

Inference under the entropy-maximizing Bayesian model of sufficient evidence

The Third International Conference on Mathematical and
Computational Medicine

18 May 2016

David Bickel
University of Ottawa



Loss function example

$$\ell(\theta, \delta) = \begin{cases} 0 & \text{if } \delta = \theta \in \{0, 1\} \\ \ell_{\text{I}} & \text{if } \delta = 1, \theta = 0 \\ \ell_{\text{II}} & \text{if } \delta = 0, \theta = 1 \end{cases}$$

What act is appropriate?

- Given **multiple priors** of sufficient evidence, agreement with the data
- With or without **confidence sets**

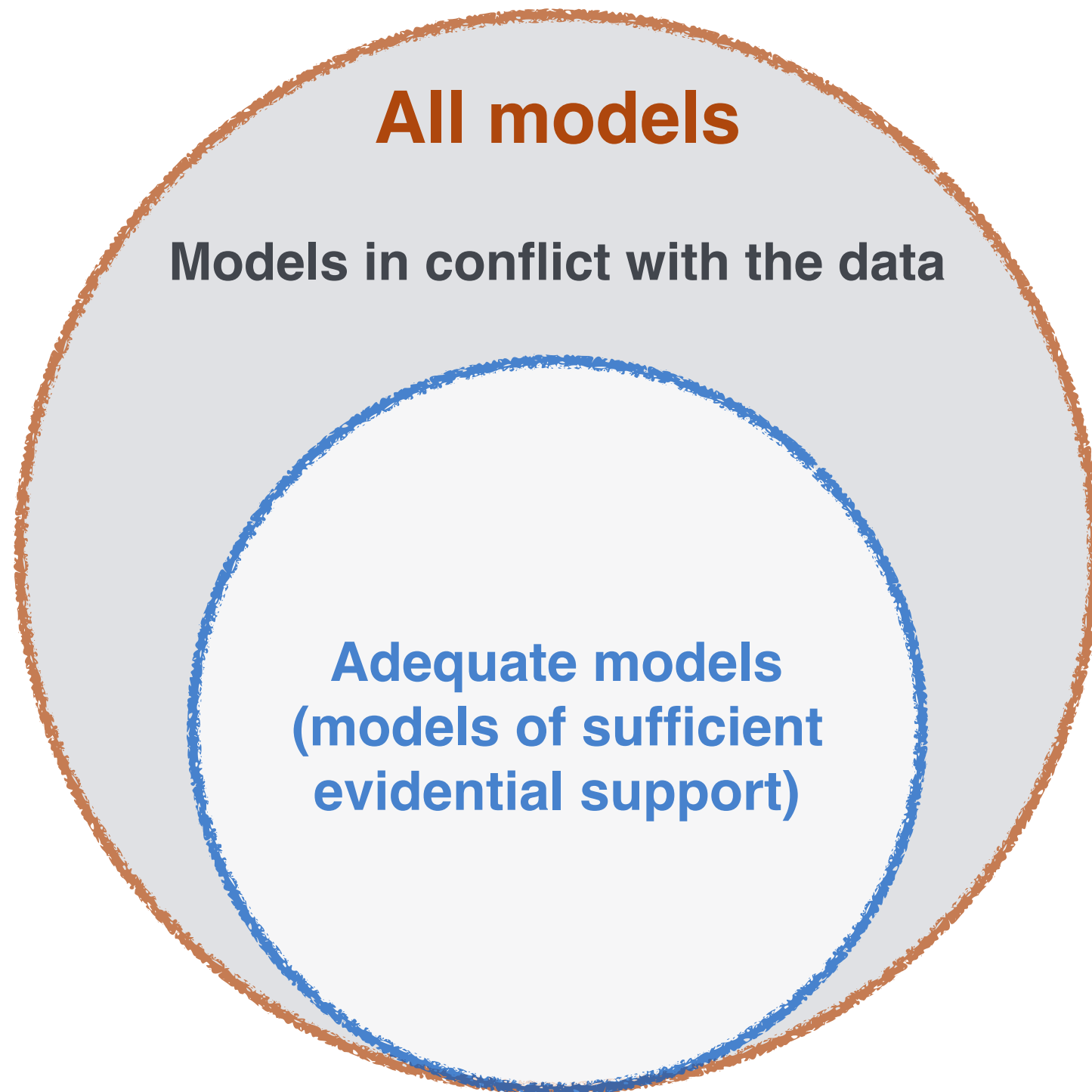
Adequate models have
sufficient evidence

$$B(M; x) = \frac{f_M(x)}{f_{\neg M}(x)} = \frac{\int f_M(x|\theta) \pi_M(\theta) d\theta}{\int f_M(x|\theta) \pi_{\neg M}(\theta) d\theta}$$

