Language Dependence in Philosophy of Science and Formal Epistemology

Branden Fitelson

Department of Philosophy

&

Center for Cognitive Science (RuCCS)

Rutgers University

branden@fitelson.org

- Suppose we have two false hypotheses H_1 and H_2.
- Sometimes, we would like to be able to say that H_1 is *closer to the truth* than H_2 (*e.g.*, Newton's hypothesis *vs.* Ptolemy's).
- Various accounts of "closeness to the truth" (*verisimilitude*) have been proposed in the literature (see [12], [10], [11]).
- One of the many accounts is a naive, syntactical explication, which "counts numbers of true conjuncts" ([15], [16]).
- **Problematic Example.** Suppose the truth about the weather is that it is *H*ot and *R*ainy and *W*indy ($T = H \,\&\, R \,\&\, W$).
- And, consider the following two false hypotheses:

  (H_1) $\sim H \,\&\, R \,\&\, W$.

  (H_2) $\sim H \,\&\, \sim R \,\&\, \sim W$.
- If we "count numbers of true conjuncts", then we get the verdict that H_1 is closer to the truth than H_2 ($H_1 \prec H_2$).
- A problem with this account is that the orderings it imposes on hypotheses are *language dependent* [9]. Let me explain.

- In our example, we adopted the HRW-language ($\mathscr{L}_1$).
- Consider this *expressively equivalent* HMA-language ($\mathscr{L}_2$).
  - $\mathscr{L}_2$ has atoms H and M, A, where M, A are such that:
    - $M \dashv\vDash H \equiv R$. ["*M*innesotan" weather]
    - $A \dashv\vDash H \equiv W$. ["*A*rizonan" weather]
- $\mathscr{L}_1$ and $\mathscr{L}_2$ can express exactly the same propositions. This is what it means for $\mathscr{L}_1$, $\mathscr{L}_2$ to be *expressively equivalent*.
- In $\mathscr{L}_2$, the claims T, H_1, and H_2 are expressed as follows:

  (T) $H \,\&\, M \,\&\, A$.

  (H_1) $\sim H \,\&\, \sim M \,\&\, \sim A$.

  (H_2) $\sim H \,\&\, M \,\&\, A$.
- Thus, if we "count numbers of true conjuncts" — *in* $\mathscr{L}_2$, then we get a *reversal* of the ordering we got in $\mathscr{L}_1$. That is:
  - In $\mathscr{L}_1$, we have $H_1 \prec H_2$.
  - In $\mathscr{L}_2$, we have $H_2 \prec H_1$.

☞ This shows that verisimilitude — on this "counting true conjuncts" definition — is *language dependent* ([9], [10]).

- Let $\mathcal{P}(x) \stackrel{\text{def}}{=}$ the set of predicates that an object x falls under.
- We seek a measure $\mathfrak{s}(x, y)$ of the "degree to which x and y are similar". A naive way to measure similarity is by "counting the predicates x and y share (both fall under)".
- Here is the naive "shared predicate" $\mathfrak{s}$imilarity measure:

$$\mathfrak{s}(x, y) \stackrel{\text{def}}{=} \left|\mathcal{P}(x) \cap \mathcal{P}(y)\right|$$

- That is, $\mathfrak{s}(x, y)$ is just the size (cardinality) of the intersection of the sets of predicates that x and y (respectively) fall under. [Like "counting conjuncts" again!]
- Carnap *et. al.* ([3], [4]) say things which suggest the following "analogy by similarity" principle for inductive logic (where, it is not known whether a falls under the predicate ϕ):

  ($\mathscr{A}$) If $n > m$, then $\Pr[\phi a \mid \phi b \,\&\, \mathfrak{s}(a, b) = n] > \Pr[\phi a \mid \phi b \,\&\, \mathfrak{s}(a, b) = m]$.
- In words, ($\mathscr{A}$) says the following. Suppose that b falls under ϕ. Then, the probability that a also falls under ϕ *increases as the (known) similarity between a and b* [$\mathfrak{s}(a, b)$] *increases*.

