is done. Now you can give your talk! Title: "How to Learn From Humans: With Methods That Meet Them Where They Are."

Charles: Great! So I'd love to give this talk, but of course, we're out of time.

4.2 Learning Theory

Now for some Learning Theory.

4.2.1 On the ERM Principle with Networked Data [112]

The speaker is Yuyi Wang, joint work with Yuanhong Wang, Xingwu Liue, and Juhua Fu.

Problem: Predicting edge relations in a social network.

That is: Each node is a person, we get features for each person. Model as a graph with 0s and 1s, want to predict if any pair of people are connected. Assume symmetry:

$$\ell(r, (x_i, x_j, y_{i,j})) = \ell(r, (x_j, x_i, y_{i,j})). \tag{30}$$

With a complete graph, use empiricial risk minimization (ERM). That is, choose r that minimizes:

$$L_m(r) = \frac{1}{n(n-1)} \sum_{i,j} \ell(r, (x_j, x_i, y_{i,j})). \tag{31}$$

But: we don't get a complete graph. Lots of $y_{i,j}$ are not known. If we still use ERM, then risk bounds are $O\left(\sqrt{X^{\cdot}/m}\right)$, with $X^{\cdot}$ the fractional chromatic number of D_G, the graph.

Issue: equal weighting of nodes leads to trivial bounds. So, they instead use a weighted ERM from [111]. Yields a risk bound of $O\left(1/\sqrt{||w||_1}\right)$, where w is just the vector of weights of the graph. Dave: But wait: $||w||_1$ doesn't depend on sample size?. Someone asked this Q, too: speaker said $||w||_1$ is a function of m. So This bound is tighter and works for general graphics, not just complete graphs.

Conclusion: don't use unweighted ERM for networked data. Can get a tighter universal risk bound for networked data when data graph is incomplete.

4.2.2 A Provable Approach for Double-Sparse Coding [79]

Presented by Than Nguyen, joint with Raymond Won, Chinmay Hegde.

Problem: **sparse coding**. The goal is to learn a sparse representation from high dimensional data.

One model: dictionary learning. Highly adaptive.

Applications of sparse coding: image processing, regularization, pre-training, comp. neuroscience.

Definition 6 (Dictionary Learning): *In the sparse coding problem, the dictionary learning method is given: p data samples, wants to find a dictionary A and sparse code X that minimizes the objective:*

$$\min_{A,X} L(A, X) \tag{32}$$

Challenges:

- *Theoretical*: extremely non convex

- *Practical*: memory and running-time issues as the signal dimension n is big.

  Solution: Impose further structure on the dictionary (separable, sparse, convolutional).

One variant is **double-sparse coding**. That is, the dictionary is sparse in some known basis Φ. They achieve the first known bounds for double-sparse coding.
Conclusion: probable statistical algorithm for double sparse coding that is (1) robust to noise, (2) practically useful, and (3) neurally plausible.

4.2.3 Sample Efficient Learning of Mixtures [7]

The speaker is Shai Ben-David, joint work with Hassan Ashtiani and Abbas Mehrabian.

Problem: PAC learning of probability distributions (a.k.a. density estimation):

Definition 7 (Density Estimation): *Given an i.i.d. sample generated from an unknown density g^*, find a density $\hat{h}$ that is "close" to g^*.*

In the most general case, hopeless. No finite sample guarantees. So, some assumption: Assume g^* belons to (or can be approx by)a fixed class of distr. $\mathcal{F}$.

New Q: What is the sample complexity of learning with respect to $\mathcal{F}$.
Specifically, mixture:

$$k_{mix}(\mathcal{F}) = \sum_{i=1}^{k} w_i f_i \tag{33}$$

Suppose $\mathcal{F}$ class of mixture of k Gaussian distributions over $\mathbb{R}$.

Then, problem: consider the PAC setting where the learned mixture is close with high probability.

For 1-dimensional mixture of k Gaussians, we need $O\left(\frac{k + \log 1/\delta}{\varepsilon^2}\right)$.
Main result:

Theorem 4.1. *For any natural class $\mathcal{F}$, the class of k-mixtures of $\mathcal{F}$ can be learned with:*

$$O\left(\frac{k \log k}{\varepsilon^2} m_{\mathcal{F}}(\varepsilon, \delta/3k)\right), \tag{34}$$

samples, where $m_{\mathcal{F}}()$ is the sample complexity of learning $\mathcal{F}$.

So, by going from one component to k components, the sample complexity increases at most by a factor of $(k \log k)/\varepsilon^2$.

Generic result but surprisingly tight! The lower bound almost matches this.

Note: this is distinct from *parameter learning*. In that case there are issues of representation, parameterization, identifiability, separability.

The goal is to have $d(g, \hat{g}) < \varepsilon$, not guess the parameters. Many choices for disimilarity measure.

How to measure distance: (1) total variation, (2) KL, (3) Hellinger. Lots of measures. In their work, they address the total variation:

$$||g - \hat{g}||_{TV} = \sup_{A \subset \mathbb{R}^d} |g(A) - \hat{g}(A)|. \tag{35}$$

Can also handle the agnostic case, in which the distribution does not come from the family, just end up iwth a constant multiplied by the best entry in the family.

Conclusion:

- If a class is learnable so is its mixture.

- Sample complexity of learning $k_{mix}(\mathcal{F})$ is at most a factor of $\tilde{O}(k/\varepsilon^2)$.

4.2.4   Algorithms for Generalized Topic Modeling [12]

The speaker is Nika Patel, joint work with Avrim Blum.

Problem: topic modeling.

> **Definition 8** (Topic Modeling): *The topic modeling problem is a cased where, given an unlabeled corpus of documents, can we automatically learn their subject matter?*

Traditional topic modeling assumes:

- A topic is a distribution over words.

- A document is created by choosing a weighted combination of topics.

- Draw some words i.i.d. from this distribution yielding a sparse vector of words.

- Goal: From the sparse vector of words generate the list of topics.

This work: let's go back and think about the model.

One existing assumption is that words are draw i.i.d., but this isn't right. Documents contain *correlated* words and correlated topics.

Here: documents are paragraphs, where paragraphs are points in a feature space. Features are words, phrases, length of the document, and so on. A document can have correlated features.

Goal: given unlabeled documents, learn to predict topics.

New model: consists of k topics $a_1, \ldots, a_k$. Each paragraph X projected on $span(a_1, \ldots, a_k)$, a convex mixture of topics and some noise. A document is some number of paragraphs $(x^1, \ldots, x^t)$ with the same mixture.

Some assumptions for the main result:

1. Each document is two paragraphs.

2. Some paragraphs have Gaussian noise, some have no noise

3. There exists a nearly-pure documents for any topic i (like a medical textbook that every med-student must read).

Theorem 4.2. *We can recover $a_1, \ldots, a_k$ within $\varepsilon \in (0, 1]$ using $O(n/\varepsilon^2)$ samples. Further, need $\Omega(n/\varepsilon)$ to recover $a_1, \ldots, a_k$.*

Conclusion: let's rethink about our model for studying documents. Introduce the generalized topic model and some nice clean results on this new model.

Open Problem: What if every document or paragraph is perturbed?

4.3 Planning and Scheduling

Next, planning.

4.3.1 Sublinear Search Spaces for Shortest Path Planning [53]

The work is by Johannes Blum, Stefan Funke, Sabine Storandt.

Problem: given an unweighted graph, and a query $s \in V, t \in V$ we want to compute the distance of the shortest path from s to t.

