Decision Theory II: Course Outline

What we think we're going to do this semester:

- a closer look at decision rules that don't assume probability
- dynamic decision making
- decision making in a group setting

The details are still a bit fuzzy.

- (We're making it up as we go along.)

We're organizing a decision theory workshop on April 4-5 on these issues.

Representations of Uncertainty

Goal: to find (and characterize) reasonable decision rule that deal with the Ellsberg paradox.
We've already seen one: a set $\mathcal{P}$ of probabilities. Recall that

$$\underline{E}_{\mathcal{P}}(u_a) = \inf_{\Pr \in \mathcal{P}} \{E_{\Pr}(u_a) : \Pr \in \mathcal{P}\}.$$

Thus, we get the rule MMEU (Maxmin Expected Utility):

$$a_1 \leq a_2 \text{ if } \underline{E}_{\mathcal{P}}(u_{a_1}) \leq \underline{E}_{\mathcal{P}}(u_{a_2}).$$

MMEU generalizes maximin (if $\mathcal{P}$ consists of all probability measures) and expected utility (if $\mathcal{P}$ consists of just one probability measure).

Characterizing EU

Recall the Anscombe-Aumann framework:

- the objects of choice are horse lotteries.
  - functions from state space S (assume finite) to simple probability distributions (i.e. distributions with finite support) over Z (prizes)

Here were the axioms that characterized expected utility maximization:

A1. $\succ$ is a preference relation on H (horse lotteries)

A2. (Continuity:) If $f \succ g \succ h$, then there exist $\alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ g \succ \beta f + (1-\beta)h$.

A3. (Independence:) If $f \succ g$, then for all h and $\alpha \in (0,1]$, $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$.

If $X \subseteq S$, let fXg be the act that agrees with f on X and with g on X^c (the complement of X).

A4. (Monotonicity:) If p and q are probabilities on prizes such that $f(S - \{s\})p \succ f(S - \{s\})q$, then $f(S - \{s'\})p \succ f(S - \{s'\})q$ for all nonnull $s' \in S$.

A5. (Nondegeneracy:) There exist f and g such that $f \succ g$.

Key result:

Theorem: (Anscombe-Aumann) If A1–A5 hold, then there exist a utility u on prizes and a probability Pr on states such that $\succ$ can be represented by expected utility.

- Can associate with each horse lottery h a random variable u_h:
  - $u_h(s)$ is the expected utility of the lottery $h(s)$ on prizes (i.e., $u_h(s) = \Sigma_{z \in Z}\, h(s)(z)u(z)$)
- $f \succ g$ iff $E_{\Pr}(u_f) > E_{\Pr}(u_g)$.

Moreover, Pr is unique and u is unique up to affine transformations.

Claim: A1, A2, and A4 hold for MMEU (homework), but A3 fails.

A3. (Independence:) If $f \succ g$, then for all h and $\alpha \in (0,1]$, $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$.

Example: Suppose that

- $S = \{s_1, s_2\}$
- $\mathcal{P} = \{\Pr_1, \Pr_2\}$; $\Pr_1(s_1) = 1/3$, $\Pr_2(s_1) = 2/3$
- $f = (4.2, 4.2)$ (i.e. $f(s_1) = 4.2$; $f(s_2) = 4.2$), $g = (6,3)$, $h = (3,6)$.
- $\underline{E}(f) = 4.2$ and $\underline{E}(g) = 4$, so $f \succ g$.
- $f/2 + h/2 = (3.6, 5.1)$; $g/2 + h/2 = (4.5, 4.5)$
- $\underline{E}(f/2 + h/2) = 4.1$ and $\underline{E}(g/2 + h/2) = 4.5$, so $g/2 + h/2 \succ f/2 + h/2$.

Characterizing MMEU

[Gilboa and Schmeidler:] Independence doesn't hold; we replace it by:

A3′. (Certainty-Independence:) If $f \succ g$, h is a constant function, and $\alpha \in (0, 1]$, then $\alpha f + (1 - \alpha)h \succ \alpha g + (1 - \alpha)h$.

- A3 just says "if ... then"; "iff" follows from other axioms.

Instead of A4, GS use:

A4′. (Monotonicity:) If $f(s) \succeq g(s)$ for all $s \in S$, then $f \succeq g$.

- This doesn't quite mean that f beats g at every state. Think of $f(s)$ as the constant horse lottery that returns $f(s)$ at every state. It means the constant $f(s)$ beats the constant $g(s)$.