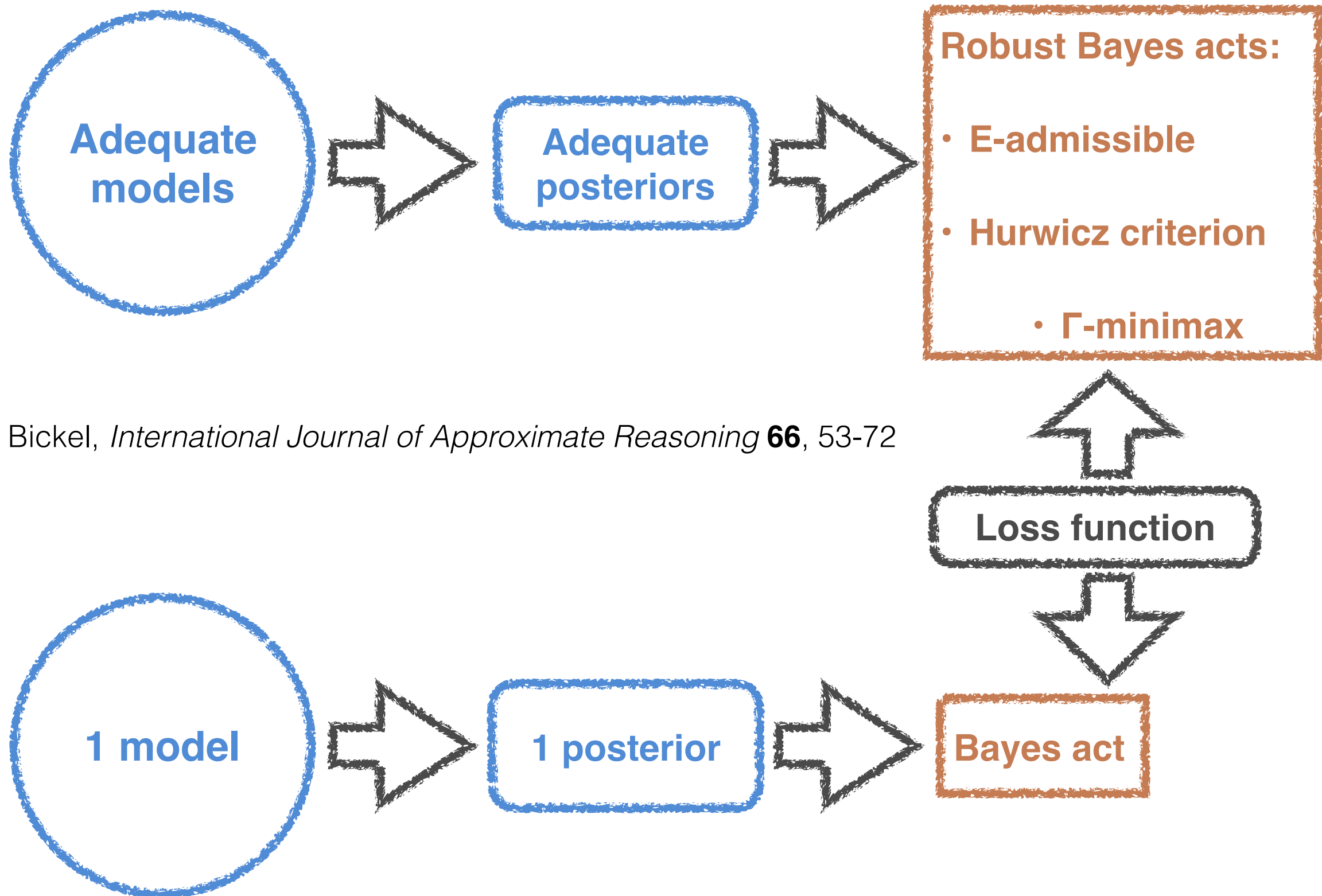
$$w(M) = \log B(M; x)$$

$$\mathcal{M}(a) = \{M \in \mathcal{M} : w(M) > a\}$$

Assessing multiple models



Decisions under robust Bayes rules applied to models of sufficient evidence



Bayesian model assessment

Does the prior or parametric family conflict with the data?