But: isn't this Dijkstra's? runtime: $O(n \log n + m)$, with n number of nodes and m number of edges. Actually: too slow in practice.

Q: What if we add a preprocessing step? Some ideas:

- Compute transit nodes

- Contraction hierarchies

- Hub labels

This preprocessing steps speed things up in practice *a lot*. But can we prove how much it improves? Not really, in general, because of pathological cases.

So, let's focus on a subset of all graphs. One common assumption made to yield bounds is to suppose a finite *highway dimension*, as in Abraham et al. [4].

Another analysis is on bounded growth, as in Funke and Storandt [37]. Here we assume exists a constant $c \in \mathbb{R}$ such that for all nodes and all radii r:

$$|\{w \in V : d(v,w) = r\}| \le c \cdot r \tag{36}$$

Transit nodes [9]: $T = \uplus_{v \in V} AN(v)$, where $AN(v)$ are "access nodes". Something about the connectivity structure in the graph. Idea here is to precompute some distances in the graph which gives you these node types. This precomputation results in a ε-Net Set System (U, S), a data structure that captures the results of this pre-processing phase.

4.3.2 Risk-aware Proactive Scheduling [114]

The work is by Wen Song, Donghun Kang, Jie Zhang, and Hui Xi.

Focus: Resource Constrainted Project Scheduling Problem (RCPSP).

Idea: actions $A = \{a_1, \ldots, a_N\}$, have a duration $d_i \in \mathbb{N}$ and require some resource, $r_i \in \mathbb{N}$.

Goal is the Risk-Aware objective: achieve a good execution with high probability.

Use a model callled a Partial-Order Schedule (POS) from **?**]. Using this model, they do some risk-aware optimization.

4.3.3 LatPlan: Classical Planning in Deep Latent Space [6]

The speaker is Masataro Asai, joint work with Alex Fukunaga.

In this work they introduce LatPlan, which combines symbolic classical planning with neural planning.

Main limitations of current general AI agent architectures: transition model and representation typically given to the agent.

Classical planning, like PDDL can solve 8puzzle in 0.01 second. But: they can't solve an *image based* version of 8 puzzle. Thus, classical symbolic planners are impractical in many environments!

So: we need to automate two processes:

1. **Symbol Grounding**: Given an image based dataset of transitions, we need a PDDL model.

2. **Action Model Acquisition** (AMA): Describe the transitions with symbols.

This paper: solves the second (long standing open problem) with a latent-space planer called Lat-Plan.

Task: solve an image-based 8-puzzle *without any prior explicit knowledge.* Thus, they offer a domain-independent image-based classical planner.

The system is given 2 types of input and returns the visualized plan:

1. Training Input: randomly sampled transitions of image pairs of (s, s'). No explicit state description, no expert traces, no access to simulator, no rewards, no action symbols.

2. Planning Input: inital image (init state) and the goal image.

LatPlan archiecture:

- Takes the two inputs described above.

- Has a domain independent classical planner in the inner loop.

- Uses a State Autoencoder, which yields a propositiaonl symbol grounder.

- Next step: learn the action model from transitions, converting them into a PDDL model.

- Then, simulate a plan to yield the intermediate states.

Q: Does the SAE produce sound propositions for reasoning?

They test with an oracular Action Model Acquitisition (AMA) model. AMA_1 encodes all image transitions, for each transition, and generates PDDL (does not generalize: encodes every transition). But: AMA_1 is impractical, and requires explicit construction of actions and their preconditions/effects. In this work, they introduce use a newer model, AMA_2m, a neural architecture that automatically builds action-precondition-effect models via learning a successor function.

Conclusion: Introduce LatPlan, the first method for turning unstructured images into crisp symbolic plans.

4.3.4 Solution to Scenario Generation for Risk Management [102]

The speaker is Shririn Sohrabi, joint work with Anton Riabov, Michael Katz, and Octavian Udrea.

Scenario Planning: method that organizations use to prepare them for the future. Analyze things like technology, currency, social order, corruption, market disruptors, and so on.

Contribution: Characterize the scenario plan problem as a plan recognition and bring tools to bear on solving these problems.

Figure 9: Game complexity

Plan recognition: given a set of goals and observations, want to find a probability distribution over plans or goals.

Build on the plan recognition framework of Ramırez and Geffner [86].

Take home:

1. Many application of plan recognition: scenario planning is one example.

2. Knowledge engineering is challenging, but things don't have to be *perfect*.

4.3.5 Minesweeper With Limited Moves [93]

The speaker is Kevin Tran, joint with Serge Gaspers, joint work with Stefan Rummele, and Abdallah Saffidine.

Q: How difficult are games?

A: Two ways to measure: (1) Empirically! Just play it. (2) Prove that certain games fall somewhere in the complexity hierarchy.

Three axes for determining the difficulty: [a] How long the game is (how many actions), [b] How many players, [c] How much information you have when you play.
The first two [a] and [b], are well understood. Here they focus on [c] in the context of Minesweeper.

Results on Minesweeper:

- Deciding if Minesweeper board is consistent is NP-Complete [55]

- Counting number of consistent mine placement is #P-Complete.

- Winning with constant probability is PP-Hard.

Their work: `Input`: A partially played Minesweeper grid. `Output`:Is it possible to detemrine the location of all the mines with probability y with x clicks.

Theorem 4.3. *If the number of clicks is input, and the goal is to win with probability 1, the problem is PSpace-complete.*

4.4   IAAI: Utilities and Transportation

Next up is my session of IAAI on utilities and transportation.

4.4.1   A Water Demand Prediction Model for Indiana [94]

The speaker is Setu Shah, joint work with Mahmood Hosseini, Zina Ben Miled, Rebbecca Schaefer, and Steve Berude.

Increaseing need for water demand modeling for both short term (operational purposes) and long term (policy).

Their focus: three models on two time-scales. Inputs to the water are the daily water demands from 1997-2016, the service area, the number of customers per year, weather, calendar info (holidays, day of week), and so on.

The accuracy of the daily model was 1.69%, the monthly model was around 2%.

They test with multiple linear regression and a few neural networks for prediction. For evaluation, an average error metric with some weights on various features.

Finding: The average error of their daily-modified-RNN (recurrent net) achieved average error of 3.17, achieved the best performance. Did some analysis of the importance of correlations between features (like precipitation as a function of the month). Tested it on real data from last few years, achieves high accuracy on these test sets.

4.4.2   Optimal Pricing for Distance-Based Transit Fares [49]

The speaker is Richard Hoshino, joint work with Jeneva Beairsto.

Connect policy to transit networks. Their university doesn't declare majors but *questions*. Dave: How cool! Jeneva's question: how can we optimize transportation in cities?

Pricing system in Vancouver, BC: unfair fare system for public transit!

- Zone boundary system has been in place for over 30 years.

- 1-zone trips longer than 30km, and 3-zone trips shorter than 15km.

- High fares for short trips disincentivis ridership.

Draw inspiration from cities in Europe/Asia that have less congestion. Point: they use *distance based fares*. The more travel the more you pay.

Challenge: how can we do a distance based fare?

Organization behind Vancouver's transit: TransLink. How do we convert zone fares into distance based fares? Is there an optimal way to do it?

Simple ideas:

1. *Price elasticity*: Suppose a $+f\%$ change in the fare causes the ridership to change by $-r\%$. Then $k = \frac{-r}{f}$ is the transit fare elasticity.