It turns out that in the context of A1, A2, and A3, A4 and A4′ are equivalent.

One more property is needed:

A6. (Uncertainty Aversion:) If $\alpha \in (0,1)$ and $f \approx g$, then $\alpha f + (1-\alpha)g \succeq f$.

- For EU, A6 holds with $\approx$ (follows from A1–A3).
- Can have $\alpha f + (1 - \alpha g) \succ f$ with MMEU
  - Consider previous example: $g = (6,3)$, $h = (3,6)$. Then $g \approx h$, but $g/2 + h/2 \succ g$
- A6 models hedging.

Theorem: (Gilboa-Schmeidler) If A1, A2, A3′, A4′, A5, and A6 hold, then there exist a utility u on prizes and a closed convex set $\mathcal{P}$ of probability measures on states such that $\succ$ can be represented by MMEU.

- $f \succ g$ iff $\underline{E}_{\mathcal{P}}(u_f) > \underline{E}_{\mathcal{P}}(u_g)$

Moreover, $\mathcal{P}$ is unique and u is unique up to affine transformations.

- All you really need are the extreme points in $\mathcal{P}$; requiring that $\mathcal{P}$ be closed and convex makes it unique.

Other Representations of Uncertainty

Why is probability the “right” way to represent uncertainty?

- It’s not so good at representing ignorance.
- or extremely unlikely events.

Many alternatives considered in the literature:

- sets of probabilities
- non-additive probabilities
- belief functions
- lexicographic probabilities
- possibility measures
- ranking functions
- plausibility measures
- …

Some of these approaches are closely related. We'll focus on sets of probabilities, non-additive probabilities, and belief functions.

- If want more, take CS 677!

Non-additive probabilities

A *non-additive probability* [Choquet, Schmeidler] ν on S is a function mapping subsets of S to $[0, 1]$ such that

N1. $\nu(\emptyset) = 0$

N2. $\nu(S) = 1$

N3. If $E \subseteq F$, then $\nu(E) \leq \nu(F)$.

These constraints are pretty minimal. For example, suppose $S = \{s_1, s_2\}$ and

- $\nu_\alpha(\emptyset) = 0$
- $\nu_\alpha(s_1) = \nu_\alpha(s_2) = \alpha$
- $\nu(S) = 1$.

Then ν_α is a nonadditive probability for each $\alpha \in [0, 1]$.

We may want more constraints ...

Expectation with respect to a nonadditive probability

Suppose that f is a random variable with finite range.

- Suppose that the values of f are $x_1 < \ldots < x_n$.

Then the expectation of f with respect to ν is defined as follows [Choquet]:

$$E_\nu(f) = x_1 + (x_2 - x_1)\nu(f > x_1) + \cdots + (x_n - x_{n-1})\nu(f > x_{n-1}).$$

Why is this the right definition of expectation?

- Some good news: it coincides with the standard definition if ν is a probability measure.

But why not use the more obvious generalization of probabilistic expectation?

$$E'_\nu(f) = \sum_{s \in S} \nu(s) f(s)$$

Stay tuned ...

Nonadditive Expected Utility

Nonadditive expected utility rule:

- Given a utility function u on prizes and a nonadditive probability ν on states, then

$$f \succ g \text{ iff } E_\nu(u_f) > E_\nu(u_g)$$

Comonotonic Independence

Acts f and g are *comonotonic* if for all states s and t:

$$f(s) \succ f(t) \text{ iff } g(s) \succ g(t)$$

- f and g are comonotonic, if you're happier to be in state s than state t when doing f, then you're happier to be in state s than state t when doing g (and vice versa).
- If h is a constant act, then f and h are comonotonic for all acts f (since we never have $h(s) \succ h(t)$).

A3$''$. (Comonotonic Independence:) If f and h and g and h are both comonotonic and $f \succ g$, then for all $\alpha \in (0, 1]$, $\alpha f + (1 - \alpha)h \succ \alpha g + (1 - \alpha)h$.

Idea: comonotonic independence tries to avoid the kind of application of independence that gives Ellsberg's paradox.

- Note: A3$''$ is stronger than A3$'$.

Representation Theorem

Theorem: (Schmeidler) If A1, A2, A3$''$, A4$'$, and A5 hold, then there exist a utility u on prizes and a nonadditive probability ν on states such that $\succ$ can be represented by NEU.