- Yes, if it has low evidential support compared to alternatives
- **Measures of support** satisfying the criteria of Bickel, *International Statistical Review* **81**, 188-206:
 - Likelihood ratio (limited to comparing two simple models)
 - Bayes factor
 - Penalized likelihood ratio

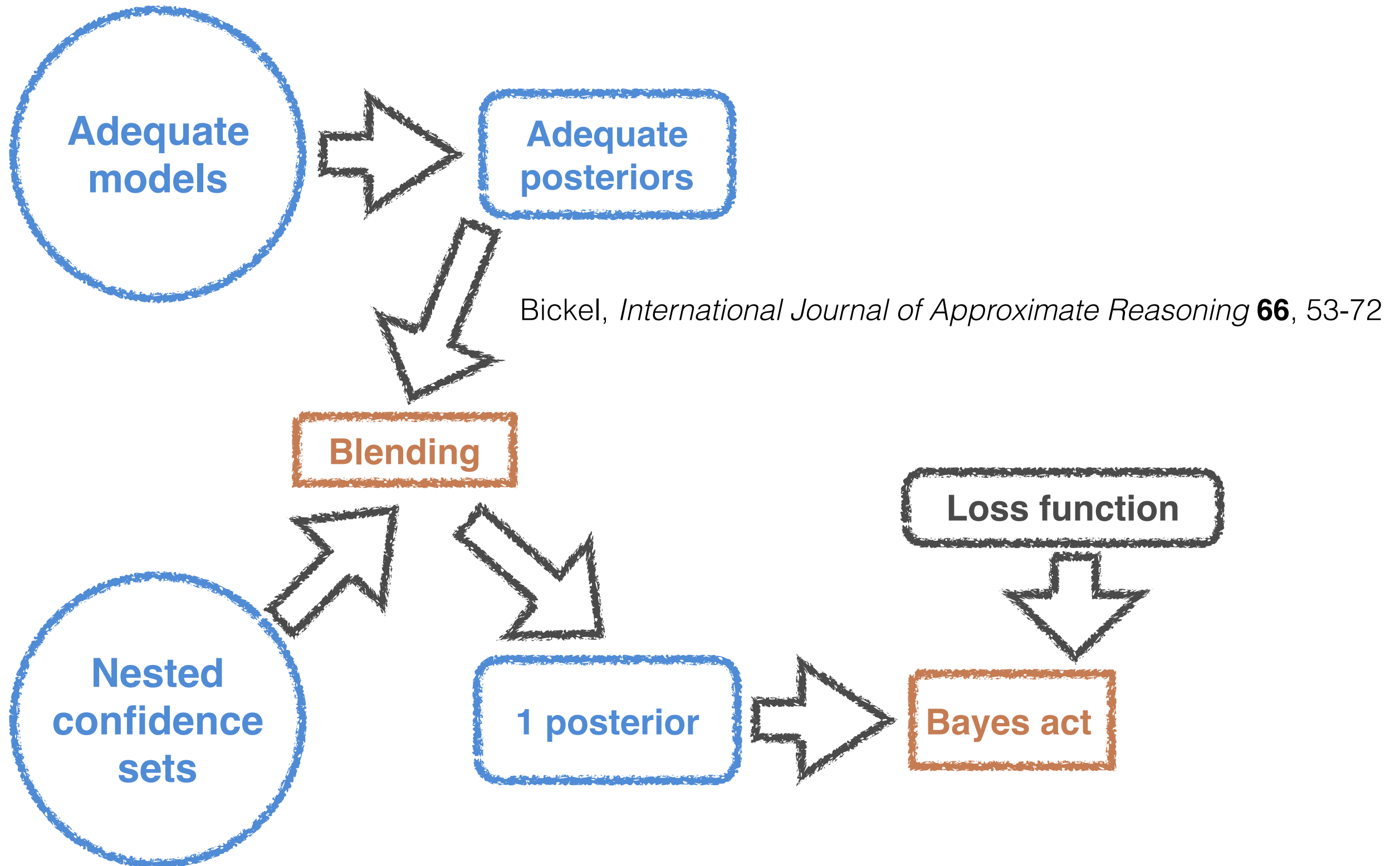
No

Yes

Use the model to take Bayes act
(minimize posterior expected loss)
even if other models are adequate

Change the model
*to an adequate model, a
model of sufficient support*

Decisions under the entropy-maximizing Bayesian model of sufficient evidence



Posterior distributions of the parameter

- Set $\dot{\mathcal{P}}$ of adequate posterior distributions on (Θ, \mathcal{A})
 - Subjective interpretation: interval levels of belief
 - Objective interpretation: physical constraints
- Set $\ddot{\mathcal{P}}$ of confidence distributions
 - Confidence distribution \ddot{P} on (Θ, \mathcal{A})