   New ridership: $\left(1 - k\frac{x-4}{4}\right) \cdot 100$ New revenue: $\left(1 - k\frac{x-4}{4}\right) \cdot 100 \cdot x$

   Maximize this with calculus.

2. *Cauchy-Schwarz Inequality*: They have a theorem leveraging the inequality where, with m zones, n distance based fares, $-k$ be elasticity, $r_{i,j}$ be the number of riders who pay X_i by travelling a distance of i units, then they can compute the exact assignments to the distance based pairs that maximize total revenue.

The city of Vancouver is moving to a distance based fare as a result of this research!

4.4.3 Upping the Game of Taxi Driving in the Age of Uber [97]

The speaker is Shashi Shekhar Jha.

Taxi industry in Singapore. In 2012, 500ish Uber/private hire cars, around 30k taxis. In 2018, 41k Ubers/private hire cars and 26k taxis. Further, 20% drop in average daily taxi trips in the last 1.5 years.

Goal: how can we help taxi drivers get more trips?

For Taxis: 75% of trips come from normal/stree-hail trips (with 15% pre-booked and 10% app-based). So, this work focuses on street-hailing.

Study imbalances of supply and demand in various zones in the city. The hope is that they can inform taxi drivers *where to go* in order to maximize number of trips. Developed a Driver Guidance System (DGS). Collect some data on supply/demand of taxis.

Main Q: How should we predict the demand (over time and location)? Address this Q with multi-layer logistic regression using time-elapse data. Compare their performance to a Poisson based model.

Experimented with a multiagent simulation, TaxiSim [20], results look good. Launched a mobile phone app in singapore for taxi drivers.

Dave: Next up is my talk (which I won't take notes on!). But, our paper is here [2]. We investigate improving solar panels with contextual bandit methods.

....

5 Tuesday February 6th

This morning begins with the presentation of the awards.

5.1 Awards

Best Paper Award: "Memory-Augmented Monte Carlo Tree Search" by Chenjun Xiao, Jincheng Mei, and Martin Muller [116].

Best Paper Honorable Mention: "Generalized Adjustment under Confounding and Selection Bias" by Juan Correa, Jin Tian [23].

Best Student Paper: "Counterfactual Multi-Agent Policy Gradients" by Jakon Foerster, Gregory Farquhar and Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson [32].

Best Student Paper Honorable Mention: "Adapting a Kidney Exchange Algorithm to Align with Human Values" by Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John P. Dickerson, Vincent Conitzer [35].

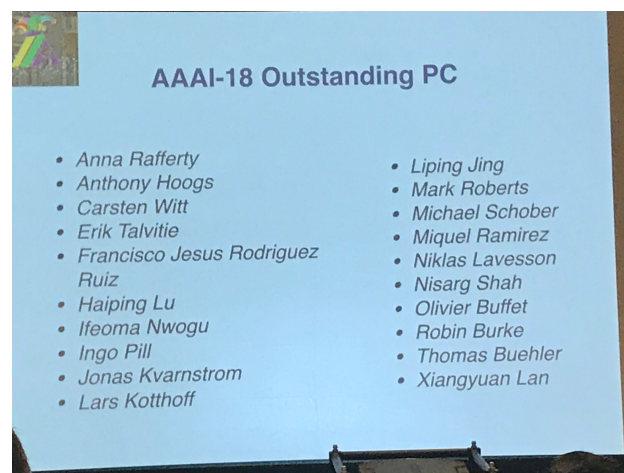**Outstanding PC Award**: Jörg Hoffman.

Outstanding PC:

Figure 10: Outstanding Program Committee!

Best Student Abstract: Ellis Hoa and Janardhan Doppa for the poster: "Bayesian Optimization Meets Search Based Optimization"

Best Student Abstract Honorable Mention: Chengqiang Huang (Dave: and colleagues, I missed their names!): "Toward Experienced Anomaly Detection through RL".
Next up: plenary talk!

5.2 Cynthia Dwork on Discrimination & Bias

Motivation: Humanity is diverse! Lots of ethnic, religious, medical, gender, and class diversity.

Problem: bias in *training data*, fairness in traditional 0/1 classification.

Framework: traditional supervised learning setup where the training data describes individual people. The outcome impacts user actions, as in college admissions, health care, insurance rates, and so on.

Some group fairness notions [28]:

- *Statistical Parity*: demographics of those receiving label o is the same as the demographics of the population.

- *Equal False Positive/Negative Rates Across Groups*: Fraction of negatives erroenously labeled "yes"/"no".

- *Equalized Odds (Conditional Parity)*: That is: $\Pr(\text{outcome} = o \mid \text{truth} = 0)$.

5.2.1 Statistical Parity

Recall: demographics of those receiving label o is the same as the demographics of the population. So:

$$\Pr(x \in S \mid \text{outcome} = 0) = \Pr(x \in S). \tag{37}$$

Or:

$$\Pr(x = 0 \mid x \in S) = \Pr(x = 0 \mid x \in S^c). \tag{38}$$

This does permit bad outcomes, such as intentional targeting of subset of s.

Consider reoffense probability p. Result from Chouldechova [21] and Kleinberg et al. [57]: No imperfect classifier can simultaneously ensure equal FPR, FNR, PPV unless base rates equal:

$$FPR = \frac{p}{1-p}\frac{1-PPV}{PPV}(1 - FNR), \tag{39}$$

with p the base rate of reoffending in the population.

Fairness: Group notions of fairness fail to capture the feelings of individuals. *A statistic in general doesn't help individual fairness, especially when other confounding factors are at play.* Studied in Dwork et al. [28]

Let's think about credit score. What if we want two similar individuals to be treated fairly? Randomize! How about we apply a Lipschitz constraint, with $M : V \mapsto \Delta(0)$:

$$||M(x) - M(y)|| \leq d(x, y). \tag{40}$$

That is: people who are similar for a particular classification task should receive similar treatment on that task. Sort of a differential privacy flavor.

Suppose we do manage to ensure statistical parity iff groups are similarly distributed.

Locus of unfairness: Let x be the probability of some event occurring through interaction with the environment, like getting a raise, getting sick, and so on. That is: π_x is the probability, over randomness in x and environment, that x will recidivate. We ensure that if $|\pi_x - \pi_y| \le \varepsilon$, then x and y should be classified as hvaing similar probabilities of being classified as high risk.

But! Hostility of environment differs across groups. So the above reasoning doesn't quite hold.
Open Problem Area: How can we use AI to help find an individual fairness metric?

Q: What *should* the metric capture? Success probability? Talent?

One thought: Learning fair representations by Zemel et al. [119]. They learn *prototypes*:

$$\mu_{x,z} = \frac{e^{-d(x,z)}}{\sum_{z'} e^{-d(x,z')}}, \tag{41}$$

that "forget" group membership,m yielding statistical parity, and permit reconstruction of other attributes: $x \approx \sum_z \mu_{x,z} z$, and preserve $y = g(x)$.

Here, we now need a *meaningful distance metric*, and strive to statistical parity.

Adversarial LFR [5] learns a censored mapping to a lower dimensional space Z:

- Encoder hides membership bit, learn classifier on Z

- Decoder tries to reconstruct x from $z = Enc(x)$.

- Adversary tries to distinguish $Enc(x \in S)$ from $Enc(x \in S^c)$.

Key intuition: if the predictor is not more powerful than the adversary, than the predictions will be relatively unbiased.

