Bright and Kinney 2021: Elite-Group Ignorance and Risk Aversion

Kevin Dorst 24.223

I. Willful Ignorance

Societal elites seem to be stubbornly ignorant of the degree to which they are privileged by society.

E.g. white ignorance (Mills 2007).

This seems to redound badly on those elites. Perhaps it shows that they have corrupted values. Perhaps shows that they're irrational.

"Motivated irrationality"?

How to establish?

B&K use the *value of evidence* result from Bayesian decision theory to first bolster this charge. Then show if people are risk-averse, they might escape it.

I'm inclined to think they should make more hay of the moral constraints on values they can impose, and still have the argument go through.

So the challenge will be that such elites must *either* have immoral values or are irrational. And then risk-aversion offers a rejoinder.

II. The Challenge: Value of Evidence

Begin with standard decision theory. Worlds; probabilities of worlds; acts as functions from worlds to numbers (i.e. random variables).

$A(w)$ = the amount of utility that *would* be realized if you were to do A in world w.

Expected value of A: $\mathbb{E}(A) = \sum_t P(A = t) \cdot t$.

The Value of Evidence: Evidence is valuable because it helps make our actions more sensitive to the world, and hence allows us to make better choices, given our values.

Good 1967

Example:

- Jill, a new CEO, must decide whether to push hard for *C*hanges to hiring practices (C), or not.
- Doing so would either *H*elp the company and community (H), or would instead be a *W*aste of resources and political capital (W) that would harm them.
- Alternatively, she could stick with the *S*tatus quo (S).

Either $C \boxrightarrow H$ (Changes would *H*elp) or $C \boxrightarrow W$ (Changes would *W*aste).

Options, taking into account only morally-relevant values:

	$C \boxrightarrow H$	$C \boxrightarrow W$
*C*hange	$+100$	-100
*S*tatus quo	0	0

A relevant consideration is whether the current hiring practices are *B*iased (B) or not. If they are, it's likely the changes would help; if not,

it's unlikely they would.

$P(\mathrm{C} \boxright \mathrm{H}|B) = 0.8$, while $P(\mathrm{C} \boxright \mathrm{H}|\neg B) = 0.2$.

Currently, Jill (reasonably, let's suppose) thinks it's unlikely that practices are biased ($P(B) = 0.3$), but at little or no cost she can order a review to find out if this is so. Should she?

Currently S is the expectedly best option, with value 0:

$$\begin{aligned} P(C \boxright H) &= P(B) \cdot P(C \boxright H|B) + P(\neg B) \cdot P(C \boxright H|\neg B) \\ &= 0.3 \cdot 0.8 + 0.7 \cdot 0.2 = 0.38, \text{ so:} \\ \mathbb{E}(C) &= 0.38 \cdot 100 + 0.62 \cdot (-100) = -24 \end{aligned}$$

Meanwhile, $\mathbb{E}(S) = 0$.

But what happens if she orders a review? If she learns B, then the best option will be C; if she learns $\neg B$, her best option will be S. So the expected value of ordering a review (O) is:

$\mathbb{E}(C|B) = 0.8 \cdot 100 + 0.2(-100) = 60$ which is greater than $0 = \mathbb{E}(S|B)$.

$$\begin{aligned} \mathbb{E}(O) &= P(B) \cdot \mathbb{E}(C|B) + P(\neg B) \cdot \mathbb{E}(S|\neg B) \\ &= 0.3 \cdot 60 + 0.7 \cdot 0 \\ &> 0.3 \cdot 0 + 0.7 \cdot 0 = \mathbb{E}(S). \end{aligned}$$

First Upshot: Learning whether B allows her to make her decision sensitive to further factors—it gives her more options! And so is expectedly better.

E.g. "Do C if B and do S if $\neg B$.

This is *always* true in (Savage) decision theory: if you are offered the chance to freely learn an answer to a question, then if this ever affects your decision, you should expect doing so to lead to a better decision.

= learn which cell of a partition is true.

Second Upshot: If Jill *doesn't* order the review, what can we conclude? That either she has some other (bad) values, or she's irrational!

