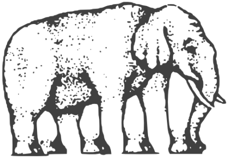
Measuring coherence in reasoning

Nicole Cruz
University of Innsbruck



Reasoning and uncertainty:
probabilistic, logical, and psychological perspectives
10th August 2022, Regensburg, Germany & online

Outline

1. What is coherence and why is it relevant?
2. How is coherence measured?
3. How can sensitivity to coherence be measured?

What is coherence and why is it relevant?

Coherence: Generalisation of logical consistency

From binary truth / falsity to probabilities

- Consistency: The truth values assigned to two statements are consistent iff they can both be true (or both false) without creating a contradiction.
- Coherence: The numeric values assigned to two statements are coherent (and are hence probabilities) iff they follow the axioms of (classical) probability theory.
- The axioms of probability are followed iff there is no risk of a Dutch book: A series of bets on logically interrelated events that leads to a sure loss to one side.

(de Finetti, 1937/1980; Ramsey 1926/1990; Stalnaker, 1970; Vineberg, 2022).

Coherence: Generalisation of logical consistency

From binary truth / falsity to probabilities

- Consistency: The truth values assigned to two statements are consistent iff they can both be true (or both false) without creating a contradiction.
- Coherence: The numeric values assigned to two statements are coherent (and are hence probabilities) iff they follow the axioms of (classical) probability theory.
- The axioms of probability are followed iff there is no risk of a Dutch book: A series of bets on logically interrelated events that leads to a sure loss to one side.

(de Finetti, 1937/1980; Ramsey 1926/1990; Stalnaker, 1970; Vineberg, 2022).

Coherence: Generalisation of logical consistency

From binary truth / falsity to probabilities

- Consistency: The truth values assigned to two statements are consistent iff they can both be true (or both false) without creating a contradiction.
- Coherence: The numeric values assigned to two statements are coherent (and are hence probabilities) iff they follow the axioms of (classical) probability theory.
- The axioms of probability are followed iff there is no risk of a Dutch book: A series of bets on logically interrelated events that leads to a sure loss to one side.

(de Finetti, 1937/1980; Ramsey 1926/1990; Stalnaker, 1970; Vineberg, 2022).

Why have coherent beliefs?

Coherence...

- Helps us advance towards our goals & reduce losses.
- Is foundation for knowledge & understanding.
- Does not apply only to formal bets.

> *It is based fundamentally on betting, but this will not seem unreasonable when it is seen that all our lives we are in a sense betting. Whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home.* *(Ramsey, 1926/1990, p. 23).*

(de Finetti, 1937/1980; Good, 1971; Ramsey 1926/1990; Stalnaker, 1970; Vineberg, 2022).

How is coherence measured?

Coherence intervals for one-premise inferences

Consider an inference with some initial information, or *premise*, from which a *conclusion* is drawn. P(premise) = x. How does this constrain P(conclusion)?

- Equivalences
  - *Not(T & C). Therefore not-T or not-C.*
  - P(not both tea & coffee)=.8. $\Rightarrow$ P(not-tea or not-coffee)=.8.
- Contradictions
  - *T. Therefore not-T.*
  - P(tea)=.6. $\Rightarrow$ P(not-tea)=.4.
- Set-subset relations
  - *T or C. Therefore C.*
  - P(tea or coffee)=.4. $\Rightarrow$ P(coffee) $\in$ [0, .4].

$\rightarrow$ C.f. conjunction & disjunction fallacies

(Tversky & Kahneman, 1983; Bar-Hillel & Neter, 1993).

Coherence intervals for one-premise inferences

Consider an inference with some initial information, or *premise*, from which a *conclusion* is drawn. P(premise) = x. How does this constrain P(conclusion)?

- Equivalences
  - *Not(T & C). Therefore not-T or not-C.*
  - P(not both tea & coffee)=.8. $\Rightarrow$ P(not-tea or not-coffee)=.8.
- Contradictions
  - *T. Therefore not-T.*
  - P(tea)=.6. $\Rightarrow$ P(not-tea)=.4.
- Set-subset relations
  - *T or C. Therefore C.*
  - P(tea or coffee)=.4. $\Rightarrow$ P(coffee) $\in$ [0, .4].

