

Coherent testing and safe testing

Royal Statistical Society discussion of “Safe testing” by
Peter Grünwald, Rianne de Heide and Wouter Koolen

January 24, 2024

David R. Bickel

Informatics and Analytics
University of North Carolina at Greensboro

davidbickel.com

What does science need?

- **Safe** testing!
 - Practical tool for replicable research
 - Checking Bayesian models <https://bit.ly/3vSmlIh>
 - Paradigm shift in the foundations of statistics?
- **Coherent** testing!
 - Local-global coherence
 - Logical coherence

Consistency and coherence

- Logical **consistency**

- $\neg (A \text{ or } B) \Leftrightarrow \neg A \text{ and } \neg B$

- Logical **coherence**

- \exists substantial evidence against $(A \text{ or } B)$

\Leftrightarrow

\exists substantial evidence against A
and
 \exists substantial evidence against B

Fossaluza et al. (2017)

"Coherent hypothesis testing,"
American Statistician **71**, 242-248

Hansen and Rice (2023) "Coherent
tests for interval null hypotheses,"
American Statistician **77**, 20-28

null hypotheses

- A
- B
- $A \text{ or } B$

Coherent e-values

- Logically coherent e-variable
- $E_{A \text{ or } B} \leq \min(E_A, E_B)$ almost surely
- Logically coherent e-variable of **maximum power**
- $E_{A \text{ or } B} = \min(E_A, E_B)$ almost surely
- **e-possibility** measure (Vovk and Wang, §5)

null hypotheses

- A
- B
- $A \text{ or } B$

- Wasserman et al. (2020) "Universal inference" *PNAS* **117**, 16880-90
- Shafer (2021) "Testing by betting ..." *JRSS A* **184**, 407-431
- Vovk and Wang (2023) "Confidence and discoveries with e-values," *Stat. Sci.* **38**, 329-354

GRO e-values

- GRO e-variable (discussion paper)
 - More **powerful** than e-possibility
 - $E_{A \text{ or } B} > \min(E_A, E_B)$ almost surely (in those cases)
 - \exists substantial evidence against $(A \text{ or } B)$
 \Rightarrow
 \exists substantial evidence against A
and
 \exists substantial evidence against B
- **Power-coherence tradeoff** among safe tests

Ramdas et al. (2023) "Game-theoretic statistics and safe anytime-valid inference," *Stat. Sci.* **38**, 576-601

"**Odd.**" —Royall (1997) *Statistical Evidence: A Likelihood Paradigm*. CRC Press, New York, §3.7

null hypotheses

- A
- B
- $A \text{ or } B$