EXTERNALISM AND THE VALUE OF INFORMATION

NILANJAN DAS
UNC CHAPEL HILL/NYU SHANGHAI

1. Plan

- *Two Principles.*

  VALUE OF INFORMATION. If evidence is available to an agent (for gathering and use) at a negligible cost, then it is instrumentally rational for that agent to gather that evidence and use it for making decisions.

  EVIDENCE EXTERNALISM. An agent's evidence may include non-trivial propositions about the external world.

For decision-theoretic arguments for VALUE OF INFORMATION, see Peirce (1967), Ramsey (1990), and Good (1967).

McDowell (1995, 2011), Williamson (2000), and Goldman (2009) defend EVIDENCE EXTERNALISM.

- *Aim*. I want to show that when we formulate these principles within a certain framework, a tension emerges between them.

2. Good's Argument

- *An Example*.

  Example 1. You work in a chemical laboratory. You want to determine the chemical properties of a certain solution: you know that it is either acidic or alkaline, but you currently have neither more nor less reason to think that it is acidic rather than alkaline. You don't want to misclassify the solution: this will make certain experiments go wrong. You have at your disposal a blue litmus paper and a red litmus paper. If the blue litmus paper turns red when brought in contact with the solution, you will learn that the solution is acidic. If the red litmus paper turns blue when brought in contact with the solution, you will learn that the solution is alkaline. Should you test the solution using these pieces of litmus paper before you decide where to store the solution?

  The answer of course is "Yes." And this supports VALUE OF INFORMATION. To vindicate this idea, Good (1967) offered an argument for VALUE OF INFORMATION.

- *The Formal Framework*. Start with a formal framework for representing the agent's evidence, her credences, and her preferences.

  1. *Frames*. A frame $\mathcal{F}$ is a structure $\langle W, E \rangle$ where W is a finite set of worlds and E is a function that maps worlds in W to sets of worlds, which represent the relevant agent's *evidence* at those worlds.

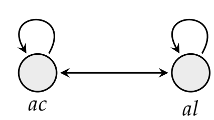

Figure 1.1: Your Evidence Before the Test

Figure 1.2: Your Evidence After the Test

2. *Decision-Problems*. A decision problem D is a structure $\langle W, E, A, \pi, \mu, f\rangle$ where, besides $\langle W, E\rangle$ being a frame,

   (a) A is a set of acts;
   (b) π is a *regular ur-prior* (i.e., an initial credence function) that is rationally permissible for the agent to have independently of all empirical evidence;
   (c) μ is a utility function that maps pairs of acts in A and worlds in W to real numbers; and
   (d) f is a function that takes any world in W as input and outputs an act in A that the agent most prefers relative to her evidence and her utility function in that world.

π assigns values to propositions, or sets of worlds in W. But somewhat loosely, I will often write $\pi(w)$ instead of $\pi(\{w\})$.

- *Properties of Frames and Decision-Problems*. We shall need three properties of frames and decision-problems.

  1. *Partitionality*. A frame $\mathcal{F} = \langle W, E\rangle$ is partitional if and only if it is *reflexive*, *transitive*, and *euclidean*.
  2. *Coarsening*. For any two frames $\langle W, E\rangle$ and $\langle W, F\rangle$, E is *coarser* than F if and only if, for any $w \in W$, $F(w) \subseteq E(w)$.
  3. *Bayesian Rationality*. A decision problem $D = \langle W, E, A, \pi, \mu, f\rangle$ satisfies *Bayesian rationality* if and only if, at any world $w \in W$, two conditions are satisfied.

     (a) CONDITIONALIZATION. Relative to her total body of evidence $E(w)$, the agent's credence function is the conditional credence function $\pi(.|E(w))$.
     (b) EXPECTED UTILITY MAXIMIZATION. The agent most prefers an act that maximizes expected utility relative to her credence function and her utility function. More formally,
     $$(\forall w \in W)(f(w) \in \arg\max_{a \in A} \sum_{w' \in W} \pi(w'|E(w))\mu(a, w')).$$

Partitionality roughly corresponds to the following structural properties of evidence.

FACTIVITY. If an agent's evidence entails a proposition X in a world w, then X is true in w.

POSITIVE ACCESS. If an agent's evidence entails a proposition X in a world w, then her evidence in w entails that her evidence entails X.

NEGATIVE ACCESS. If an agent's evidence doesn't entail a proposition X in a world w, then her evidence in w entails that her evidence doesn't entail X.