- $f \succ g$ iff $E_\nu(u_f) > E_\nu(u_g)$

Moreover, ν is unique and u is unique up to affine transformations.

- Moving from additive probability to nonadditive probability results in weakening independence to comonotonic independence.

Nonadditive Probability and Sets of Probabilities

Given a set $\mathcal{P}$ of probabilities, define $\mathcal{P}_*$ to be the lower probability of $\mathcal{P}$ and $\mathcal{P}^*$ to be the upper probability:

$$\mathcal{P}_*(X) = \inf_{\Pr \in \mathcal{P}} \Pr(X)$$
$$\mathcal{P}^*(X) = \sup_{\Pr \in \mathcal{P}} \Pr(X).$$

- $\mathcal{P}_*$ and $\mathcal{P}^*$ are both nonadditive probabilities; moreover
$$\mathcal{P}^*(A) = 1 - \mathcal{P}_*(A^c)$$

Is every nonadditive probability $\mathcal{P}_*$ (or $\mathcal{P}^*$) for some set $\mathcal{P}$ of probabilities?

- Simple counterexample: Let $S = \{s_1, s_2\}$.
  If $\nu_1(s_1) = 2/3$, $\nu_1(s_2) = 2/3$, then $\nu_1 \neq \mathcal{P}_*$.
  If $\nu_2(s_1) = 1/3$, $\nu_2(s_2) = 1/3$, then $\nu_2 \neq \mathcal{P}^*$.

Some properties of $\mathcal{P}_*$ and $\mathcal{P}^*$:

$$\begin{aligned} \mathcal{P}^*(A) + \mathcal{P}^*(B) &\geq \mathcal{P}^*(A \cup B) \text{ if } A \cap B = \emptyset \\ \mathcal{P}_*(A) + \mathcal{P}_*(B) &\leq \mathcal{P}_*(A \cup B) \text{ if } A \cap B = \emptyset \\ \mathcal{P}_*(A) + \mathcal{P}_*(B) &\leq \mathcal{P}_*(A \cap B) + \mathcal{P}^*(A \cup B) \\ &\leq \mathcal{P}^*(A) + \mathcal{P}^*(B) \end{aligned}$$

Any nonadditive probability that is of the form $\mathcal{P}_*$ must satisfy these properties. But these properties do *not* characterize $\mathcal{P}_*/\mathcal{P}^*$. The following *covering property* does:

A set $\mathcal{A}$ of subsets of S *covers* $A \subseteq S$ *exactly* k *times* if every element of A is in exactly k sets in $\mathcal{A}$.

If $\mathcal{A} = \{A_1, \ldots, A_k\}$ covers A exactly $m+n$ times and A^c exactly m times, then $\Sigma_{i=1}^k \nu(A_i) \leq m + n\nu(A)$.

Theorem: (Anger-Lembcke) ν satisfies the covering property iff $\nu = \mathcal{P}_*$ for some set $\mathcal{P}$ of probability measures. In fact, can take $\mathcal{P} = \mathcal{P}_\nu$, where

$$\mathcal{P}_\nu = \{\Pr : \Pr(A) \geq \nu(A) \text{ for all } A \subseteq S\}.$$

[Get $\mathcal{P}^*$ by replacing $\leq$ by $\geq$ in covering property.]

Convexity

A nonadditive probability ν is *convex* if

$$\nu(A) + \nu(B) \leq \nu(A \cup B) + \nu(A \cap B).$$

- $\mathcal{P}_*$ is not necessarily convex (neither is $\mathcal{P}^*$).
  - Let $S = \{s_1, s_2, s_3, s_4\}$, $\mathcal{P} = \{\Pr_1, \Pr_2\}$
  - $\Pr_1(s_1) = \Pr_1(s_3) = 1/2$
  - $\Pr_2(s_2) = \Pr_2(s_4) = 1/2$
  - Then $\mathcal{P}_*(\{s_1, s_2\}) = \mathcal{P}_*(\{s_2, s_3\}) = \mathcal{P}_*(\{s_1, s_2, s_3\}) = 1/2$; $\mathcal{P}_*(s_2) = 0$
  - $\mathcal{P}_*(\{s_1, s_2\}) + \mathcal{P}_*(\{s_2, s_3\}) > \mathcal{P}_*(s_2) + \mathcal{P}_*(\{s_1, s_2, s_3\})$

Proposition: Convexity implies the covering property. (So if ν is convex, then $\nu = (\mathcal{P}_\nu)_*$.)