- Unfortunately, ($\mathscr{A}$) leads to a *language dependent* theory of inductive probability. A Miller-style trick proves this.
  - The $ABCD$–language ($\mathscr{L}_1$) has four predicates A, B, C, D.
  - The $AXYD$–language ($\mathscr{L}_2$) also has four predicates A, X, Y, D, where X and Y are subject to the following Miller-esque semantic constraints: $Xx \dashv\vDash Ax \equiv Bx$, $Yx \dashv\vDash Bx \equiv Cx$.
    - Note: *$\mathscr{L}_1$ and $\mathscr{L}_2$ are expressively equivalent*!
  - Now, consider two objects a and b such that (speaking $\mathscr{L}_1$):

$$Aa \,\&\, Ba \,\&\, Ca$$
$$Ab \,\&\, {\sim}Bb \,\&\, Cb$$

  - In $\mathscr{L}_2$, a and b have the following (equivalent) descriptions:

$$Aa \,\&\, Xa \,\&\, Ya$$
$$Ab \,\&\, {\sim}Xb \,\&\, {\sim}Yb$$

  - Now, we see that our measure $\mathfrak{s}$ is *language dependent*:

$$\mathfrak{s}_{\mathscr{L}_1}(a,b) = 2 \neq 1 = \mathfrak{s}_{\mathscr{L}_2}(a,b)$$

- We can use this example to establish the language dependence of any probability theory that satisfies ($\mathscr{A}$).

- Recall the Carnapian principle of "analogy by similarity":

  ($\mathscr{A}$) If $n > m$, then $\Pr[\phi a \mid \phi b \;\&\; \mathfrak{s}(a,b) = n] > \Pr[\phi a \mid \phi b \;\&\; \mathfrak{s}(a,b) = m]$.

- Applying this principle within *both $\mathscr{L}_1$ and $\mathscr{L}_2$* yields:

(1) $\Pr[Da \mid Db \,\&\, (Aa \,\&\, Ba \,\&\, Ca \,\&\, Ab \,\&\, {\sim}Bb \,\&\, Cb)] > \Pr[Da \mid Db \,\&\, (Aa \,\&\, Ba \,\&\, Ca \,\&\, Ab \,\&\, {\sim}Bb \,\&\, {\sim}Cb)]$

(2) $\Pr[Da \mid Db \,\&\, (Aa \,\&\, Xa \,\&\, Ya \,\&\, {\sim}Ab \,\&\, Xb \,\&\, Yb)] > \Pr[Da \mid Db \,\&\, (Aa \,\&\, Xa \,\&\, Ya \,\&\, Ab \,\&\, {\sim}Xb \,\&\, {\sim}Yb)]$

- It is useful here to adopt the following *abbreviations*:

$\mathsf{x} \overset{\text{def}}{=} Db \,\&\, (Aa \,\&\, Ba \,\&\, Ca \,\&\, Ab \,\&\, {\sim}Bb \,\&\, Cb)$ $[\mathsf{x} \overset{\text{def}}{=} Db \,\&\, \mathfrak{s}_{\mathscr{L}_1}(a,b) = 2]$

$\mathsf{y} \overset{\text{def}}{=} Db \,\&\, (Aa \,\&\, Xa \,\&\, Ya \,\&\, Ab \,\&\, {\sim}Xb \,\&\, {\sim}Yb)$ $[\mathsf{y} \overset{\text{def}}{=} Db \,\&\, \mathfrak{s}_{\mathscr{L}_2}(a,b) = 1]$

$\mathsf{z} \overset{\text{def}}{=} Db \,\&\, (Aa \,\&\, Xa \,\&\, Ya \,\&\, {\sim}Ab \,\&\, Xb \,\&\, Yb)$ $[\mathsf{z} \overset{\text{def}}{=} Db \,\&\, \mathfrak{s}_{\mathscr{L}_2}(a,b) = 2]$

$\mathsf{u} \overset{\text{def}}{=} Db \,\&\, (Aa \,\&\, Ba \,\&\, Ca \,\&\, Ab \,\&\, {\sim}Bb \,\&\, {\sim}Cb)$ $[\mathsf{u} \overset{\text{def}}{=} Db \,\&\, \mathfrak{s}_{\mathscr{L}_1}(a,b) = 1]$

- Thus, we have $\mathsf{x} \dashv\vDash \mathsf{y}$ and $\mathsf{z} \dashv\vDash \mathsf{u}$, but ($\mathscr{A}$) implies *both* (1) $\Pr(Da \mid \mathsf{x}) > \Pr(Da \mid \mathsf{u})$, *and* (2) $\Pr(Da \mid \mathsf{z}) > \Pr(Da \mid \mathsf{y})$.
- This violates the axioms of conditional probability.

☞ *Either* ($\mathscr{A}$) must go, *or* we need some *restriction on the choice of language* — in order to block inferring (1) *and* (2).

- Suppose we have a panel of three judges (J_1, J_2, J_3). This panel will vote on an *agenda*, which stems from:

  Question. In the reunited Germany, should the German parliament and the seat of government move to Berlin or stay in Bonn?