Coarsening corresponds to a *no information loss* requirement: when $\langle W, E\rangle$ and $\langle W, F\rangle$ represent the agent's possible current and future bodies of evidence, if E is *coarser* than F, then that means that the agent doesn't lose any evidence over time in any world.

- *Good's Argument*

  GOOD'S THEOREM (GOOD 1967). Suppose there are two decision problems $D_1 = \langle W, E, A, \pi, \mu, f\rangle$ and $D_2 = \langle W, F, A, \pi, \mu, g\rangle$, which are based on partitional frames and satisfy Bayesian rationality, such that E is coarser than F. Then, for any world $w \in W$,
  $$\sum_{w' \in W} \pi(w'|E(w))\mu(f(w'), w') \leq \sum_{w' \in W} \pi(w'|E(w))\mu(g(w'), w'),$$
  with strict inequality unless, for all $w' \in E(w)$, $f(w') = g(w')$.

  This inequality is sometimes called *Good's inequality*.

- *Relaxing the Assumptions*. In what follows, I ask what happens if we give up the assumption that agent's present and future bodies of evidence satisfy *Partitionality*.

Some writers (prominently, Skyrms 1990, Kadane et al 2008, Buchak 2010, Hutteger 2014, Bradley and Steele 2016, Campbell-Moore and Salow ms) have tried to see what happens to Good's inequality when you relax some of the *other* relevant assumptions. The only other writers who have explored the connection between *Partitionality* and Good's inequality are Ahmed and Salow (forthcoming) and Dorst (ms). Also, cf. Schoenfield (forthcoming).

3. Externalism and the Access Principles

- *Externalism*. Some think that we should embrace EVIDENCE EXTERNALISM if we want to escape skepticism about the external world.

- *Access Principles*. If we accept EVIDENCE EXTERNALISM and FACTIVITY, then we should reject NEGATIVE ACCESS. Some defenders of EVIDENCE EXTERNALISM also reject POSITIVE ACCESS, but that's more controversial.

- *Good's Inequality without Access Principles*.

  Example 2. You are about to enter a room, and look at a wall. Your current evidence entails that the wall is going to be one of three shades of red: crimson, rusty red, and cardinal red. On the basis of your current evidence, you are 0.1 confident that the wall is going to be crimson, 0.8 confident that it is going to be rusty red, and 0.1 confident that it is going to be cardinal red. You are also certain that you can discriminate crimson from cardinal red, and vice-versa, but you can't discriminate rusty red from either crimson or cardinal red. You know you will be offered a gamble where you stand to gain $100 if the wall is rusty red, and lose $450 if it's not. Should you make a decision about this before you enter the room?

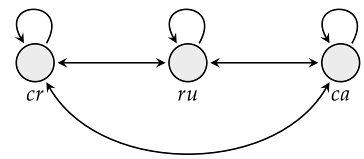

Figure 2.1: Your Current Evidence in *Example 2* given by $\langle W, E\rangle$

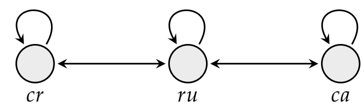

Figure 2.2: Your Future Evidence in *Example 2* given by $\langle W, F\rangle$

	Worlds		
Acts	*cr*	*ru*	*ca*
Accept	-450	100	-450
Reject	0	0	0

Table 1: Payoffs for *Example 2*

	Worlds		
Credence Functions	*cr*	*ru*	*ca*
$\pi(.\|E(cr))$	0.1	0.8	0.1
$\pi(.\|E(ru))$	0.1	0.8	0.1
$\pi(.\|E(ca))$	0.1	0.8	0.1
$\pi(.\|F(cr))$	1/9	8/9	0
$\pi(.\|F(ru))$	0.1	0.8	0.1
$\pi(.\|F(ca))$	0	8/9	1/9

Table 2: Your Credences in *Example 2*

  In this scenario, both POSITIVE ACCESS and NEGATIVE ACCESS will fail in the worlds where the wall is crimson or cardinal red. If you satisfy Bayesian rationality everywhere, Good's inequality will also fail.

- *The Question*. Since an externalist needn't reject POSITIVE ACCESS, this still doesn't demonstrate any tension betwen EVIDENCE EXTERNALISM and VALUE OF INFORMATION. Can the externalist preserve Good's inequality while rejecting NEGATIVE ACCESS?