Theorem: (Schmeidler) If ν is convex, then $E_\nu(f) = \underline{E}_{\mathcal{P}_\nu}(f)$: the two definitions of expectation coincide!

Belief Functions

(Dempster-Shafer) *belief functions* (also known as *infinitely monotone Choquet capacities*) are a well-studied notion of convex nonadditive probabilities.

A *belief function* Bel on S is a function $\mathrm{Bel} : S \rightarrow [0, 1]$ satisfying the following three properties:

B1. $\mathrm{Bel}(\emptyset) = 0$.

B2. $\mathrm{Bel}(S) = 1$.

B3. $\mathrm{Bel}(\cup_{i=1}^{n} A_i) \geq \Sigma_{i=1}^{n} \Sigma_{\{I \subseteq \{1,\ldots,n\}:|I|=i\}} (-1)^{i+1} \mathrm{Bel}(\cap_{j \in I} A_j)$, for $n = 1, 2, 3, \ldots$.

B3 seems like a strange property. But remember the *inclusion-exclusion property* of probability:

- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\mathcal{P}(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(A \cap B) + \Pr(A \cap B \cap C)$
- $\Pr(A \cup B \cup C \cup D) = \ldots$

B3 is just inclusion-exclusion with = replaced by $\geq$.

- B3 implies convexity (with $n = 2$)
- Define $\text{Plaus}(A) = 1 - \text{Bel}(A^c)$
  - By B3, $\text{Bel}(A) \leq \text{Plaus}(A)$.
  - $\text{Bel}(A) = (\mathcal{P}_{\text{Bel}})_*(A)$; $\text{Plaus} = \mathcal{P}^*_{\text{Bel}}(A)$
  - $\text{Bel}(A)$ and $\text{Plaus}(A)$ can be viewed as lower and upper bounds on the likelihood of A.
  - If Bel is a probability measure, then $\text{Bel}(A) = \text{Plaus}(A)$ for all A.

Mass Functions

A *mass function* on S is a function $m : 2^S \rightarrow [0,1]$ satisfying:

M1. $m(\emptyset) = 0$.

M2. $\Sigma_{A \subseteq S} m(A) = 1$.

Intuition:

- $m(A)$ describes the extent to which the evidence supports A.
- Another way to think about it: Consider an agent making observations.
    - Can identify an observation with a subset of S (the states where the observation could have been made).
    - $m(A)$ is the probability of observing A.
    - Must have $m(\emptyset) = 0$ (can't observe the empty set) and $\Sigma_{A \subseteq S} m(A) = 1$ (you have to observe something).

Mass Functions and Belief Functions

Given a mass function m, define the belief function based on m, Bel_m, by taking

$$\mathrm{Bel}_m(A) = \Sigma_{\{A' : A' \subseteq A\}} m(A').$$

- Intuition: $\mathrm{Bel}_m(A)$ is the sum of the probabilities of the evidence or observations that guarantee that the actual world is in A.

Define

$$\mathrm{Plaus}_m = \Sigma_{\{A' : A' \cap A \neq \emptyset\}} m(A').$$

Theorem: Bel_m is a belief function and Plaus_m is the corresponding plausibility function (i.e., $\mathrm{Plaus}(A) = 1 - \mathrm{Plaus}(A^c)$).

The converse is also true in finite spaces:

Theorem: Given a belief function Bel on W, there is a unique mass function m on W such that $\mathrm{Bel} = \mathrm{Bel}_m$.

Example

Suppose that $S = \{s_1, s_2, s_3\}$. Define m as follows:

- $m(s_1) = 1/4$
- $m(\{s_1, s_2\}) = 1/4$
- $m(\{s_2, s_3\}) = 1/2$
- mass of all other sets is 0

Then

- $\text{Bel}_m(s_1) = 1/4$; $\text{Bel}_m(s_2) = \text{Bel}_m(s_3) = 0$; $\text{Bel}_m(\{s_1, s_2\}) = 1/2$; $\text{Bel}_m(\{s_2, s_3\}) = 1/2$; $\text{Bel}_m(\{s_1, s_3\}) = 1/4$; $\text{Bel}_m(\{s_1, s_2, s_3\}) = 1$;
- $\text{Plaus}_m(s_1) = 1/2$; $\text{Plaus}_m(s_2) = 3/4$; $\text{Plaus}_m(s_3) = 1/2$; $\text{Plaus}_m(\{s_1, s_2\}) = 1$; $\text{Plaus}_m(\{s_2, s_3\}) = 3/4$; $\text{Plaus}_m(\{s_1, s_3\}) = 1$; $\text{Plaus}_m(\{s_1, s_2, s_3\}) = 1$.