Information theory

- Information divergence of P with respect to Q on (Θ, \mathcal{A}) :

$$P \ll Q \implies I(P||Q) = \int dP \log \left(\frac{dP}{dQ} \right)$$

$$P \not\ll Q \implies I(P||Q) = \infty$$

- Inferential gain of Q relative to \ddot{P} given \dot{P} :

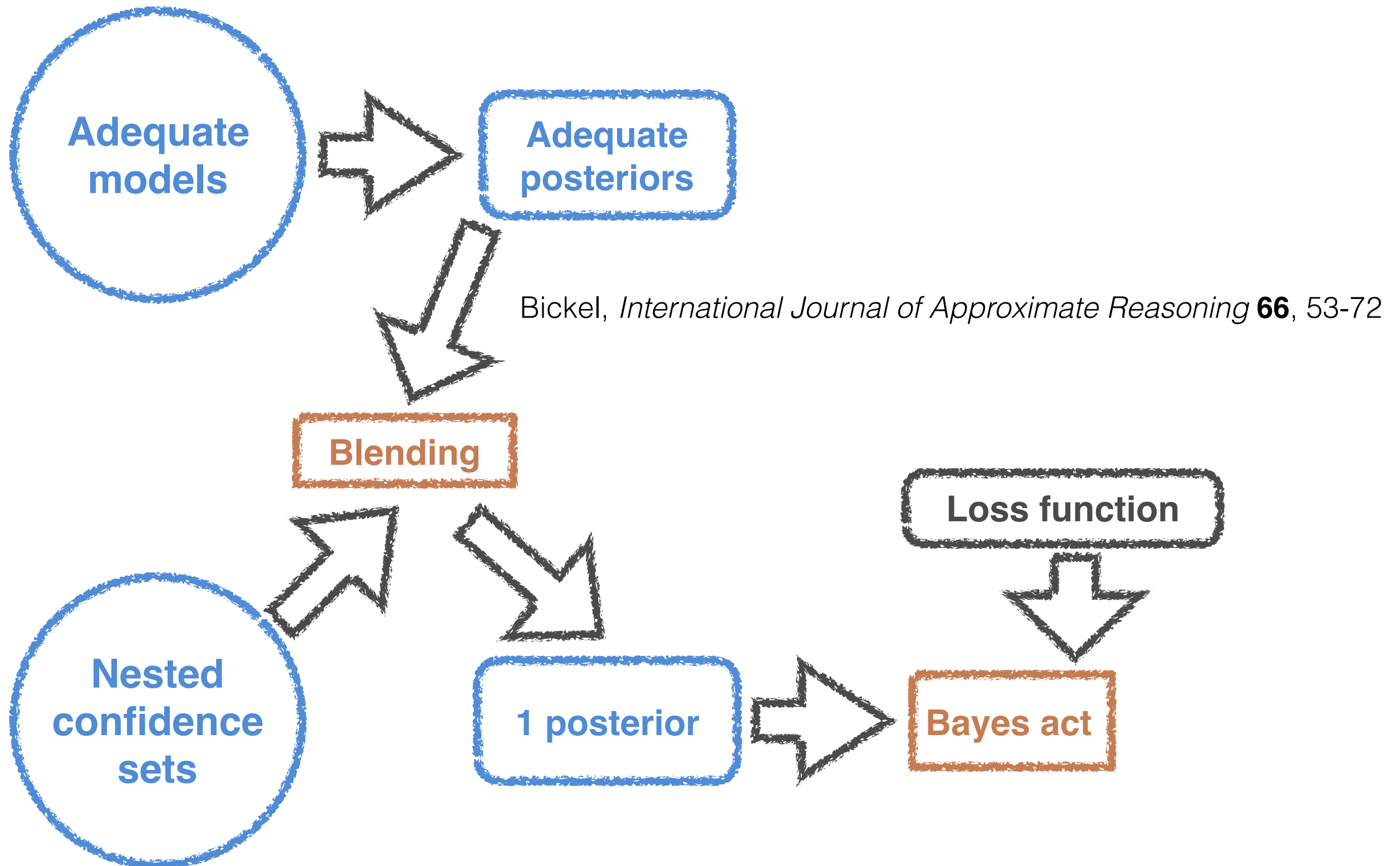
$$I(\dot{P}||\ddot{P} \rightsquigarrow Q) = I(\dot{P}||\ddot{P}) - I(\dot{P}||Q)$$

- Subset $\dot{\mathcal{P}}(\ddot{P}) = \left\{ \dot{P} \in \dot{\mathcal{P}} : I(\dot{P}||\ddot{P}) < \infty \right\}$ of Bayes posteriors