  $\rightarrow$ C.f. conjunction & disjunction fallacies

(Tversky & Kahneman, 1983; Bar-Hillel & Neter, 1993).

Coherence intervals for one-premise inferences

Consider an inference with some initial information, or *premise*, from which a *conclusion* is drawn. P(premise) = x. How does this constrain P(conclusion)?

- Equivalences
  - *Not(T & C). Therefore not-T or not-C.*
  - P(not both tea & coffee)=.8. $\Rightarrow$ P(not-tea or not-coffee)=.8.
- Contradictions
  - *T. Therefore not-T.*
  - P(tea)=.6. $\Rightarrow$ P(not-tea)=.4.
- Set-subset relations
  - *T or C. Therefore C.*
  - P(tea or coffee)=.4. $\Rightarrow$ P(coffee) $\in$ [0, .4].

$\rightarrow$ C.f. conjunction & disjunction fallacies

(Tversky & Kahneman, 1983; Bar-Hillel & Neter, 1993).

Coherence intervals for one-premise inferences

Consider an inference with some initial information, or *premise*, from which a *conclusion* is drawn. P(premise) = x. How does this constrain P(conclusion)?

- Equivalences
  - *Not(T & C). Therefore not-T or not-C.*
  - P(not both tea & coffee)=.8. $\Rightarrow$ P(not-tea or not-coffee)=.8.
- Contradictions
  - *T. Therefore not-T.*
  - P(tea)=.6. $\Rightarrow$ P(not-tea)=.4.
- Set-subset relations
  - *T or C. Therefore C.*
  - P(tea or coffee)=.4. $\Rightarrow$ P(coffee) $\in$ [0, .4].

$\rightarrow$ C.f. conjunction & disjunction fallacies

(Tversky & Kahneman, 1983; Bar-Hillel & Neter, 1993).

Coherence intervals for one-premise inferences

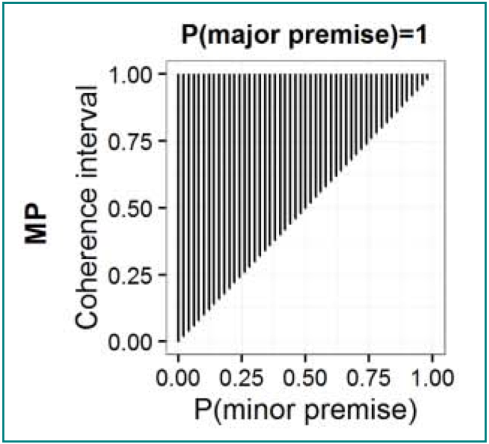
Consider an inference with some initial information, or *premise*, from which a *conclusion* is drawn. P(premise) = x. How does this constrain P(conclusion)?

- Equivalences
  - *Not(T & C). Therefore not-T or not-C.*
  - P(not both tea & coffee)=.8. $\Rightarrow$ P(not-tea or not-coffee)=.8.
- Contradictions
  - *T. Therefore not-T.*
  - P(tea)=.6. $\Rightarrow$ P(not-tea)=.4.
- Set-subset relations
  - *T or C. Therefore C.*
  - P(tea or coffee)=.4. $\Rightarrow$ P(coffee) $\in$ [0, .4].

  $\rightarrow$ C.f. conjunction & disjunction fallacies

(Tversky & Kahneman, 1983; Bar-Hillel & Neter, 1993).

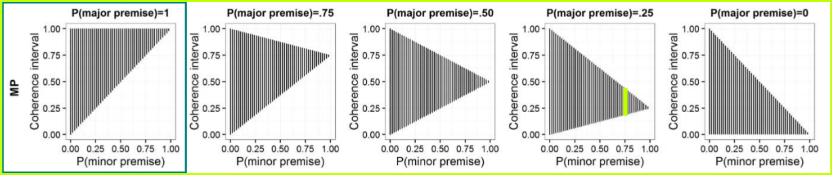
Coherence intervals for complexer inferences



If T then C (major). T (minor). $\Rightarrow$ C (conclusion).
Given P($C|T$) and P(T), P(C) $\in$ [P($C|T$)*P(T), P($C|T$)*P(T) + (1 - P(T))].

