SOME APPLICATIONS OF A SIMPLE THEORY OF CHOICE UNDER AMBIGUITY

MAXWELL B. STINCHCOMBE

Abstract.

Choice under uncertainty is an umbrella term for situations in which actions do not lead deterministically to outcomes.

(1) A choice problem is risky if choices lead to probability distributions over consequences.
(2) A choice problem is ambiguous if choices lead to sets of probabilities.
(3) Sets of probabilities are identified with incomplete descriptions of the distribution on consequences.

This work applies continuous, linear extensions of expected utility preferences on probabilities to sets of probabilities to

(1) a complete separation of attitudes toward risk and toward uncertainty,
(2) a theory of efficient allocation of ambiguity when attitudes toward risk and ambiguity differ, and
(3) a theory of efficient contracts when some actors have more control over ambiguity than others.

The centrality of new goods and techniques

There is at least a century's worth of social science commentary on the critical role of new goods and production techniques in capitalist expansion. Some name-dropping —

(1) Marx on the role of positive feedback cycles in expansion (1864 (1906) Vol. I, Ch. XXIV, esp. p. 663 *et seq.*).

(2) Weber (1930) discussing the congruence between Puritan ideals and the accumulation and expansion of capital.

(3) Knight (1921 (1971) Ch. 11) on the role of uncertainty in social progress.

(4) Schumpeter (1942, Ch. 12) on the entrepreneurial function in expansion.

Definitionally, a **new** good or production process is one for which the distribution of the rewards associated is not and can not be known. This work is an attempt to trace through Knight's arguments about how ambiguity leads to extra-economic profits, profits above and beyond those that show up in competitive equilibria with risk.

Preferences of innovators and contracts

Schumpeter (1942, p. 132) briefly, and Knight (1921 (1971) Part III) much more extensively discussed the psychological makeup of people who introduce new products and new methods of production, people who make decisions not knowing the distribution of the consequences.

My model-based version of this argument: X is a random reward to be efficiently split between two agents, one less ambiguity averse than the other. The distribution of X belongs to some set. This is an ambiguous problem if the set contains more than one point. An allocation is a mapping $x \mapsto (f_1(x), f_2(x))$. In principle, $f_i(x)$ can be a distribution over rewards.

The question is "What are the efficient allocations?"

In general, the efficient allocations are extreme points of the set of allocations. They reflect tradeoffs between attitudes towards risk and ambiguity.

In special cases, arguably the ones that Knight had in mind, one achieves Knight's result that if i is more ambiguity averse than j, then $f_i(x) \equiv c$ is efficient. In other words, those most tempermentally suited to dealing with ambiguity bear all of it. They become the residual claimants, the entrepreneurs.

Two comments:

(1) The extra-economic profit is the premium, above and beyond the risk premium, that the residual claimant receives for absorbing the ambiguity.

(2) This is a preference based argument for a particular pattern. Self-selection into roles is implicit.

Roles and contracts

Knight also gave a role-based argument for who will be the residual claimant. If there is more ambiguity attached to the choices of one agent, then one expects that agent to be the residual claimant. In modern terms, I believe that Knight's argument is about incentives to exercise the effort to control the ambiguity, and about the other party to the contract to be unwilling to be subject to ambiguity as well as risk due to moral hazard problems.

My model-based version of this argument: Two players much choose their actions/effort levels simultaneously. The choices of one agent have larger effects on the ambiguity of the outcome. This structure defines the roles of the two players.

In general, tradeoffs between incentive effects and ambiguity effects are more complicated than the tradeoffs between incentive effects, and those are already very complicated. One can, however, see the patterns that lead to (something like) the residual claimant being the one whose actions have a large effect on the ambiguity of the outcome.

Ambiguity is not probabilistic

The Ellsberg paradox is an example of people strictly prefering to know the distribution of random variables. If we believe that this is a real phenomenom, then preferences, linear or non-linear, over probability distributions, are not the right tool for choosing in all uncertain situations.

An urn is known to contain 90 balls, 30 of which are Red, each of the remaining 60 can be either Green or Blue. The decision maker is faced with the urn, the description just given, and two pairs of choice situations.