- Suppose the panel votes on these two (atomic) *premises*:
  - $P \overset{\text{def}}{=}$ the parliament should move.
  - $G \overset{\text{def}}{=}$ the seat of government should move.
- There is also the following "*conclusion*" whose truth-value is *determined by* the truth-values of the premises:
  - $B \overset{\text{def}}{=}$ both the parliament and the seat of government should move.
- Suppose the judges render the following judgments (votes):

	P?	G?	B?
J_1	yes	no	no
J_2	no	yes	no
J_3	yes	yes	yes

- For each judge, the conclusion column is *determined by* the premise columns (*i.e.*, we assume each judge is *consistent*).

- Example of *doctrinal paradox/discursive dilemma* ([6], [13]).

Doctrinal Paradox/Discursive Dilemma

	P?	G?	B?
J_1	yes	no	no
J_2	no	yes	no
J_3	yes	yes	yes
Majority	yes	yes	**yes** *&* **no**?

☞ *Naive majority rule* for aggregating *all* judgments can lead to *inconsistent aggregations of premises + conclusions*.

- Various alternative aggregation procedures have been proposed, so as to ensure overall consistency. Example:
  - **Premise-Based Procedure**. Use majority rule on the premises, and then just *enforce the **logical** conclusion*.
- The premise-based procedure seems reasonable (esp. if the premises make up the agenda that is explicitly voted on).

Premise-Based Procedure

	P?	G?	B?
J_1	yes	no	no
J_2	no	yes	no
J_3	yes	yes	yes
Majority:	yes	yes	(ignore)
Logical Conclusion:			**yes**

- So, the premise-based procedure is a way of restoring logical consistency to a naive majority aggregation rule.
- But, the premise-based procedure (and *many* other procedures) faces another problem: *language dependence*.
- To see this, look at what happens when we move from the PG–language ($\mathscr{L}_1$) to the following PY–language ($\mathscr{L}_2$), where Y satisfies the following (Miller-style) semantic constraint:
  - $Y \dashv\vDash P \equiv G$.

Language Dependence of the Premise-Based Procedure

$\mathscr{L}_1$:	P	G	$P \equiv G$	$P \,\&\, G$
$\mathscr{L}_2$:	P	$P \equiv Y$	Y	$P \,\&\, Y$
J_1	yes	no	no	no
J_2	no	yes	no	no
J_3	yes	yes	yes	yes
Majority in $\mathscr{L}_1$:	yes	yes		**yes**
Majority in $\mathscr{L}_2$:	yes		no	**no**

☞ The aggregate conclusion judgment depends on which (expressively equivalent) language we use to express the premises (*i.e.*, the agenda).

- This shows that the premise-based procedure is language dependent, in basically the same sense that our accounts of verisimilitude and inductive probability (above) were.
- Which (consistent) procedures are language *in*dependent?

- A recent paper by Cariani (one of my former students, now at NU) *et. al.* gives a precise answer to this question [2].
- To understand their central theorem, I'll need to introduce a little more terminology. Two properties of procedures:
  - **Decisiveness**. An aggregation procedure is *decisive* iff it is *consistent and complete (and, hence, deductively closed)*.
  - **Anonymity**. An aggregation procedure is *anonymous* iff it is invariant under permutations of the (names of the) judges.
- Consistency and closure are basic logical constraints. Completeness is stronger than closure (given consistency).
- Anonymity basically says that *it shouldn't matter who the judges are* (*i.e.*, *which judgment profiles belong to which judges*). This is a substantive (non-logical) assumption. But, it's often plausible (*e.g.*, in our German Capital case).
- I will return to the completeness assumption, below.
- But, first, here is their new impossibility theorem [2].

Central Impossibility Theorem of Cariani *et. al.* [2]

Let n be the number of judges on the panel, and k be the number of atomic sentences in our language(s) $\mathscr{L}$. Provided that $k \geq \log_2(n+2)$, *no* decisive aggregation procedure can be *both*:

(i) anonymous,
and
(ii) language independent.

☞ In other words, for languages with sufficiently many atomic sentences, *it is impossible for any aggregation procedure to be decisive, anonymous, and language independent.*

  - In fact, all decisive, language independent aggregation procedures must be *rolling dictatorships* [2].

- ∴ *Most of the judgment aggregation procedures that have been proposed in the literature are language dependent.*
- There are various ways to avoid this. *E.g.*, *incompleteness*.