One way to understand belief functions is in terms of sets of probabilities. Others (like Shafer) view it as a different way of representing uncertainty, and do not want to reduce it to probability.

- Shafer views belief functions as a representation of *evidence*.
  - This leads to different notions of conditioning, but that's beyond the scope of this course ...

Open question?

What axioms characterize decision making with respect to a belief function (using Choquet expectation):

- Must have at least A1, A2, A3$''$, A4$'$, and A5
  - these hold for any nonadditive probability

  and A6
  - This holds for lower expectation with a set of probabilities, and therefore for convex nonadditive probabilities.
- These axioms characterize convex nonadditive probabilities.
- What extra properties are needed for belief functions?

References

- Axiomatizing MMEU
  I. Gilboa and D. Schmeidler, Maxmin Expected Utility with a Non-unique Prior, *Journal of Mathematical Economics* 18, 1989, pp. 141–153.

- Axiomatizing nonadditive probability:
  D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* 57, 1989, pp. 571–587.

- Nonadditive probability and MMEU:
  D. Schmeidler, Integral Representation without Additivity, *Proc. of the Amer. Math. Soc.* 97:2, 1986, pp. 255-261.

- General results on uncertainty
  - belief functions and other representations of uncertainty
  - relating nonadditive probability to lower probability
  - updating representations of uncertainty
  - general notions of expectation

  J. Y. Halpern, *Reasoning About Uncertainty*, MIT Press, 2003 (out in June, I hope)
- Case-based decision theory (coming next):
  I. Gilboa and D. Schmeidler, *A Theory of Case-Based Decisions*, Cambridge University Press, 2001.

Applying Expected Utility: Conceptual Problems

To apply expected utility maximization (or any of its variants), you need to have a clear notion of what the states are, and what the outcomes are.

- This is not always the case

Example 1: From Savage's book:

- You want to make a 6-egg omelet, and have already cracked 5 eggs into the bowl. You suspect the 6th egg might be bad. Do you
  - crack the sixth egg into the bowl, or
  - crack it into a separate dish

  In this case, the states and outcomes are quite clear. It seems pretty straightforward to apply EU.

Example 2: You're trying to choose between three nannies. The set of acts is clear, but what are the possible states?

- nanny may smoke
- nanny may have a boyfriend she stays out late at night with, so is late in the morning
- nanny may be a serious jogger who, instead of taking the kids on a leisurely stroll, jogs at top speed with their stroller

There are many unforeseen contingencies.

Example 3: President Bush is trying to decide whether to attack Iraq. The set of possible acts is not so clear, and the set of possible states is even less clear.

- Whatever the set of states is, it's huge.
- Again, many unforeseen contingencies ...

There has been a lot of work on unforeseen contingencies.

- Certainly nothing definitive ...

Case-Based Decision Theory

It's clear that people don't use utility maximization for complex problems where the state space (and set of actions) is large and unclear.

More typically, they look at similar scenarios and see what worked/didn't work:

- Get letters of reference for nannies.
- Consult with friends who have hired nannies.
- Look at effects of attacking/not attacking in other scenarios:
  - Vietnam
  - Iraq I
  - Bosnia
  - ...

Make your new decision based on past history of cases.

A Formal Model

P: set of problems

- Choose a nanny, decide what to do about Iraq, ...

A: set of actions

- Hire nanny 1, hire nanny 2, bomb Iraq, ...

R: a set of results, or outcomes

- The nanny was great, pretty good, a disaster, ...

$C = P \times A \times R$ is the set of *cases*.

- A case describes a problem, an action, and a result.
  - (p, a, r): r is the result of having taking action a in problem p.

A and R are as in the EU approach.

- No analogue to states.
- Emphasis on P is new

Memory and Similarity

The decision maker is assumed to have a *memory* $M \subseteq C$.

- Intuitively, these are the cases that you remember, or have access to.
  - These may include cases that you read about, or hear about from others.

Assume that there is a similarity metric

$$s : (P \times A) \times (P \times A) \rightarrow [0, 1]$$

- How similar is (p, a) to (p', a')?
- 0 means "completely unalike; 1 means identical.
- Seems reasonable to assume that s is symmetric.