1. Single ticket choices:
   (a) The choice between the Red and the Green ticket.
   (b) The choice between the Red and the Blue ticket.
2. Pairs of ticket choices:
   (a) The choice of the R&B or the G&B pair.
   (b) The choice of the R&G or the B&G pair.

In each situation, after the DM makes her choice, one of the 90 balls will be picked at random. If the ball's color matches the color of (one of) the chosen ticket(s), the decision maker gets $\$1,000$, otherwise they get nothing.

Easy case

Typical preferences are

$$R \succ G \text{ and } R \succ B,$$

$$R\&B \prec G\&B \text{ and } R\&G \prec B\&G.$$

There is no possible Bayesian explanation for these preferences. If there was, we'd have both

$$P(R) > P(G) \text{ and } P(R) > P(B),$$

as well as

$$P(R) + P(B) < P(G) + P(B), \; P(R) + P(G) < P(B) + P(G).$$

Note that this argument encompasses *any* story about beliefs about the distribution of the numbers of Blues and Greens.[1]

[1] Well, okay, it doesn't encompass the idea that someone is cheating and re-arranging the balls after you've picked a color.

Preferences over sets of probabilities

The general model is that choices lead to sets of probabilities, and preferences over sets of probabilities induce preferences over choices.

The probability that the Red ticket wins is $\frac{1}{3}$. That is, the action "choose Red" is risky, with the associated probability $\frac{1}{3}$. The actions "choose Blue" and "choose Green" are ambiguous, leading to the interval of probabilities $[0, \frac{2}{3}]$.

Choosing the Blue&Green pair is risky, $\frac{2}{3}$, choosing the other two pairs is ambiguous, $[\frac{1}{3}, 1]$.

As noted, the typical preferences are

$$\{\tfrac{1}{3}\} \succ [0, \tfrac{2}{3}] \text{ and } \{\tfrac{2}{3}\} \succ [\tfrac{1}{3}, 1].$$

People prefer knowing the center of the interval to the interval itself.

Linear preferences on intervals of probabilities

We add sets in a vector space with $A + B = \{a + b : a \in A,\ b \in B\}$. We multiply by non-negative constants with $\lambda A = \{\lambda a : a \in A\}$. Convex and other linear combinations simply paste these two together.

Every interval $[a, b]$ is of the form $[c - r, c + r]$ for $c = (a + b)/2,\ r \geq 0$. Adding intervals involves adding the centers and the radii.

Under study are non-trivial, continuous, affine preferences, $\succeq$, on, $\mathbb{K} = \mathbb{K}([0, 1])$, the closed convex subsets of $[0, 1]$. The singleton subsets of $[0, 1]$ will be denoted either $[p, p]$ or $\{p\}$. Non-trivial, continuous, affine preference can be represented by a function $U : \mathbb{K} \to \mathbb{R}$ such that

1. (Continuous) $c^n \to c$ and $r^n \to r$, implies $U([c^n - r^n, c^n + r^n]) \to U([c - r, c + r])$,
2. (Affine) $U(\alpha[c - r, c + r] + (1 - \alpha)[c' - r', c' + r']) = \alpha U([c - r, c + r]) + (1 - \alpha)U([c' - r', c' + r'])$, and
3. (Non-triviality, normalized) $U(\{0\}) = 0$ and $U(\{1\}) = 1$.

The unique representation is $U([c - r, c + r]) = c - vr$. Any $v > 0$ represents the urn preferences above.

Finitely many consequences

This work develops the theory of preferences over compact convex sets of probability distributions over an arbitrary finite number of consequences. Given time, I will explain why a theory with infinite sets of consequences will need to be different, maybe substantially.

There are two answers to the question, “Why convexity?”

1) Compact convex set of probabilities are e.g. given by linear inequalities: With two consequences, “The best consequence is at least 3 times as likely as the worst,” corresponding to the interval $[\frac{3}{4}, 1]$. With three consequences, an example is “The middle consequence happens at least half the time,” corresponding to the set $\{(p_1, p_2, p_3) \geq 0 : p_1 + p_2 + p_3 = 1,\ p_2 \geq \frac{1}{2}\}$.