Theorem 5.1. *For any tasks T, T', with not identiacal, non-trivial metrics D, D' such that there exist individually fair classifiers C, C' that, when naively composed, violate multi-task fairness. Formally: $\exists_{u,v \in U} f(T) \wedge \neg f(T')$, where $f(T)$ denotes the predicate "is fair".*

Open Problem Area: Improve on the economics of the simple solution.

Open Problem Area: Explore functional composition properties of fair representations.

Some philosophical issues:

- Consequences of leniency when most violineve is intra-racial (Berk).

- Invest, invest, invest, in education of disadvantaged issues (Roemer).

- Accuracy/correctness is the wrong lens through which to access wrongfulness and ds

5.3 Reinforcement Learning

Next up, RL session (mostly Deep RL, it seems).

5.3.1 Counterfactual Multi-Agent (COMA) Policy Gradients [33]

The speakers are Jakob N. Foerster and Gregory Farquher, joint with Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson.

Motivation: cooperative multi-agent settings are important.

Problem Setting: Cooperative, partially observable multi-agent RL. They use a Deep Multi-Agent Actor-Critic algorithm.

Main Idea: Centralized critic, decentralized actors, counterfactual advantage for credit assignment. Uses REINFORCE [115]:

$$g = \mathbb{E}_{s_0:\infty, u_t \sim \pi} \left[\sum_{t=1}^{T} R_T \nabla_{\theta^\pi} \log \pi(u_t \mid s_t) \right], \tag{42}$$

where u_t is the action because a will be agent.

Recall the advantage function:

$$A^\pi(s_t, u_t) = Q^\pi(s_t, ut_) - V^\pi(s_t). \tag{43}$$

They use the Counterfactual Advantage:

$$A^a(s, u) = Q(s, u) - \sum_{u'^a} \pi^a(u'^a \mid \tau^a) Q(s, (u^{-a}, u'^a)). \tag{44}$$

Experimenting on Starcraft 1 micromanagement tasks. Control a team of five or so units, each of which only gets local observations, fighting against five or so other agents. Benefit comes both from centralization and the counterfactual piece. Biggest empirical result: achieve comparable performance a centralized, fully observable controllers (like DQN).

5.3.2 Rainbow: Combining Improvements in Deep RL [47]

The speaker is Matteo Hessel, joint work with Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver.

Problem: Deep RL agents are mostly evaluated against a single outdated baseline (DQN).

Focus: how can we best combine improvements of Deep RL agents? Study combination of recent techniques for Deep RL.

Mixture of methods:

- Distributional DQN [10]: Improves robustness.
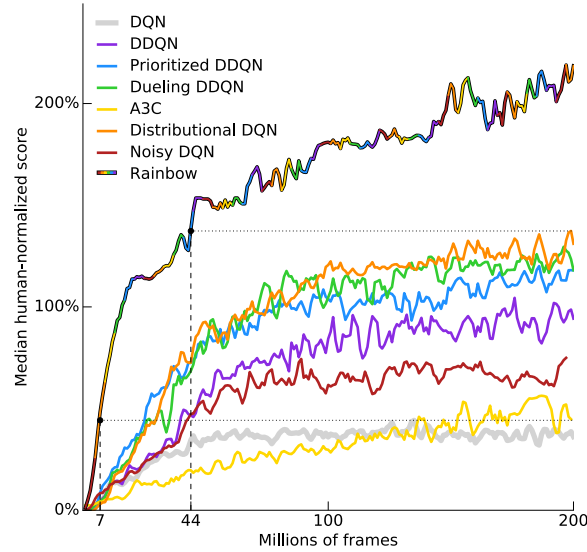
Figure 11: Rainbow Results compared to individual DQN improvements from Hessel et al. [47].

- Double Q-Learning [44]: Addresses over estimation.

- Dueling DQN [113]: Addresses stability.

- Prioritized Experience Replay [91]: improves data efficiency.

- Stochastic NNs [34]: For exploration.

- Multi-step updates [104, 76]

Did some experimentation and ablation studies with each component in Atari, main result averaged over all atari games:
Conclusions:

1. DQN's performance can be achieved in hours, but do this, we need comprehensive agents.

2. We should test how a single idea shifts the boundary of what is possible with RL?

5.3.3   Safe RL via Formal Methods [36]

The speaker is Nathan Fulton, joint work with Andre Platzer.

Setting: autonomous safety-critical systems.

One approach: model-based verification. Here, prove that control software achieves a specification with respect to a model.

- **Benefits**: Strong safety guarantees, automated analyses. **Drawbacks**: Control policies are typically non-deterministic. Models answer "what-is-safe" not "what-is-useful". Further, assumes accurate model.

Other approach: RL.

- *Benefits* No need for complete model, optima policies.

- *Drawbacks*: No strong safety guarantees, proof are obtained by hand, decades-long proof development.

Goal: Provably correct RL

1. Learn safety

2. Learn a safe policy

3. Justify claims of safety

Use a model-based method, to prove RL is safe. Why? Because accurate and analyzable models often exist!

Problem: Model-based verification isn't enough because perfect models don't exist.

Their contribution: **justified speculative control**, an approach for provably safe RL learning that:

1. Learns to resolve non-determinism without sacrificing safety.

2. Directs speculation when model mismatch occurs

Main theorem:

Theorem 5.2. *If the differential equations that characterize the model are accurate, then proofs transfer from the non-deterministic model to the learned deterministic policy.*

Shortcoming: the assumption that the ODEs are accurate is too strong. They relax this assumption and do additional analysis under this relaxation.

Dave: I had to step out the rest of the day to work on ICML submissions. I did come back for AAAI games night! Hosted by Michael Bowling and Michael Littman – it was great! What a fun tradition. We played an AI rendition of Hollwood Squares and a game called Codename(?).

....

6 Wednesday February 7th

First: two awards for the demo track.

Best Demo Honorable Mention: "Perception-Action-Learning system for Mobile Social-Service Robots using Deep Learning", by Beom-Jin Lee, Jinyoung Choi, Kyung-Wha Park, et al.

Best Demo: "Cognitive Assistant for Visualizing and Analyzing Exoplanets" by Jeffrey Kephart, Victor Dibia, Jason Ellis, Biplac Srivastava, et al.

Next, Percy Liang is giving a talk on Evaluation in Machine Learning.

6.1 Percy Liang: Evaluation in Machine Learning

Kilian introducing Percy: ML is focusing too much on accuracy, and so leaves to many other things on the table. Percy Liang will talk to us about this movement!

I always get the question in the first day of class: How is AI different from ML? Here, we'll say AI is a set of behaviors that a system can have, like perception and language, while machine learning is the set of tools we have.

Lots of success stories in machine learning in speech, game playing, vision, language. But, adversarial examples, such as those from Eykholt et al. [30] and Goodfellow et al. [38], poke holes in this success.

$$x \quad\quad \text{sign}(\nabla_x J(\boldsymbol{\theta}, \boldsymbol{x}, y)) \quad\quad \begin{array}{c} \boldsymbol{x} + \\ \epsilon\text{sign}(\nabla_x J(\boldsymbol{\theta}, \boldsymbol{x}, y)) \end{array}$$

x	$\text{sign}(\nabla_x J(\boldsymbol{\theta}, \boldsymbol{x}, y))$	$\boldsymbol{x} + \epsilon\text{sign}(\nabla_x J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"panda"	"nematode"	"gibbon"
57.7% confidence	8.2% confidence	99.3 % confidence
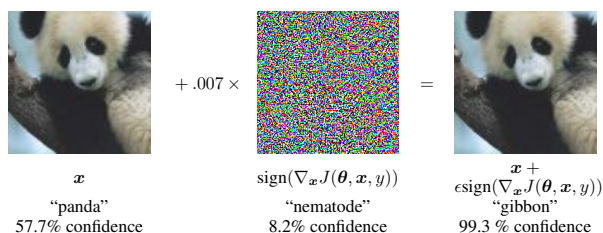
Figure 12: An adversarial example from Eykholt et al. [30].