(Cruz, 2018).

Coherence intervals for complexer inferences



If T then C. T. $\Rightarrow$ *C.*

Given $\mathrm{P}(C|T) = .25$ and $\mathrm{P}(T) = .75$,

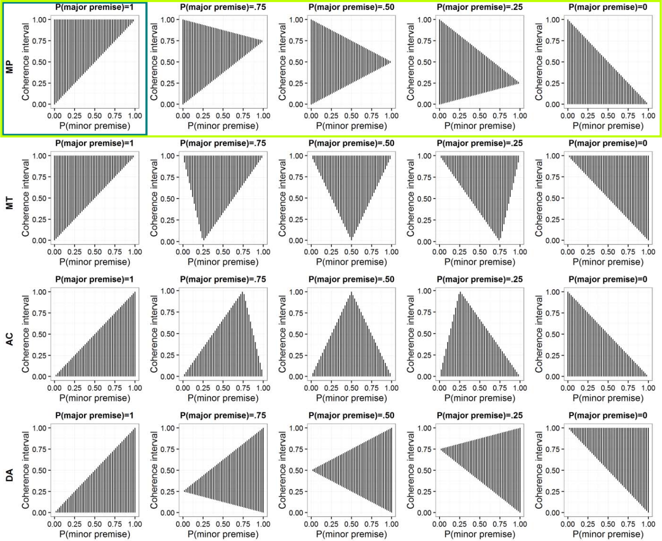
$\mathrm{P}(C) \in [\mathrm{P}(C|T)*\mathrm{P}(T), \mathrm{P}(C|T)*\mathrm{P}(T) + (1 - \mathrm{P}(T))]$.

$\mathrm{P}(C) \in [.25*.75, .25*.75 + (1 - .75]$.

$\mathrm{P}(C) \in [0.1875, 0.4375]$.

(Cruz, 2018).

Coherence intervals for complexer inferences

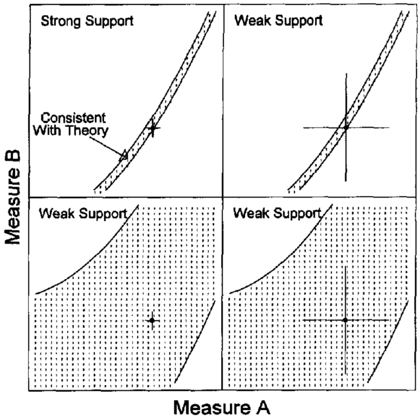


(Cruz, 2018).

How can sensitivity to coherence be measured?

Informative tests & plausible falsifiability

"Four possible relationships between theory and data. [...] Only when both theory and data provide substantial constraints does this provide significant evidence for the theory."



(Roberts & Pashler, 2000; Vanpaemel, 2020).

Chance & above-chance rates

- Above-chance coherence: Commonly measured as observed coherence rate - coherence interval width (Evans et al., 2015).
- How good is this measure and why? How does it compare to alternatives?
- For above-chance coherence to be detectable, the chance rate must be sufficiently low.

P(bank-teller)=1 $\Rightarrow$ P(bank-teller & feminist) $\in$ [0, 1].

(Cruz, 2018; Evans et al., 2015; Politzer, 2015; Singmann et al., 2014).

Chance & above-chance rates

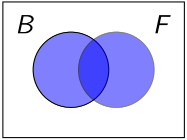
- Above-chance coherence: Commonly measured as observed coherence rate - coherence interval width (Evans et al., 2015).
- How good is this measure and why? How does it compare to alternatives?
- For above-chance coherence to be detectable, the chance rate must be sufficiently low.

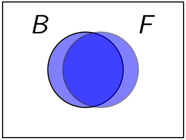
P(bank-teller)=1 $\Rightarrow$ P(bank-teller & feminist) $\in$ [0, 1].

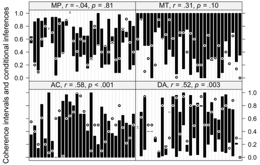
(Cruz, 2018; Evans et al., 2015; Politzer, 2015; Singmann et al., 2014).