2) There is no real loss in restricting attention to convex sets: If A is a closed subset of $[0, 1]$, and α_k is a convex set of weights, then $\sum_k \alpha_k A$ is approximately convex if $\max_k \alpha_k$ is small. Formally, the Starr-Shapley-Folkman theorem plus linearity and continuity gives

Lemma 1. *If U is a continuous affine mapping on the compact subsets of $\mathbb{R}^n$, then for all compact A, $U(A) = U(\mathbf{co}\,(A))$.*

Finitely many consequences

$\mathfrak{C}$ is a finite set of consequences, $\Delta(\mathfrak{C})$ the set of probabilities on $\mathfrak{C}$. I assume that $\succeq$ on $\mathbb{K}(\Delta(\mathfrak{C}))$ is representable by a continuous, non-trivial, and affine function U.

Given continuity and non-triviality, the independence axiom behind this is $\forall A \succeq B$, $\forall \alpha \in (0,1)$ and $\forall C$,

$$\alpha A + (1-\alpha)C \succeq \alpha B + (1-\alpha)C.$$

There are two kinds of uncertainty in these convex combinations, the risky kind, in the α, and the ambiguous kind, in the sets A, B and C.

Note that the average of a collection of sets if dimension d is another set of dimension d and the diameter does not shrink to 0. In particular, $\sum_k \alpha_k A \equiv A$ for all convex weights $(\alpha_k)_{k=1}^K$. This means that there is NOT an intuition about uncertainty averaging out, as there might be if a Bayesian thought that there was a distribution over the probabilities in A.

Centers of convex sets

To generalize the two consequence case, we need a definition of the center of a compact convex set. Baricenters (are not continuous), Tchebyshev centers (not continuous, and do not reflect important changes in sets), and Steiner points.

$\boldsymbol{U} := \{\boldsymbol{u} \in H : \boldsymbol{u} \cdot \boldsymbol{u} = 1\}$.

$K \mapsto h_K(u) := \max\{u{\cdot}x : x \in K\}$. $h_K(u)$ is a continuous function on $\boldsymbol{U}$. For $1 < p < \infty$,

$$d_p(K, K') = \left(\int_{\boldsymbol{U}} |h_K(\boldsymbol{u}) - h_{K'}(\boldsymbol{u})|^p \, d\mu(\boldsymbol{u}) \right)^{1/p},$$

μ being normalized Lebesgue measure on $\boldsymbol{U}$.

Define $C_p(K)$ as the necessarily unique singleton set minimizing $d_p(\{c\}, K)$. The Steiner point is $C_2(K)$. The other C_p are also nice, but violate the Axiomatics of Steiner points:

It is known [17, Theorem 3.4.2, p. 167] that the mapping $A \mapsto St_A$ is the unique mapping from $\mathbb{K}$ to $\mathbb{R}^n$ that is linear, continuous, and equivariant under rigid motions (i.e. if $R : \mathbb{R}^n \to \mathbb{R}^n$ is a rigid motion, then $R(St_A) = St_{R(A)}$).

Centers of convex sets

For present purposes, the important observations are that

1. St_A is in the relative interior of A,
2. $[A^n \to A] \Rightarrow [St_{A^n} \to St_A,\ St_{A^n - St^n_A} \to St_{A-St_A}]$, and
3. $St_{A-St_A} = 0$.

Another way to understand Steiner points begins with the observation that for Lebesgue almost all $\boldsymbol{u} \in \boldsymbol{U}$, the problem $\max\{u \cdot x : x \in K\}$ has a unique solution, $x_K(\boldsymbol{u})$. The Steiner point can be defined as $St_K = E\, x_K$ where the expectation is taken w.r.t. μ.

The basic model

A decision problem is a triple (A, C, K) where A is a (usually finite) set of actions, C is a finite set of consequences, and K is a mapping from A to $\mathbb{K}(\Delta(C))$, the non-empty, compact convex subsets of $\Delta(C)$. The decision rule under study is

$$a^*(K, U) = \text{argmax}\, \{U(K(a)) : a \in A\}.$$

There is an informative decomposition of $U(\cdot)$.

Decomposition

We work in H, the smallest linear subspace of $\mathbb{R}^C$ containing $\Delta - St_\Delta$, and identify Δ with $\Delta - St_\Delta$, viewed as subset of the vector space H.

