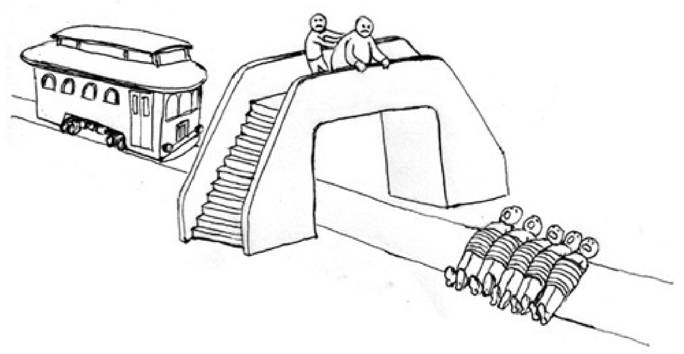
Reverse inference, Bayesian confirmation, and the neuroscience of moral reasoning

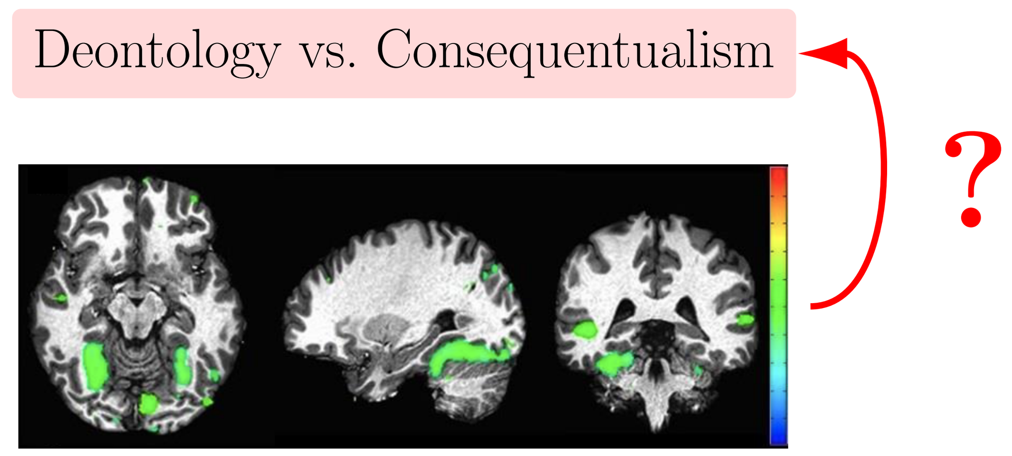
Gustavo Cevolani and Davide Coraci

IMT School for Advanced Studies Lucca, Italy — MoMiLab (Molecular Mind Laboratory)
{gustavo.cevolani,davide.coraci}@imtlucca.it

The "normativity" problem

"Footbridge" trolley dilemma





What is the normative relevance of neuroscientific evidence? [1]

The "inference problem"

"Reverse inference" (RI) in fMRI research [3]:

(P1) In the literature, when cognitive process P is engaged, brain area A is active;
(P2) In the present study, brain area A is active;

(C) Therefore, in the present study the cognitive process P in engaged.

- logically invalid (fallacy of "affirming the consequent")
- lack of selectivity problem (same area A active for many different processes $P, P', \ldots$)
- defensible as a form of "abductive" reasoning (from effects to causes or explanations)

Bayesian analysis of reverse inference

$$p(P|A) = \frac{p(A|P) \times p(P)}{p(A|P)p(P) + p(A|\neg P)p(\neg P)}$$

- $p(P|A)$ = posterior probability of P engaged given A active;
- $p(A|P)$ and $p(A|\neg P)$= likelihoods of P vs. $\neg P$ given A;
- $p(P)$ and $p(\neg P)$ = prior probabilities of engagement (e.g., set at 0.5).

Bayes factor (BF)

A common "support" measure:

$$BF(P, A) = \frac{o(P|A)}{o(P)} = \frac{p(A|P)}{p(A|\neg P)}$$

where $o(P) = \dfrac{p(P)}{p(\neg P)}$ are the odds of P

Bayesian confirmation

Confirmation as measure of evidential support [2, 4]:

- Hypothesis H is confirmed by evidence E iff
$$p(H|E) > p(H)$$
(Carnap: confirmation as increase in probability)
- BF is a measure of confirmation

Reverse inference as Bayesian confirmation

Given current research practice, RI seems best construed in terms of Bayesian confirmation; but: confirmation is different from probability, i.e.:

- $p(P|A)$ may be high even if $BF(P, A)$ is low (P is probable but not confirmed)
- $BF(P, A)$ may be high even if $p(P|A)$ is low (P is confirmed but still not likely)

References

[1] S. Berker. "The normative insignificance of neuroscience". In: *Philosophy & Public Affairs* 37.4 (2009), pp. 293–329.
[2] V. Crupi. "Confirmation". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. 2020.
[3] R. A. Poldrack. "Can cognitive processes be inferred from neuroimaging data?" In: *Trends in Cognitive Sciences* 10.2 (2006), pp. 59–63.
[4] J. Sprenger and S. Hartmann. *Bayesian philosophy of science*. OUP, 2019.