Adversarial Example Takeaway: the existence of these examples is suggestive of there being something wrong with how we're thinking about machine leanring.

So, how do we close this gap?

- Tenenbaum: probabilistic programming

- Bengio: consciousness prior!

- Hinton: capsules!

Percy: well, we need an objective evaluation metric. It used to be accuracy but we need something new.

Hector Levesque: "We want multiple-choice questions that people can answer easily, but we also want to avoid as much as possible questions that can be answered using **cheap tricks** (aka heuristics). " (2013).

So, proposed Winograd schema: "The dog chased the cat, which ran up a rate, it waited at the top." vs. "The dog chased the cat, which ran up a rate, it waited at the bottom."

Q: What does "it" refer to in each question?

Statistics don't really care about things like commonsense knowledge. So, the central problem is that our existing philosophy is insufficient.

Point: Status quo is that we split our data set into train/test and we evaluate on test.

Any expressive enough model with enough data will do the job. And we have lots of data! So we can fit *something* to our data.

Claim: What we need is a harder metric. Such as: extrapolation.

That is: to extrapolate, and be robust, we must get a more "correct" model.

Example 1: Laws of physics:

- *Train*: small objects, the past.
- *Test*: large objects, the future.

Example 2: Language/comp sem:

- *Train*: the blue block
- *Test:* right of the blue block on top of the red block.

Two ideas:

1. Harder data.
2. Models that extrapolate.

6.1.1 Harder Data

They designed Codalab[9], an website for sending in models to be evaluated on traditional training tasks. They also then evaluated each model on adversarial training task (added a sentence to a paragraph in the training set). The performance of *all* models plummet on the adversarial setting. Conversely, humans don't get fooled (basically no performance drop).

Issue: you cna find plausible answers with cheap tricks.

---

[9]http://codalab.org/

Q: How can we create tests that *can't* be solved by cheap tricks?

Asked Turkers to create "negative" questions with plausible answers, that is.

Example: "Victoria has 60,000 full-time teachers". Q: "How many janitor's does Victoria have?". Models answer: "60,000". Clear case where the model's learn cheap tricks.

NLP Task: good or bad story ending? Given some contet to a story (one sentence), you're asked to predict which of two endings is the correct ending. But, if you completely ignore the context and just look at the endings, an n-gram model gets around 70% accuracy.

Solution: Come up with a reweighted version of the training set based on some superficial features. Key idea: model using only superficial features gets bad performance. Doing this leads to lowered performance for existing models (down to around 60%).

Takeaway: Yes, there are security implications to adversarial examples, but also the models are learning the wrong thing.

6.1.2 Robustness Against Attacks

Propose a game in which the adversary can perturb the input by at most ε in each feature/pixel.

Optimal Attack: $A_{opt}(x) = \arg\max_{\tilde{x}} f(\tilde{x})$. But, this is intractable, so folks use the Fast Gradient Sign Method (FGSM): $A_{fgsm}(x) = x + \varepsilon\text{sign}(\nabla f(x))$.

Can we upper bound an optimal attack? Idea: uniformly bound grads:

$$f(\tilde{x}) \leq f(x) + \varepsilon \max_{\tilde{x}} ||\nabla f(\tilde{x})||_1. \tag{45}$$

Still intractable. Do some additional work to simplify the computation, yielding a certificate based attack. Idea is to *include* the certificate in the objective of the training. So, something like:

$$\min_{W,v} \sum_{i=1}^{n} L(x_i, W, v) + K, \tag{46}$$

where K is a certificate acting as a regularizer. The bound is still vacuous, but if you *train* with the certificate, it's somewhat reasonable. Similar idea explored in Kolter and Wong [59].
Summary of the last two sections:

- Convex relaxation obtains upper bounds, protects against **all attacks**.

- Verification is hard in general, easy if we can optimize the certificate.

- Adversarial training minimizes lower bound

- Lots of room to go!

6.1.3 Models That Extrapolate

We need inductive bias to do effective extrapolation. Neural Nets sort of do this, as in CNNs (convolution as prior for locality) and Attention-based mechanisms (attention sa prior for sparsity).

New Task. Train is a (review, sentiment) pair. On test, we're going to take a positive review, and see if the model can transform it into a negative review.

So: the *objective* is different between train and test.

Inductive bias (potentially) present in this task: *attribute/style* are localized in text.

Had people evaluate the resulting translated reviews based on (1) grammar, (2) preservation of content, (2) has target attribute.
Q: Can neural networks do logical reasoning? So, given $(x_1, \lor x_2) \land (\neg x_1 \lor x_3)$, can decide SAT $\{0, 1\}$.

Train: fix some distribution of SAT/¬SAT "minimal" pairs (a "hard" distribution of SAT problems). That is, if you change one literal, its satisfiability flips.

Test: Same distribution (for now).

Test accuracy: 88%. Surprisingly high!

Using the activations of the network, can actually decode the satisfying assignment, too. Also extrapolate to other SAT distributions, still works quite well. dnoteVery cool.

Conclusion:

- We need to measure progress empirically in the right way.

- Existing models work because of our attachment to the training/test paradigm. Single successful examples should be taken with a grain of salt.

- If we change our evaluation metrics to be more rigorous, we can get models that learn the right thing.

- *Goodhardt's Law*: When a measure becomes a target, it ceases to be a good measure.

6.2 Game Theory and Human Interaction

The final session I'm attending: Game theory and Human Interaction, starting with Anson's talk!

6.2.1 Ranking Wily People Who Rank Each Other [54]

The speaker is Anson Kahng, joint work with Yasmine Kotturi, Chinmay Kulkarni, Davic Kurokawa, and Ariel D. Proccacia.
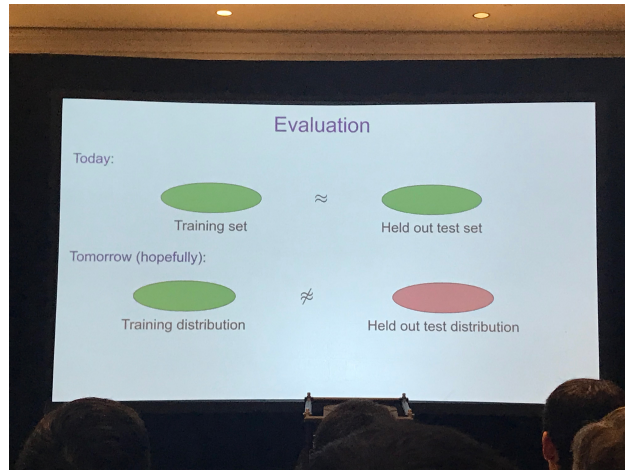
Figure 13: Percy's call for a new evaluation paradigm for ML.

Challenge: in online labor markets (like contracting a painter/technician), hiring is really hard. Often requires domain expertise and lots of time to find an appropriate worker.

Idea: Leverage applicants' expertise to have them evaluate each other. Problem here, though: strategic lying! (you can benefit yourself).

Their solution: impartiality.

Definition 9 (Impartial Rank): *A randomized rank aggregation rule is impartial if, for each voter i, the distribution of her final rank does not depend on her report. Intuitively: "Your report has no bearing on your final rank".*

Model:

- Set of voters the set of alternatives.