Let $\mathbb{K}_S$ be the class of singleton subsets of H and $\mathbb{K}_0 = \{A \in \mathbb{K}(H) : St_A = 0\}$. It is immediate that $\mathbb{K}(H) = \mathbb{K}_S \oplus \mathbb{K}_0$, giving the direct sum decomposition $\mathbb{K}(\Delta) \subset \mathbb{K}_S \oplus \mathbb{K}_0$.

Any continuous affine function, U, on $\mathbb{K}(\Delta)$ can be expressed as the sum of a continuous affine function, u, on $\mathbb{K}_S$, and a linear function, $-v$, on $\mathbb{K}_0$.

This completely separates attitudes toward ambiguity, embodied in v, from attitudes toward risk, embodied in u. $U(A) = u(St_A) - v(A - St_A)$, and if $v \geq 0$, then the DM is ambiguity averse. Whether or not $v \geq 0$ is entirely separated from the attitude toward risk embodied in u.

Rewriting the decision rule

Rewriting the decision rule yields

$$a^*(K,U) = \text{argmax}\{u(K_S(a)) - v(K_0(a)) : a \in A\}, \text{ or}$$

$$a^* = \text{argmax}\{u(a) - v(a) : a \in A\},$$

where $K_S(a) := St_{K(a)}$, $K_0(a) := K(a) - St_{K(a)}$, and the obvious notational changes have been made.

The analyses often turn on changes in $u(\cdot)$ and $v(\cdot)$ changing a^*. Expected utility theory tells us everything there is to know about $u(\cdot)$. Time to look at $v(\cdot)$.

Comparative ambiguity aversion

Studying $v(\cdot)$.

Sets of probabilities are part of the primitive description of a choice problem. Differing reactions to sets contain the information necessary to distinguish between degrees of ambiguity aversion. The starting point is the normalization that $\min_{c \in C} u(c) = 0$ and $\max_{c \in C} u(c) = 1$. Let w and b be the best and worst c's.

We assume some bounds on risk-ambiguity tradeoffs. There are a couple of equivalent ways of talking about the assumptions, allowability, balancedness, and betweenness.

Comparative ambiguity aversion

Studying $v(\cdot)$.

Define the **risk equivalent of** $A \in \mathbb{K}$ is that number $p_A \in \mathbb{R}$ that satisfies $p_A U(\{b\}) + (1 - p_A)U(\{w\}) = U(A)$. It is allowable to dislike ambiguity, but overwhelming hate is not allowable.

Definition 1. *Preferences are* **allowable** *if for all* $A \in \mathbb{K}$, $0 \leq p_A \leq 1$.

Define $\underline{u}(A) = \min\{u(a) : a \in A\}$ and $\overline{u}(A) = \max\{u(a) : a \in A\}$. These represent the preferences of the most wildly pessimistic and the most wildly optimistic possible people. Balanced people are those not beyond either extreme.

Definition 2. *The risk-ambiguity tradeoff is* **balanced** *if for all* $A \in \mathbb{K}$, $\underline{u}(A) \leq U(A) \leq \overline{u}(A)$.

Define $A \vee B = \mathbf{co}\,(A \cup B)$.

Definition 3. *Preferences satisfy* **betweenness** *if for all* $A, B \in \mathbb{K}$, $[U(A) > U(B)] \Rightarrow U(A) \geq U(A \vee B) \geq U(B)$.

I assume that all of these conditions hold. I could have been less profligate with my assumptions.

Studying $v(\cdot)$.

Theorem 1. *The following are equivalent:*

1. *the risk-ambiguity tradeoff is allowable,*
2. *the risk-ambiguity tradeoff is balanced, and*
3. *preferences satisfy betweenness.*

In the normalized, two consequence case, the utility of $[c-r, c+r]$ *is* $c - vr$ *and (1)-(3) are equivalent to*

4. $-1 \leq v \leq 1$.

There must be another condition equivalent to these having to do with $|\partial v(\lambda A)/\partial\lambda|$ for $A \in \mathbb{K}_0$. I'm looking.

Comparative ambiguity aversion

Studying $v(\cdot)$.