E.g. it would be painful for her to learn that B is true.

III. The Reply: Risk Aversion

Buchak 2013

Real people are *risk averse*. Moreover, this seems reasonable.

Which would you prefer?

- *No Bet:* \$100 for sure.
- *Bet:* 50% shot at \$200.

If you prefer *No Bet*, suggests EUT is wrong to weight the value of \$200 by $0.5 \times$ the utility of \$100.

How else could we weight it? Using a *risk function* R.

$R(0) = 0$, $R(1) = 1$, and R is monotonically increasing. E.g. $R(x) = x^2$ or $R(x) = \sqrt{x}$.

How to apply?

Suppose $t_1 < t_2, ..., < t_n$ are the possible values of an act A. Then:

$$\begin{aligned}\mathbb{E}(A) &= P(A = t_1)(t_1) + P(A = t_2)(t_2) + \cdots + P(A = t_n)(t_n) \\ &= P(A \geq t_1)(t_1) + P(A \geq t_2)(t_2 - t_1) + \cdots P(A \geq t_n)(t_n - t_{n-1}) \\ &= t_1 + P(A \geq t_2)(t_2 - t_1) + \cdots P(A \geq t_n)(t_n - t_{n-1})\end{aligned}$$

Get to *risk-weighted* expected utility by applying the risk-function to each of these probabilities of increases:

$$\begin{aligned}\mathbb{RE}(A) &= \boldsymbol{R}(P(A \geq t_1))(t_1) + \boldsymbol{R}(P(A \geq t_2))(t_2 - t_1) + \cdots \boldsymbol{R}(P(A \geq t_n))(t_n - t_{n-1}) \\ &= t_1 + \boldsymbol{R}(P(A \geq t_2))(t_2 - t_1) + \cdots \boldsymbol{R}(P(A \geq t_n))(t_n - t_{n-1})\end{aligned}$$

So for example if $R(x) = x^2$, then

- $\mathbb{RE}(\textit{No Bet}) = \$100$, while
- $\mathbb{RE}(\textit{Bet}) = 0 + R(P(\textit{Bet} \geq 200))(200 - 0) = 0.5^2(200) = \$50$

Why does this matter? Because risk-averse decision theory leads to failures of the value of evidence.

Suppose Jill is risk-averse with $R(x) = x^2$. Then even if her values are the morally kosher ones, she can rationally think it's not worth the risk to find out whether B.

Finding out whether B might lead her to take riskier actions (like pushing for change); so she prefers not to do so.

In detail:

- If she learns B, she'll push for changes:

$$\begin{aligned}\mathbb{RE}(C|B) &= -100 + R(P(C \geq 100|B)) \cdot (100 - (-100)) \\ &= -100 + 0.8^2(200) = -100 + 128 > 0 = \mathbb{RE}(S|B)\end{aligned}$$

- If she learns $\neg B$, she won't.
- So if she orders a review, the possible values are -100 (if she learns B, so does C, and it wastes resources), 0 (if she learns $\neg B$ and so does S), or 100 (if she learns B, so does C, and it helps). Thus:

$$\begin{aligned}\mathbb{RE}(O) &= -100 + R(P(O \geq 0))(0 - (-100)) + R(P(O \geq 100))(100 - 0) \\ &= -100 + R(P(\neg B) + P(B)P(C \boxdot\!\!\rightarrow H|B))(100) + R(P(B) \cdot P(C \boxdot\!\!\rightarrow H|B))(100) \\ &= -100 + (0.7 + 0.24)^2(100) + (0.3 \cdot 0.8)^2(100) \\ &= -5.88 < 0 = \mathbb{RE}(S).\end{aligned}$$

So in fact, if risk-averse in this way, Jill would prefer to avoid finding out whether B, *even though* her values are morally kosher.

References

Buchak, Lara, 2013. *Risk and rationality*. Oxford University Press.

Good, I J, 1967. 'On the Principle of Total Evidence'. *The British Journal for the Philosophy of Science*, 17(4):319–321.

Mills, Charles W., 2007. 'White ignorance'. *Race and Epistemologies of Ignorance*, 13–38.