- Opinions express as ranking over alternatives.

In general, voting rules are not strategy-proof.
In their setting: the set of voters is exactly the set of alternatives. Assuming self-interested parties. Goal: develop impartial rank aggregation methods. However, impartiality on its own isn't necessarily useful.

New Goal: develop *impartial* and *accurate* rank aggregation methods. Here, "accuracy" means: closely emulates standard rank aggregation rules.

Three measures of accuracy

1. *Forward Error*: everyone close to where they should be,

2. *Backward Error*: everyone is exactly where they could b

3. *Mixed Error*: everyone is close to where they could be.

Their work, three rule-algorithms:

1. *Bipartite*: Randomly split n players into two sets. Have each set rank the players in the other set. Deterministically interleave the two halves.

   No error guarantees, but works in practice.

2. *k-Partite*: Generalized bipartite. Randomly split n players into k sets.

   Backward error guarantees

3. *Committee*: Randomly choose a committee of size k. Place each commitee member near where other committee members.

   Mixed error guarantees. "Most interesting and well behaved of the three."

6.2.2 Adapting a Kidney Exhcange Algorithm to Align with Human Values [35]

The speaker is Rachel Freidman, winner of Best Student Paper Award Honorable Mention, joint work with Jana Schaich Borg, Walter Sinnott-Armstrong, John P. Dickerson, Vincent Conitzer.

Problem: Kidney transplant waiting lists have high demand, low supply, often long waiting lists.

Kidney has a big transplant waiting list. Around 100,000 patients on US cadaver transplant waiting list. 3,000 patients added each month on average.

Solution: Kidney Exchanges [90].
Kidney exchanges allow patients with willing but incompatible live donors to swap donors.

Previous work finds optimal matching [3].

Problem: find the highest cardinality set of disjoint cycles in the donation graph with a maximum cycle length of some constant L. In practice, can formulate and solve as a integer linear program (ILP).

But: which matchines are optimal?

This work: incorporate human values to break ties in a way that aligns with societal values.

Steps to their research:

1. Which patient attributes matter most? Did a Turk study querying about which moral considerations go into this decision. Results: Age, Health (behavioral, like how much they drink), Health (general), were the three most important factors.

2. Follow up study on fictional patient profiles: How do different factors effect their decision of who to prioritize?

Use Bradley-Terry Model:

$$P(i > j) = \frac{p_i}{p_i + p_j},\tag{47}$$

to estimate the inherent "value" of saving each profile.

3. Incorporate these values into the ILP kidney matching algorithm by weighting edges according to the Bradley-Terry model.

Experimental findings: the introduced value-weighting changing assignments quite a lot. The main groups that were effected were O-patient, AB-donor, or both, which were previously under-demanded pairs.

Q: Should we go out and apply this to a real kidney exchange?

A: Not yet! So far we've shown no theoretical obstacles to doing this. Still need several improvements before real world deployment: (1) Need more attributes such as dependents, criminal record, life remaining, (2) Expert Advice, (3) Stakeholder input.

6.2.3 Anchors: High-Precision Model-Agnostic Explanations [70]

The speaker is Marco Tulio Ribeiro, joint work with Sameer Singh and Carlos Guestrin.

Two Trust Challenges

1. Once you train an ML model, is it really working? Does it respect rules?

2. Should I act based on the predictions of the model?

One solution: Treat model as a black box and extract explanations from the model (Model agnostic explanations). Using these explanations we can try to attend to thes two trust challenges.

Two key assessment concepts: Coverage P(trust a model), Precision P(correct | trust a model). This work focuses on Precision.

This work: instead of an explanation/linear model, they offer "anchors". That is:

> **Definition 10** (Anchor): *An anchor is a sufficient condition for the model's prediction.*

Examples:

- Predicting if a load will default or be late. Anchor: $\{FICO < 649\}$, model predicts *bad loan*, same for good but > 699. FICO works for extremes.

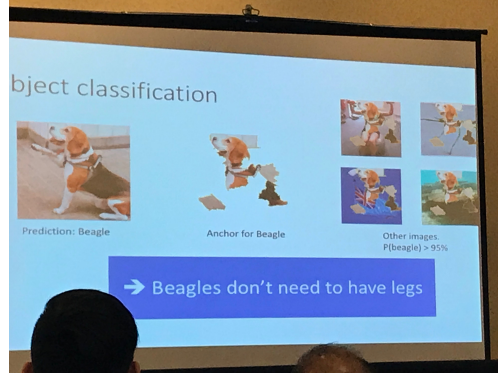- Object classification. Prediction of an image is that there's a beagle in it. "Anchor" for beagle":

Figure 14: Beagle anchors in object classification.

Q: How do we compute anchors?

A: We want Pr(positive prediction | A) > 95% (high precision). Bottom up construction of anchors. Start with {}, evaluate precision. Add a candidate piece of the anchor, recompute precision. Eventually find the highest precision piece, add it to {}, try different additions, keep adding until above 95%. Effectively an exploration problem.

Also did a user study comparing anchors with other explanations, find anchors are informative.

6.2.4 FILE: Predicting Social Status in Signed Networks [117]

The speaker is Xiaoming Li, joint work with Hui Fang and Jie Zhang.

A Signed Social Network (SSN) better reflects real-life relationships by capturing friend/foe or trust/distrust links. Common problem is link prediction in SSNs (that is, guess if links are positive/negative).

Q: Is it true that there are only two types of social status? Not really. They extend SSNs to include "no-relation" status, which exists widely.

Problem formulation: given an SSN, we want to rank all pairs (i, j) with $S_{ij} = 0$ (neutral/unknown assignment), according to the probability of transforming to [a] positive links, [b] maintain no-relation, and [c] becoming a negative link. Focus on ranking, not classification.

FILE: Framework of Integrating Latent and Explicit Features for this problem, centering around the link score function:
$$\mathcal{L}(i, j) = N(u_i^T \cdot v_j) + \sum_k \alpha_w \cdot N(f_{ij}^w). \tag{48}$$

Here, u_i and v_j are features describing each user – they do some feature engineering that incorpo-

rates things like a user's activity, popularity, and so on. Then, they assign links by:

$$\ell_{ij} = \begin{cases} \text{positive} & \mathcal{L}(i,j) > 1 \\ \text{negative} & \mathcal{L}(i,j) < -1 \\ \text{neutral} & \text{otherwise} \end{cases} \tag{49}$$

Dave: And that's a wrap! I need to head out to work on my ICML submissions.

....

References

[1] An agile ethical/legal model for the international and national governance of ai and robotics. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[2] David Abel, Edward C. Willams, Stephen Brawner, Emily Reif, and Michael L. Littman. Bandit-based solar panel control. In *Proceedings of the Thirtieth Conference on Innovative Applications of Artificial Intelligence*, 2018.

[3] David J Abraham, Avrim Blum, and Tuomas Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 295–304. ACM, 2007.

[4] Ittai Abraham, Amos Fiat, Andrew V Goldberg, and Renato F Werneck. Highway dimension, shortest paths, and provably efficient algorithms. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 782–793. Society for Industrial and Applied Mathematics, 2010.

[5] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.

[6] Masataro Asai and Alex Fukunaga. Classical planning in deep latent space: Bridging the subsymbolic-symbolic boundary. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[7] Hassan Ashtiani, Shai Ben-David, and Abbas Mehrabian. Sample-efficient learning of mixtures. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[8] Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. In *AAAI*, pages 1726–1734, 2017.