- Incompleteness may make sense in some contexts [7]:

  > If a group is prepared to refrain from making a collective judgment on some propositions . . . then it may use an aggregation procedure such as the 'unanimity procedure', whereby the group makes a judgment on a proposition iff the group members unanimously endorse that judgment. Propositions judged to be true by all members are collectively judged to be true; and ones judged to be false by all members are collectively judged to be false; *no collective judgment is made on any other propositions.*

- Other incomplete/$\mathscr{L}$–independent procedures (see [8] for survey):
  - A **conclusion-based** procedure, which returns no aggregate judgment for the premises. In our example, it returns "no" for the conclusion and it returns *nothing* for the premises.
  - **Approval voting** [1], which returns as the group judgment set the most popular of the voters' judgment sets (where judges can vote on *any number* of agenda items that they like). In our example, it returns *all* the judgment sets (that is to say, *no definitive aggregate judgment whatsoever*).

- So far, we've seen "logical"/"syntactical" manifestations of LD. But, the underlying phenomenon is *much more general*.
- As Popper [14, Appendix 2] and Miller [10, Ch. 11] show, the problem also arises in the context of *quantitative* theories.
- For instance, suppose we have two (real-valued) parameters ϕ and ψ, and we have two hypotheses H_1 and H_2, each of which implies estimates/predictions of their values. To wit:

	ϕ	ψ	α	β
H_1	0.150	1.225	0.925	2.000
H_2	0.100	1.000	0.800	1.700
T	0.000	1.000	1.000	2.000

- Here, H_2 is "closer to the truth" (T) *with respect to both* ϕ and ψ. But, this *reverses with respect to* α and β, where:

$$\alpha = \psi - 2\phi \qquad \beta = 2\psi - 3\phi$$
$$\phi = \beta - 2\alpha \qquad \psi = 2\beta - 3\alpha$$

☞ The ϕ/ψ and α/β languages are *expressively equivalent*. But, assessments of "closeness to T" *depend on language*.

- Jim Joyce [5] has recently argued for probabilism, by appeal to the following sort of "verisimilitude theorem".
  - **Theorem** (Joyce). If S's credence function b (construed as providing *estimates of truth-values*) fails to be probabilistic, then there exists a *probabilistic* estimate b' of truth-values that is *closer to the truth than* b*, in all possible worlds.*
- As it turns out, one can also prove the following (*via* translations slightly more complex than Popper-Miller's).
  - **Theorem** (Me). For any pair of credence functions $\langle b, b' \rangle$ of the sort mentioned in Joyce's theorem (and for any way of measuring "closeness" you like), there exists a *translation* τ such that $\tau(b)$ is *closer to the truth than* $\tau(b')$ *— with respect to the "*τ*-truth-values" — in all possible worlds.*

☞ There are *some aspects of* "the truth" with respect to which (any) incoherent b is bound to be less accurate than (some) coherent b', but — for any such pair $\langle b, b' \rangle$ — there are also *some aspects of* "the truth" on which *the opposite is the case.*

- Why can we *ignore some aspects of* the truth, but not others?

[1] S. Brams and P. Fishburn, *Approval Voting* (2ed), Springer, 2007.

[2] F. Cariani, M. Pauly and J. Snyder, *Decision Framing in Judgment Aggregation*, *Synthese*, 2008.

[3] R. Carnap, *A Basic System of Inductive Logic, Parts I & II.* In R. Carnap & R. Jeffrey (*eds.*), *Studies in Inductive Logic and Probability, Vols I & II*, University of California Press, 1971 & 1980.

[4] M. Hesse, *Analogy and Confirmation Theory*, *Philosophy of Science*, 1964.

[5] J. Joyce, A*ccuracy and coherence: Prospects for an alethic epistemology of partial belief*, in *Degrees of Belief*, F. Huber & C. Schmidt-Petri (*eds.*), 2009.

[6] L. Kornhauser and G. Sager, *Unpacking the Court*, *The Yale Law Journal*, 1986.

[7] C. List, *Group knowledge and group rationality: a judgment aggregation perspective*, *Episteme*, 2005.

[8] C. List and B. Polak, *Introduction to Judgment Aggregation*, *Journal of Economic Theory*, 2010.

[9] D. Miller, *The Accuracy of Predictions*, *Synthese*, 1975.

[10] D. Miller, *Out Of Error: Further Essays on Critical Rationalism*, Ashgate, 2006.

[11] I. Niiniluoto, *Truthlikeness*, *Springer*, 1987.

[12] G. Oddie, *Truthlikeness*, *Stanford Encyclopedia of Philosophy*, 2007.

[13] P. Petit, *Deliberative Democracy and the Discursive Dilemma*, *Philosophical Issues*, 2001.

[14] K. Popper, *Objective Knowledge: An Evolutionary Approach*, second edition, 1979.

[15] P. Tichý, *On Popper's definitions of verisimilitude*, *British J. for the Phil. of Science*, 1974.

[16] P. Tichý, *Verisimilitude Redefined*, *British J. for the Phil. of Science*, 1976.