Chance & above-chance rates

- Above-chance coherence: Commonly measured as observed coherence rate - coherence interval width (Evans et al., 2015).
- How good is this measure and why? How does it compare to alternatives?
- For above-chance coherence to be detectable, the chance rate must be sufficiently low.



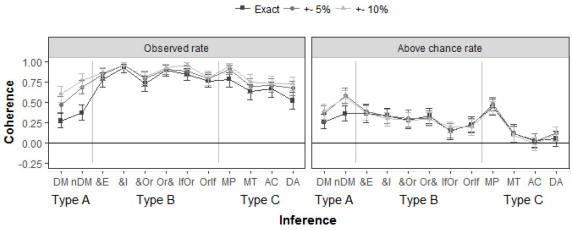




P(bank-teller)=1 $\Rightarrow$ P(bank-teller & feminist) $\in$ [0, 1].

(Cruz, 2018; Evans et al., 2015; Politzer, 2015; Singmann et al., 2014).

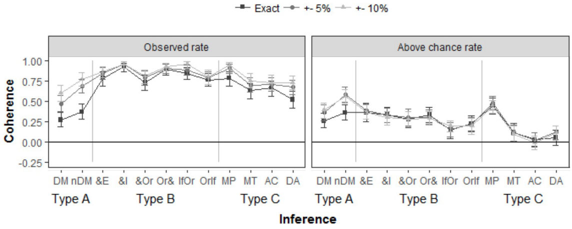
Above-chance $\neq$ high



- People's responses in reasoning tasks are typically coherent above chance $\rightarrow$ evidence of sensitivity to coherence.
- But coherent to what extent?
- How can it be quantified when coherent responses are determined by intervals rather than points?

(Costello & Watts, 2018; Cruz, 2018; Klauer et al., 2010; Oaksford et al., 2000; Politzer & Baratgin, 2016).

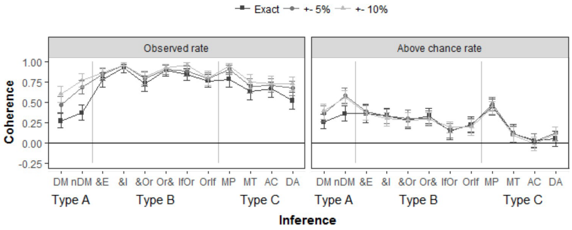
Above-chance $\neq$ high



- People's responses in reasoning tasks are typically coherent above chance $\rightarrow$ evidence of sensitivity to coherence.
- But coherent to what extent?
- How can it be quantified when coherent responses are determined by intervals rather than points?

(Costello & Watts, 2018; Cruz, 2018; Klauer et al., 2010; Oaksford et al., 2000; Politzer & Baratgin, 2016).

Above-chance $\neq$ high



- People's responses in reasoning tasks are typically coherent above chance $\rightarrow$ evidence of sensitivity to coherence.
- But coherent to what extent?
- How can it be quantified when coherent responses are determined by intervals rather than points?

(Costello & Watts, 2018; Cruz, 2018; Klauer et al., 2010; Oaksford et al., 2000; Politzer & Baratgin, 2016).

From possible to plausible falsifiability

From comparisons against chance to comparisons between theories

- Does coherence differ between statement interpretations?
  - P(*if coin flipped then heads*) = P(*heads*|*flipped*) (probabilistic approaches)
  - P(*if coin flipped then heads*) = P(*heads*|*flipped*) - P(*heads*) (relevance-based approaches)
  - P(*if coin flipped then heads*) = P(*not-heads* or *flipped*) (classical logic)
- Does coherence differ as a function of which inferences are considered deductively valid (c.f. contraposition, centering, or-to-if, transitivity)?

(Crupi et al., 2007; Cruz et al., 2015; Over & Cruz, 2018; Rott, 2019; Skovgaard-Olsen et al., 2017).