[9] Holger Bast, Stefan Funke, Domagoj Matijevic, Peter Sanders, and Dominik Schultes. In transit to constant time shortest-path queries in road networks. In *Proceedings of the Meeting on Algorithm Engineering & Expermiments*, pages 46–59. Society for Industrial and Applied Mathematics, 2007.

[10] Marc G Bellemare, Will Dabney, and Rémi Munos. A distributional perspective on reinforcement learning. *arXiv preprint arXiv:1707.06887*, 2017.

[11] Florian Benavent and Bruno Zanuttini. An experimental study of advice in sequential decision-making under uncertainty. In *32nd AAAI Conference on Artificial Intelligence*, 2018.

[12] Avrim Blum and Nika Haghtalab. Algorithms for generalized topic modeling. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[13] Ashok Goel Bobbie Eicher, Lalith Polepeddi. Jill watson doesnt care if youre pregnant: Grounding ai ethics in empirical studies. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[14] Noam Brown and Tuomas Sandholm. Libratus: the superhuman ai for no-limit poker. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.

[15] Robert R Bush and Frederick Mosteller. A stochastic model with applications to learning. *The Annals of Mathematical Statistics*, pages 559–585, 1953.

[16] John C Camillus. Strategy as a wicked problem. *Harvard business review*, 86(5):98, 2008.

[17] Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.

[18] Stephen Cave and Sen S higeartaigh. An ai race for strategic advantage: Rhetoric and risks. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[19] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582, 2015.

[20] Shih-Fen Cheng and Thi Duong Nguyen. Taxisim: A multiagent simulation platform for evaluating taxi fleet operations. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, volume 2, pages 14–21. IEEE, 2011.

[21] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163, 2017.

[22] Gianfranco Ciardo and Radu Siminiceanu. Using edge-valued decision diagrams for symbolic generation of shortest paths. In *International Conference on Formal Methods in Computer-Aided Design*, pages 256–273. Springer, 2002.

[23] Juan D Correa, Jin Tian, and Elias Bareinboim. Generalized adjustment under confounding and selection biases. 2018.

[24] Scott Niekum Daniel Brown. Efficient probabilistic performance bounds for inverse reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[25] Kristopher De Asis, J Fernando Hernandez-Garcia, G Zacharias Holland, and Richard S Sutton. Multi-step reinforcement learning: A unifying algorithm. *arXiv preprint arXiv:1703.01327*, 2017.

[26] Kristopher De Asis, J Fernando Hernandez-Garcia, G Zacharias Holland, and Richard S Sutton. Multi-step reinforcement learning: A unifying algorithm. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[27] Peter F Drucker. Pension fund socialism. *The Public Interest*, (42):3, 1976.

[28] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226. ACM, 2012.

[29] Daniel Estrada. Value alignment, fair play, and the rights of service robots. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[30] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Dawn Song, Tadayoshi Kohno, Amir Rahmati, Atul Prakash, and Florian Tramer. Note on attacking object detectors with adversarial stickers. *arXiv preprint arXiv:1712.08062*, 2017.

[31] Yueh-Hua Wu Fan-Yun Sun, Yen-Yu Chang and Shou-De Lin. Designing non-greedy reinforcement learning agents with diminishing reward shaping. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[32] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. *arXiv preprint arXiv:1705.08926*, 2017.

[33] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[34] Meire Fortunato, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Ian Osband, Alex Graves, Vlad Mnih, Remi Munos, Demis Hassabis, Olivier Pietquin, et al. Noisy networks for exploration. *arXiv preprint arXiv:1706.10295*, 2017.

[35] Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John P Dickerson, and Vincent Conitzer. Adapting a kidney exchange algorithm to align with human values. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[36] Nathan Fulton and André Platzer. Safe reinforcement learning via formal methods. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[37] Stefan Funke and Sabine Storandt. Path shapes: an alternative method for map matching and fully autonomous self-localization. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 319–328. ACM, 2011.

[38] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[39] Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. When will ai exceed human performance? evidence from ai experts. *arXiv preprint arXiv:1705.08807*, 2017.

[40] Barbara J Grosz. Collaborative systems (aaai-94 presidential address). *AI magazine*, 17(2): 67, 1996.

[41] Ross Gruetzemacher. Rethinking ai strategy and policy as entangled super wicked problems. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[42] Jean Harb, Pierre-Luc Bacon, Martin Klissarov, and Doina Precup. When waiting is not an option: Learning options with a deliberation cost. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[43] Anna Harutyunyan, Peter Vrancx, Pierre-Luc Bacon, Doina Precup, and Ann Nowe. Learning with options that terminate off-policy. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[44] Hado V Hasselt. Double q-learning. In *Advances in Neural Information Processing Systems*, pages 2613–2621, 2010.

[45] Brian W Head et al. Wicked problems in public policy. *Public policy*, 3(2):101, 2008.

[46] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, 2015.

[47] Matteo Hessel, Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. Rainbow: Combining improvements in deep reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[48] John Hooker and Tae Wan Kim. Toward non-intuition-based machine ethics. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[49] Richard Hoshino and Jeneva Beairsto. Optimal pricing for distance-based transit fares. In *Proceedings of the Thirtieth Conference on Innovative Applications of Artificial Intelligence*, 2018.

[50] Justin Huang and Maya Cakmak. Supporting mental model accuracy in trigger-action programming. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 215–225. ACM, 2015.

[51] Charles Isbell, Christian R Shelton, Michael Kearns, Satinder Singh, and Peter Stone. A social reinforcement learning agent. In *Proceedings of the fifth international conference on Autonomous agents*, pages 377–384. ACM, 2001.

[52] Nan Jiang. Pac reinforcement learning with an imperfect model. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[53] Sabine Storandt Johannes Blum, Stefan Funke. Sublinear search spaces for shortest path planning in grid and road networks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[54] Anson Kahng, Yasmine Kotturi, Chinmay Kulkarni, David Kurokawa, and Ariel D Procaccia. Ranking wily people who rank each other. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[55] Richard Kaye. Minesweeper is np-complete. *The Mathematical Intelligencer*, 22(2):9–15, 2000.

[56] Richard Kim, Max Kleiman-Weiner, Andres Abeliuk, Edmond Awad, Sohan Dsouza, Josh Tenenbaum, and Iyad Rahwan. A computational model of commonsense moral decision making. *arXiv preprint arXiv:1801.04346*, 2018.

[57] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.

[58] W Bradley Knox and Peter Stone. Framing reinforcement learning from human reward: Reward positivity, temporal discounting, episodicity, and performance. *Artificial Intelligence*, 225:24–50, 2015.

[59] J Zico Kolter and Eric Wong. Provable defenses against adversarial examples via the convex outer adversarial polytope. *arXiv preprint arXiv:1711.00851*, 2017.

[60] TK Satish Kumar, Zhi Wang, Anoop Kumar Craig Milo Rogers Craig, and A Knoblock. Load scheduling of simple temporal networks under dynamic resource pricing. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[61] Emily Larosa and David Danks. Impacts on trust of healthcare ai. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[62] Adam Lerer and Alexander Peysakhovich. Maintaining cooperation in complex social dilemmas using deep reinforcement learning. *arXiv preprint arXiv:1707.01068*, 2017.

[63] Kelly Levin, Benjamin Cashore, Steven Bernstein, and Graeme Auld. Playing it forward: Path dependency, progressive incrementalism, and the super wickedproblem of global climate change. In *International Studies Association 48th Annual Convention. Chicago, February.* Citeseer, 2007.