Just as in the Savage approach, also assume a real-valued utility function u on outcomes.

A Case-Based Decision Rule

Given a decision problem p, the "utility" of act a is

$$U(a) = \Sigma_{(q,a',r)\in M} s((q,a'),(p,a))u(r).$$

- Consider all the utility of all the previous that you're aware of, weighted according to how close they are to the current option.
- Instead of multiplying utility by probability, multiply it by similarity.
- By convention, the empty sum (which is what happens if M is empty) is 0.

Choose act a with highest utility.

Aspirations

In standard utility maximization, shifting a utility function by adding a constant to all values does not change the decision.

In case-based decision theory, shifting by a constant can make a big difference.

- Think of 0 utility as indicating the decision maker's *aspiration level.*
  - Above 0 is OK, below it is not
- Clearly with this interpretation, shifting everything up/down by a constant changes things drastically.

Changing Aspirations

Suppose we start with an empty memory:

- All acts are equal – have "utility" 0
- So an act is chosen at random
- If the act has positive utility, it will continue to be chosen, perhaps even if a better act is discovered.
  - Satisficing behavior [Simon]
  - why fix it if it ain't broke
- This reluctance to change behaviors is observable empirically.
- If aspiration levels go up, all acts are downgraded
  - Acts that were acceptable may no longer be so.
  - There is more motivation to try new acts
- If aspiration levels go down, there is more of a preference for the tried and true.

- It can be shown that if aspiration levels are adjusted appropriately over time, and the same situations are encountered repeatedly, then the decision maker eventually start choosing the optimal act (in the expected utility sense).

CBDT vs. EUT

Both expected utility theory (EUT) and case-based decision theory (CBDT) make decision by associating a weighted sum of utilities to acts. But ... there are major conceptual differences between them:

- In EUT, you have to construct all the states and guess the outcome of performing act a in state s
  - lots of counterfactual reasoning
  - if I had done a, r would have happened.
- In CBDT, you consider only cases that you know about; no counterfactuals involved.
  - All that the decision maker knows about the world is what she has observed

EUT seems most appropriate when the states and the outcomes of acts are clear.

CBDT seems most appropriate in complex, novel decision problems.

Reducing one approach to the other

- Every preference order induced by EUT can be represented by CBDT, by choosing the appropriate memory.
- Every preference order induced by CBDT can be represented by EUT, by choosing the right state space.

CBDT and EUT are best thought of as different conceptual frameworks.

Similarity vs. Probability

Where is the similarity function coming from?

- The same place as the probability function (at least, in Savage's approach)
- Both represent subjective judgments

But ...

- With probability there is a frequency interpretation, that people have some comfort with.
  - i.e., there is an "objective" correlate to the subjective judgment
- The posteriors of people with different priors will converge, given enough evidence

As yet, there are no analogous results for similarity.

On the other hand:

- [Gilboa-Schmeidler] Similarity is more basic
  - Frequency judgments depend on a notion of similarity

References

- Axiomatizing MMEU
  I. Gilboa and D. Schmeidler, Maxmin Expected Utility with a Non-unique Prior, *Journal of Mathematical Economics* 18, 1989, pp. 141–153.

- Axiomatizing nonadditive probability:
  D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* 57, 1989, pp. 571–587.

- Nonadditive probability and MMEU:
  D. Schmeidler, Integral Representation without Additivity, *Proc. of the Amer. Math. Soc.* 97:2, 1986, pp. 255-261.

- Case-based decision theory:
  I. Gilboa and D. Schmeidler, *A Theory of Case-Based Decisions*, Cambridge University Press, 2001.

- General results on uncertainty
  - belief functions and other representations of uncertainty
  - relating nonadditive probability to lower probability
  - updating representations of uncertainty
  - general notions of expectation

  J. Y. Halpern, *Reasoning About Uncertainty*, MIT Press, 2003 (out in June, I hope)

- More on belief functions:

  R. Fagin and J. Y. Halpern, A new approach to updating beliefs, in *Uncertainty in Artificial Intelligence 6*, 1991, pp. 347–374.

  J. Y. Halpern and R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, *Artificial Intelligence* **54**, 1992, pp. 275–317

  http://www.cs.cornell.edu/home/halpern/papers/ update1.pdf and . . . /update2.pdf.