From possible to plausible falsifiability

From comparisons against chance to comparisons between theories

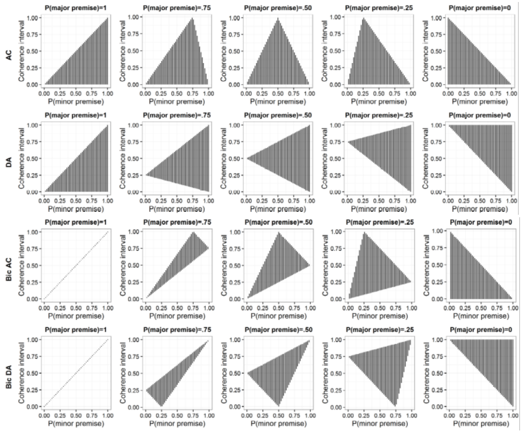
- Does coherence differ between statement interpretations?
  - P(*if coin flipped then heads*) = P(*heads*|*flipped*) (probabilistic approaches)
  - P(*if coin flipped then heads*) = P(*heads*|*flipped*) - P(*heads*) (relevance-based approaches)
  - P(*if coin flipped then heads*) = P(*not-heads* or *flipped*) (classical logic)
- Does coherence differ as a function of which inferences are considered deductively valid (c.f. contraposition, centering, or-to-if, transitivity)?

(Crupi et al., 2007; Cruz et al., 2015; Over & Cruz, 2018; Rott, 2019; Skovgaard-Olsen et al., 2017).

Is *if* interpreted as conditional or biconditional?

Conditional:

Biconditional:



Drawing on correlation vs. independence between *if* and *then* for communication and decision making.

(Cruz, 2018; Cruz & Over, in press; Lassiter, in press).

Risk of researcher incoherence through precision

- Further information e.g. about the *if-then* correlation makes it possible to narrow down intervals to points.
- But not all possible correlations will be coherent.
- If the researcher's benchmark coherence calculation is incoherent, then any comparisons with that benchmark will be uninformative: garbage-in, garbage-out.
- Example: for $P(q|p) = .8$ and P(categorical premise) = .6, $P(q|\textit{not-p})$ will be constrained as follows for the four syllogisms: MP: [0, 1], DA: [0, 1], MT; [0, .4], and AC: [0, .6]. This means that if e.g. $P(q|\textit{not-p}) = .8$, then the input to the coherence formulas for MT and AC will be incoherent, rendering their output uninterpretable.

(Singmann et al., 2014).

Risk of researcher incoherence through precision

- Further information e.g. about the *if-then* correlation makes it possible to narrow down intervals to points.
- But not all possible correlations will be coherent.
- If the researcher's benchmark coherence calculation is incoherent, then any comparisons with that benchmark will be uninformative: garbage-in, garbage-out.
- Example: for P($q|p$) = .8 and P(categorical premise) = .6, P($q|$*not-p*) will be constrained as follows for the four syllogisms: MP: [0, 1], DA: [0, 1], MT; [0, .4], and AC: [0, .6]. This means that if e.g. P($q|$*not-p*) = .8, then the input to the coherence formulas for MT and AC will be incoherent, rendering their output uninterpretable.

(Singmann et al., 2014).

Risk of researcher incoherence through precision

- Further information e.g. about the *if-then* correlation makes it possible to narrow down intervals to points.
- But not all possible correlations will be coherent.
- If the researcher's benchmark coherence calculation is incoherent, then any comparisons with that benchmark will be uninformative: garbage-in, garbage-out.
- Example: for $P(q|p) = .8$ and P(categorical premise) $= .6$, $P(q|not\text{-}p)$ will be constrained as follows for the four syllogisms: MP: [0, 1], DA: [0, 1], MT; [0, .4], and AC: [0, .6]. This means that if e.g. $P(q|not\text{-}p) = .8$, then the input to the coherence formulas for MT and AC will be incoherent, rendering their output uninterpretable.

(Singmann et al., 2014).

Risk of researcher incoherence through precision

- Further information e.g. about the *if-then* correlation makes it possible to narrow down intervals to points.
- But not all possible correlations will be coherent.
- If the researcher's benchmark coherence calculation is incoherent, then any comparisons with that benchmark will be uninformative: garbage-in, garbage-out.
- Example: for $P(q|p) = .8$ and P(categorical premise) $= .6$, $P(q|\textit{not-p})$ will be constrained as follows for the four syllogisms: MP: [0, 1], DA: [0, 1], MT; [0, .4], and AC: [0, .6]. This means that if e.g. $P(q|\textit{not-p}) = .8$, then the input to the coherence formulas for MT and AC will be incoherent, rendering their output uninterpretable.