Definition 4. *Let $\succeq$ be a continuous, affine preference relation on $\mathbb{K}$. If for all $x \in \Delta$ and all $A \in \mathbb{K}_0$ such that $x + A \subset \Delta$, and for all $0 \leq \lambda < \lambda' \leq 1$,*

$x + \lambda A \succeq x + \lambda' A$, *then $\succeq$ is* **ambiguity averse**,

$x + \lambda A \sim x + \lambda' A$, *then $\succeq$ is* **ambiguity neutral**, *and*

$x + \lambda A \preceq x + \lambda' A$, *then $\succeq$ is* **ambiguity loving**.

The decomposition $U = u - v$ and linearity of v delivers

Lemma 2. *A continuous affine $U : \mathbb{K} \to \mathbb{R}$ represents ambiguity averse (neutral, loving) preferences if and only if the associated $v : \mathbb{K}_0 \to \mathbb{R}$ satisfies $v \geq 0$ ($v \equiv 0$, $v \geq 0$).*

More or less sensitivity of v to $\lambda \cdot A$ covers more or less ambiguity aversion, etc.

Efficiency through preferences

Allocations of ambiguity and risk

It's been a while since the introduction: Knight's (1921) theory of profits above and beyond economic profits is based on ambiguity. In his analysis, entrepreneurs have extra ability or willingness to deal with the choices that must be made in the course of endeavors in which experience provides little or no guide in the assignment of probabilities.

Knight argues that ambiguity is more prevalent in economies introducing new products or production processes, which are all of the capitalist ones. A new product/process is one for which there is no actuarial experience with which to estimate the distribution of sales. For Knight, the ability to deal with such ambiguities is what defines entrepreneurs, leading them to be the residual claimants. This ability earns a premium above and beyond normal economic profits.

A model

$X \in M = \{1, \ldots, M\}$ is a random reward to be efficiently split between two agents, one less ambiguity averse than the other. $X \sim P$, $P \in A \in \mathbb{K}(\Delta(M))$.

This is an ambiguous problem if A contains more than one point. An allocation is a mapping $x \mapsto f(x)$ where $f(x)$ is a probability distribution on $\{(s_1, s_2) : s_1 + s_2 = x\}$. The set of such f's is the compact convex set F.

The two agents receive $w_i + s_i \in C_i$ where w_i is their initial wealth. I assume that $|s_i| \leq M < w_i$, I need some bounds, these seem to cover many interesting possibilities.

Remember the results when A contains only point — efficient contracts have the least risk averse agent bearing all the risk. If i is the least risk averse, then $f_j(x) \equiv s_1^\circ$ for some s_i°. The less risk averse person typically receives a premium for bearing the risk, and the more risk averse person is happy to pay it. Adding ambiguity introduces another layer of potential complication.

Efficiency through preferences

A continuation of a model

The preferences of each i is represented by U_i, a continuous affine function on $\mathbb{K}(\Delta(C_i))$.

For a distribution $\mu \in \Delta(M)$, $f_i(\mu)$ is the marginal distribution of $f(\mu)$ on C_i. $f_i(A) := \{f_i(\mu) : \mu \in A\}$. The mapping $f \mapsto f_i(A)$ is linear. Therefore the efficient f's are the solutions to

$$\max_{f \in F} \lambda U_1(f_1(A)) + (1-\lambda)U_2(f_2(A))$$

for some $\lambda \in [0,1]$.

Given all the linearity, we have

Lemma 3. *The set of efficient contracts always contains an extreme point of F.*

Go to examples and pictures, when both are risk lovers, and more general tradeoffs when $M = 2$.

Efficiency through preferences

The last gasp of the continuation of a model

It looks hard to give informative necessary and sufficient conditions for f_j =constant to be an efficient solution. Here are some sufficient conditions.

Lemma 4. *If* 1 *is much more risk tolerant and much more ambiguity tolerant than* 2*, the efficient* f*'s have* $f_2(\cdot)$ *constant.*

Intuition: if a function f 'compresses' the uncertainty, then $C = f(B)$ will satisfy $v_i(B) < v_i(C)$. Because of the tradeoff between risk and ambiguity, the class of 'compressive' functions that satisfy $v_2(f(B)) \geq v_2(B)$ becomes larger as 1 becomes more ambiguity averse. The constant functions are maximally 'compressive.' The more ambiguity-averse 2 is, the more mean income they're willing to give up to b in order to get $f(X) = constant$. So, 2 can collect a larger premium as 1 becomes more ambiguity averse. And we can replace "ambiguity" with "risk" and make these arguments even truer.