4. Good's Inequality without Negative Access

- *Hope*. The answer, one might think, is "Yes."

  Nestedness. A frame $\langle W, E\rangle$ is *nested* if and only if for any two worlds w, w' in W, if the agent's total evidence in w, $E(w)$, isn't disjoint from her evidence in w', $E(w')$, then either $E(w)$ entails $E(w')$, or $E(w')$ entails $E(w)$.

  Geneakoplos (1989) proves:

  GEANOKOPLOS' THEOREM (THEOREM 1, GEANAKOPLOS 1989). Suppose $D_1 = \langle W, E, A, \pi, \mu, f\rangle$ and $D_2 = \langle W, F, A, \pi, \mu, g\rangle$ are two decision problems which satisfy Bayesian rationality, such that

$\langle W, E\rangle$ is partitional and E is coarser than F. If the frame $\langle W, F\rangle$ satisfies reflexivity, transitivity, and nestedness, then Good's inequality holds for D_1 and D_2.

- *An Example*. There are plenty of cases where NEGATIVE ACCESS fails, but *nestedness* holds.

  Example 3. You are about to enter a room and see a wall. You don't know for sure what the colour of the wall is, but it is 0.99 likely by lights of your current evidence that the wall is red. If the wall is red, your evidence after entering the room will entail that it is red. However, there is a small probability of 0.01 that it is white, but lit up with red right. In that case, your evidence will remain the same as before. You also know that immediately afterwards, you will be offered a gamble where you stand to gain $100 if the wall is red, and lose $10000 if the wall isn't red. Should you make a decision about the gamble before entering the room?

  In this scenario, NEGATIVE ACCESS fails. But since nestedness holds, Good's inequality is preserved.

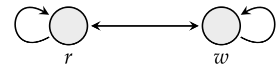

Figure 3.1: Your Current Evidence in *Example 3*

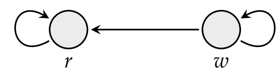

Figure 3.2: Your Future Evidence in *Example 3*

	Worlds	
Acts	r	w
Accept	100	-10,000
Reject	0	0

Table 3: Payoffs for *Example 3*

	Worlds	
Credence Functions	r	w
$\pi(.\|P(r))$	0.99	0.01
$\pi(.\|P(w))$	0.99	0.01
$\pi(.\|Q(r))$	1	0
$\pi(.\|Q(w))$	0.99	0.01

Table 4: Your Credences in *Example 3*

5. The Argument from Fallibility

- *Nestedness and Negative Access*. Nestedness should fail for the same reasons for which NEGATIVE ACCESS fails.
  - NEGATIVE ACCESS fails because the mechanisms by which we gather information about the external world are *fallible*: they sometimes provide us false information even when we have no clue that this has happened.
  - *Nestedness* should fail in cases where there are multiple such mechanisms, which malfunction in the relevant ways independently of each other.
- *A Final Example*.

  Example 4. You are about to go into a room and encounter a wooden wall. You don't know for sure what the colour of the wall is or what kind of wood it is made of. But you are rationally 0.99 confident that the wall is red (/made of sandalwood). If the wall is red (/made of sandalwood), you will see that it is red (/know by smell that it is made of sandalwood); so your evidence after entering the room will entail that the wall is red (/made of sandalwood). However, there is a small probability of 0.01 that it is white, but will be lit up with red right when you enter the room (/made of ordinary wood but smeared with sandalwood perfume). If that happens, your evidence will remain the same as before.

  Relative to your current credence function, the possibility of the wall's being red is probabilistically independent of the possibility

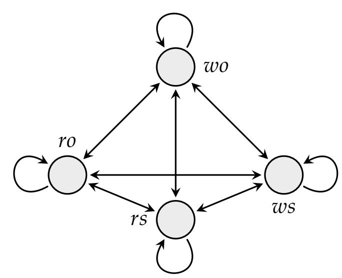

Figure 4.1: Your Current Evidence in *Example 4* given by $\langle W, E\rangle$

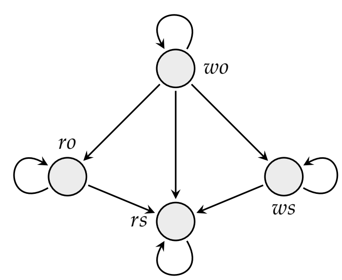

Figure 4.2: Your Future Evidence in *Example 4* given by $\langle W, F\rangle$

of its being made of sandalwood, while the possibility of the wall's being white is probabilistically independent of the possibility of its being made of ordinary wood. You also know that immediately afterwards, you will be offered a gamble where you stand to gain $100 if the wall is red and made of sandalwood, and lose $5,000 if the wall is either white or not made of sandalwood. Should you make a decision about the gamble before entering the room?