(Singmann et al., 2014).

Risk of researcher incoherence through *im*precision

- Coherence intervals depend on (a) the logical structure of an inference (likelihood), and (b) the premise probabilities (priors).
- Negations are part of the logical structure of an inference.
  - MT (*if p then q, not-q, therefore not-p*): If the child is crying then it is sad. The child is not sad. Therefore, the child is not crying.
  - **Not** MT (*if p then q, r, therefore s*): If the child is crying then it is sad. The child is happy. Therefore, the child is laughing.
- Some "negation-effects" in the literature may be an artefact resulting from comparing apples with oranges.

(Trippas et al., 2016).

Risk of researcher incoherence through *im*precision

- Coherence intervals depend on (a) the logical structure of an inference (likelihood), and (b) the premise probabilities (priors).
- Negations are part of the logical structure of an inference.
  - MT (*if p then q, not-q, therefore not-p*): If the child is crying then it is sad. The child is not sad. Therefore, the child is not crying.
  - **Not** MT (*if p then q, r, therefore s*): If the child is crying then it is sad. The child is happy. Therefore, the child is laughing.
- Some "negation-effects" in the literature may be an artefact resulting from comparing apples with oranges.

(Trippas et al., 2016).

Risk of researcher incoherence through *im*precision

- Coherence intervals depend on (a) the logical structure of an inference (likelihood), and (b) the premise probabilities (priors).
- Negations are part of the logical structure of an inference.
  - MT (*if p then q, not-q, therefore not-p*): If the child is crying then it is sad. The child is not sad. Therefore, the child is not crying.
  - **Not** MT (*if p then q, r, therefore s*): If the child is crying then it is sad. The child is happy. Therefore, the child is laughing.
- Some "negation-effects" in the literature may be an artefact resulting from comparing apples with oranges.

(Trippas et al., 2016).

Risk of researcher incoherence through *im*precision

- Coherence intervals depend on (a) the logical structure of an inference (likelihood), and (b) the premise probabilities (priors).
- Negations are part of the logical structure of an inference.
  - MT (*if p then q, not-q, therefore not-p*): If the child is crying then it is sad. The child is not sad. Therefore, the child is not crying.
  - **Not** MT (*if p then q, r, therefore s*): If the child is crying then it is sad. The child is happy. Therefore, the child is laughing.
- Some "negation-effects" in the literature may be an artefact resulting from comparing apples with oranges.

(Trippas et al., 2016).

Trying to link theory & measurement

Generic resources

- Computational cognitive modelling (e.g. hierarchical Bayesian, distribution-free methods, reinforcement learning).
- Sensitivity analysis: finding (plausible) data patterns that would disconfirm our theories.
- More open, accessible science, interdisciplinary collaboration.

Generic limitations

- Coherence applies only at a fixed point in time.
- Principles & background assumptions for dynamic reasoning & belief updating (Jeffrey conditionalisation; KL-divergence, Bregman divergence, Total divergence norm)?

(Brozzi, Capotorti, & Vantaggi, 2012; Chechile, 2020, Cruz, 2018; Dunn & Anderson, 2018; Hadjichristidis et al., 2014; Lee, 2018; Oaksford & Chater, 2013; Pearl, 2000; Zhao & Osherson, 2010).

Trying to link theory & measurement

Generic resources

- Computational cognitive modelling (e.g. hierarchical Bayesian, distribution-free methods, reinforcement learning).
- Sensitivity analysis: finding (plausible) data patterns that would disconfirm our theories.
- More open, accessible science, interdisciplinary collaboration.

Generic limitations

- Coherence applies only at a fixed point in time.
- Principles & background assumptions for dynamic reasoning & belief updating (Jeffrey conditionalisation; KL-divergence, Bregman divergence, Total divergence norm)?

(Brozzi, Capotorti, & Vantaggi, 2012; Chechile, 2020, Cruz, 2018; Dunn & Anderson, 2018; Hadjichristidis et al., 2014; Lee, 2018; Oaksford & Chater, 2013; Pearl, 2000; Zhao & Osherson, 2010).

Thank you!