Another model

A reminder: Knight also gave a role-based argument for who will be the residual claimant. If there is more ambiguity attached to the choices of one agent, then one expects that agent to be the residual claimant. There is a moral hazard argument in here.

My model-based version of this argument: Two players much choose their actions/effort levels simultaneously. The choices of one agent have larger effects on the ambiguity of the outcome. This structure defines the roles of the two players.

In general, tradeoffs between incentive effects and ambiguity effects are more complicated than the tradeoffs between incentive effects, and those are already very complicated. One can, however, see the patterns that lead to (something like) the residual claimant being the one whose actions have a large effect on the ambiguity of the outcome.

Efficiency through roles

Continuing another model

The basic simultaneous choice inefficiency when there are external effects. Turn into a random variable story. Add some K_0. Make one person's increased efforts reduce ambiguity, the other's have no effect. Note how you can improve efficiency by making one closer to a residual claimant.

Some math perversities

Vitale [14] showed that Steiner points have no continuous extensions to infinite dimensional Hilbert spaces. I can do just a little bit more and show that the same is true when we restrict ourselves to some compact subsets of H. But there's not a lot of glory in this, I didn't want the result to be true, and my proof is but a slight adaptation of his. $\Delta([0, M])$ is, however, a special compact convex subset of an infinite dimensional Hilbert space, and maybe Steiner points will work there.

References

[1] Allais, M. 1953, Le Comportment de l'Homme Rationnel Devant le Risque: Critique des Postulates et Axiomes de l'Ecole Americaine. *Econometrica* **21**, 503-546

[2] Amarante, M. 2003, States, Models and Unitary Equivalence. Working paper, Department of Economics, Columbia University, New York.

[3] Crawford, V. 1990, Equilibrium without Independence. *Journal of Economic Theory* **50**, 1, 127-54.

[4] Epstein, L. 1999, A Definition of Uncertainty Aversion. *Review of Economic Studies* **66**, 579-608.

[5] Epstein, L. and Zhang, J. 2001, Subjective Probabilities on Subjectively Unambiguous Events. *Econometrica* **69**, 265-306.

[6] Fenchel, W. and Jessen, B. 1938, Mengenfunktionen und konvexe Körper. *Danske Vid. Selskab. Mat.-fis. Medd.* **16**, 3.

[7] Ghirardato, P. Marinacci, M. 2001, Range Convexity and Ambiguity Averse Preferences. *Economic Theory* **17**(3), 599-617.

[8] Karni, E., Schmeidler, D., Vind, K. 1983, On State Dependent Preferences and Subjective Probabilities. *Econometrica* **51**(4), 1021-1032.

[9] Knight, F. H. 1921, *Risk, uncertainty and profit.* Chicago, University of Chicago Press (1971, reprint of the 1921 ed.).

[10] Körner, R. 1995, A Variance of Compact Convex Random Sets. Working paper, Institut fr Stochastik, Bernhard-von-Cotta-Str. 2 09599 Freiberg.

[11] Machina, M. 1982, 'Expected Utility' Analysis Without the Independence Axiom. *Econometrica* **50**, 277-323.

[12] Mukerji, S. 1998, Ambiguity Aversion and Incompleteness of Contractual Form. *American Economic Review* **88**(5), 1207-31.

[13] Mukerji S.; Tallon J-M. 2001, Ambiguity Aversion and Incompleteness of Financial Markets. *The Review of Economic Studies* **68**(4), 883-904.

[14] Vitale, R. 1985, Steiner Points in Infinite Dimensions, *Israel Journal of Mathematics*.

[15] Von Neumann, J. and Morgenstern, O. 1944, *Theory of games and economic behavior*. New York, Wiley, (1953, 3d ed.).

[16] Savage, L. J. 1954, *The foundations of statistics*. New York, Wiley.

[17] Schneider, R. 1993, *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge, Cambridge University Press.