[64] Siyuan Li and Chongjie Zhang. An optimal online method of selecting source policies for reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[65] Fang Liu, Joohyun Lee, and Ness Shroff. A change-detection based framework for piecewise-stationary multi-armed bandit problem. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[66] Robert Tyler Loftin, James MacGlashan, Bei Peng, Matthew E Taylor, Michael L Littman, Jeff Huang, and David L Roberts. A strategy-aware technique for learning behaviors from discrete human feedback. In *AAAI*, pages 937–943, 2014.

[67] Tania Lombrozo. The structure and function of explanations. *Trends in cognitive sciences*, 10(10):464–470, 2006.

[68] Alex John London and David Danks. Regulating autonomous vehicles: A policy proposal. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[69] Matthijs M. Maas. Regulating for normal ai accidents operational lessons for the responsible governance of ai deployment. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[70] Carlos Guestrin Marco Tulio Ribeiro, Sameer Singh. Anchors: High-precision model-agnostic explanations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[71] Ofir Marom and Benjamin Rosman. Belief reward shaping in reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[72] Vincent Conitzer Max Kramer, Jana Schaich Borg and Walter Sinnott-Armstrong. When do people want ai to make decisions? In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[73] Drew McDermott. Artificial intelligence meets natural stupidity. *ACM SIGART Bulletin*, (57):4–9, 1976.

[74] Drew McDermott, Malik Ghallab, Adele Howe, Craig Knoblock, Ashwin Ram, Manuela Veloso, Daniel Weld, and David Wilkins. Pddl-the planning domain definition language. 1998.

[75] Erez Karpas Michael Katz*, Dany Moshkovich. Semi-black box: Rapid development of planning based solutions. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[76] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*, pages 1928–1937, 2016.

[77] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

[78] Andrew Y Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *ICML*, volume 99, pages 278–287, 1999.

[79] Thanh Nguyen, Raymond KW Wong, and Chinmay Hegde. A provable approach for double-sparse coding. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[80] Judy Goldsmith. Olivia J. Erdelyi. Regulating artificial intelligence proposal for a global solution. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[81] Supratik Paul, Konstantinos Chatzilygeroudis, Kamil Ciosek, Jean-Baptiste Mouret, Michael Osborne, Shimon Whiteson, Serena Ivaldi, and Barkan Ugurlu. Alternating optimisation and quadrature for robust control. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[82] Judea Pearl. *Causality*. Cambridge university press, 2009.

[83] Nicolas Angelard-Gontier-Nan Rosemary Ke Genevieve Fried Ryan Lowe Joelle Pineau Peter Henderson, Koustuv Sinha. Ethical challenges in data-driven dialogue systems. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[84] Huao Li-Michael Lewis Ramitha Sundar Rahul Iyer, Yuezhang Li and Katia Sycara. Transparency and explanation in deep reinforcement learning neural networks.

[85] Iyad Rahwan, Jean-Francois Bonnefon, Azim Shariff, et al. Moral machine. *Scalable Cooperation, MIT Media Lab. At: http://moralmachine. mit. edu (accessed 21 August 2016)*, 2016.

[86] Miquel Ramırez and Hector Geffner. Probabilistic plan recognition using off-the-shelf classical planners. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence (AAAI 2010)*, pages 1121–1126, 2010.

[87] Andres Abeliuk Edmond Awad Sohan Dsouza Josh Tenenbaum Richard Kim, Max Kleiman-Weiner and Iyad Rahwan. A computational model of commonsense moral decision making. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[88] Horst WJ Rittel and Melvin M Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973.

[89] Benedict Wright Bernhard Nebel Robert Mattmueller, Florian Geier. On the relationship between state-dependent action costs and conditional effects in planning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[90] Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488, 2004.

[91] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015.

[92] Michael Scheessele. A framework for grounding the moral status of intelligent machines. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[93] Abdallah Saffidine Kevin Tran Serge Gaspers, Stefan Rmmele. Minesweeper with limited moves. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[94] Zina Ben-Miled Rebecca Shafer Setu Shah, Mahmood Hosseini and Steve Berube. A water demand prediction model for central indiana. In *Proceedings of the Thirtieth Conference on Innovative Applications of Artificial Intelligence*, 2018.

[95] Arjun Sharma and Kris M Kitani. Phase-parametric policies for reinforcement learning in cyclic environments. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[96] Guni Sharon, Michael Albert, Tarun Rambha, Stephen Boyles, and Peter Stone. Traffic optimization for a mixture of self-interested and compliant agents. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[97] Meghna Lowalekar Nicholas Wong Wai Hin Rishikeshan Rajendram Tran Trong Khiem Pradeep Varakantham Truong Trong Nghia Shashi Shekhar Jha, Shih-Fen Cheng and Firmansyah Rahman. Upping the game of taxi driving in the age of uber. In *Proceedings of the Thirtieth Conference on Innovative Applications of Artificial Intelligence*, 2018.

[98] Heinz Koeppl Sikun Yang. A poisson gamma probabilistic model for latent node-group memberships in dynamic networks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[99] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529 (7587):484–489, 2016.

[100] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*, 2017.

[101] Herbert A Simon. Theories of bounded rationality. *Decision and organization*, 1(1):161–176, 1972.

[102] Shirin Sohrabi, Anton V Riabov, Michael Katz, and Octavian Udrea. An ai planning solution to scenario generation for enterprise risk management. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[103] Kaushik Subramanian, Charles Isbell, and Andrea Thomaz. Learning options through human interaction. In *2011 IJCAI Workshop on Agents Learning Interactively from Human Teachers (ALIHT)*. Citeseer, 2011.

[104] Richard S Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

[105] Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112 (1-2):181–211, 1999.

[106] Luming Tang, Yexiang Xue, Di Chen, and Carla P Gomes. Multi-entity dependence learning with rich context via conditional variational auto-encoder. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[107] Arash Tavakoli, Fabio Pardo, and Petar Kormushev. Action branching architectures for deep reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[108] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[109] Dieter Vanderelst and Alan Winfield. The dark side of ethical robots. In *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics Society*, 2018.

[110] Chen Wang, Le Zhang, Lihua Xie, and Junsong Yuan. Kernel cross-correlator. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[111] Yuanhong Wang, Yuyi Wang, Xingwu Liu, and Juhua Pu. On the erm principle with networked data. *arXiv preprint arXiv:1711.04297*, 2017.

[112] Yuanhong Wang, Yuyi Wang, Xingwu Liu, and Juhua Pu. On the erm principle with networked data. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[113] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Van Hasselt, Marc Lanctot, and Nando De Freitas. Dueling network architectures for deep reinforcement learning. *arXiv preprint arXiv:1511.06581*, 2015.

[114] Jie Zhang Hui Xi Wen Song, Donghun Kang. Risk-aware proactive scheduling via conditional value-at-risk. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[115] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. In *Reinforcement Learning*, pages 5–32. Springer, 1992.

[116] Chenjun Xiao, Jincheng Mei, and Martin Muller. Memory-augmented monte carlo tree search. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[117] Jie Zhang Xiaoming Li, Hui Fang. File: A novel framework for predicting social status in signed networks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[118] Bo An Xinrun Wang, Martin Strobel, and Fookwai Kong. Catching captain jack: Efficient time and space dependent patrols to combat oil-siphoning in international waters. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[119] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. Learning fair representations. In *International Conference on Machine Learning*, pages 325–333, 2013.