	Worlds			
Acts	*rs*	*ro*	*ws*	*wo*
Accept	100	-5,000	-5,000	-5,000
Reject	0	0	0	0

Table 5: Payoffs for *Example 4*

	Worlds			
Credences	*rs*	*ro*	*ws*	*wo*
$\pi(.\|E(rs))$	0.9801	0.0099	0.0099	0.0001
$\pi(.\|E(ro))$	0.9801	0.0099	0.0099	0.0001
$\pi(.\|E(ws))$	0.9801	0.0099	0.0099	0.0001
$\pi(.\|E(wo))$	0.9801	0.0099	0.0099	0.0001
$\pi(.\|F(rs))$	1	0	0	0
$\pi(.\|F(ro))$	0.99	0.01	0	0
$\pi(.\|F(ws))$	0.99	0	0.01	0
$\pi(.\|F(wo))$	0.9801	0.0099	0.0099	0.0001

Table 6: Your Credences in *Example 4*

In this case, nestedness fails, ultimately leading to a failure of Good's inequality. So, if fallibility is the reason why the externalist rejects NEGATIVE ACCESS, they cannot rely on nestedness to preserve VALUE OF INFORMATION.

- *Responses*. There are two strategies that the externalist could use to reject *Example 4*.
  - *Self-Evident Preferences*. She could claim: one is always certain of what one prefers. That's not true in *Example 4*.

    Reply. But we can construct cases where Good's inequality fails even when an agent's preferences are always self-evident to her.
  - *Self-Evident Credences*. She could claim: one is always certain of what one's credences are. That's not true in *Example 4*.

    Reply. But, if we want to allow for failures of NEGATIVE ACCESS, then there has to be a violation of some other constraints.

6. Despair

The only other option is to reject EVIDENCE EXTERNALISM and accept a Cartesian picture of evidence. But it's by no means obvious that Good's inequality will hold on that picture.

References

Ahmed, A. and Salow, B. (forthcoming). Don't look now. *The British Journal for the Philosophy of Science*.

Bradley, S. and Steele, K. (2016). Can free evidence be bad? value of information for the imprecise probabilist. *Philosophy of Science*, 83(1):1–28.

Buchak, L. (2010). Instrumental rationality, epistemic rationality, and evidence-gathering. *Philosophical Perspectives*, 24(1):85–120.

Campbell-Moore, C. and Salow, B. (ms.). Avoiding risk and avoiding evidence. Unpublished Manuscript.

Dorst, K. (ms.). Evidence: A guide for the uncertain. Unpublished Manuscript.

Geanakoplos, J. (1989). Game theory without partitions, and applications to speculation and consensus. Technical report. Cowles Foundation Discussion Paper No. 914, Yale University.

Goldman, A. (2009). Williamson on knowledge and evidence. In Greenough, P. and Pritchard, D., editors, *Williamson on Knowledge*, pages 73–92. Oxford: Oxford University Press.

Good, I. J. (1967). On the principle of total evidence. *British Journal for the Philosophy of Science*, 17(4):319–321.

Huttegger, S. M. (2014). Learning experiences and the value of knowledge. *Philosophical Studies*, 171(2):279–288.

Kadane, J. B., Schervish, M., and Seidenfield, T. (2008). Is ignorance bliss? *Journal of Philosophy*, 105(1):5–36.

McDowell, J. (1995). Knowledge and the internal. *Philosophy and Phenomenological Research*, 55(4):877–93.

McDowell, J. H. (2011). *Perception as a Capacity for Knowledge*. Marquette University Press.

Peirce, C. S. (1967). Note on the theory of the economy of research. *Operations Research*, 15(4):643–648.

Ramsey, F. P. (1990). Weight or the value of knowledge. *The British Journal for the Philosophy of Science*, 41(1):1–4.

Schoenfield, M. (forthcoming). Conditionalization does not (in general) maximize expected accuracy. *Mind*.

Skyrms, B. (1990). The value of knowledge. *Minnesota Studies in the Philosophy of Science*, 14:245–266.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.