Game theory

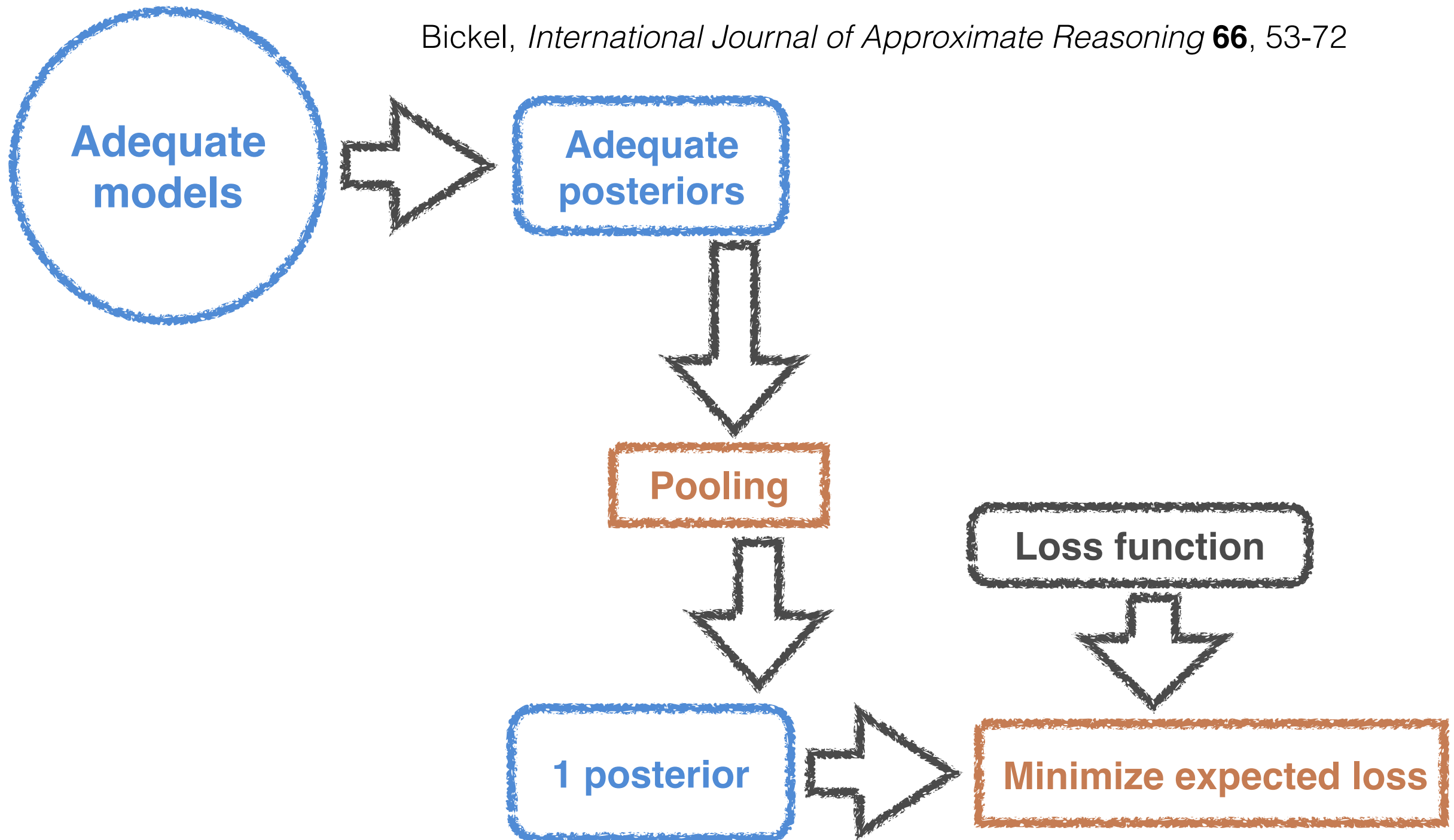
- The inferential gain of Q is $I \left(\dot{P} || \ddot{P} \rightsquigarrow Q \right) = I \left(\dot{P} || \ddot{P} \right) - I \left(\dot{P} || Q \right)$.
- The blended posterior distribution \hat{P} is defined by this game:
$$\inf_{\dot{P} \in \dot{\mathcal{P}}(\ddot{P})} I \left(\dot{P} || \ddot{P} \rightsquigarrow \hat{P} \right) = \sup_{Q \in \mathcal{P}} \inf_{\dot{P} \in \dot{\mathcal{P}}(\ddot{P})} I \left(\dot{P} || \ddot{P} \rightsquigarrow Q \right).$$
- If $I \left(\dot{P} || \ddot{P} \right) < \infty$ for some $\dot{P} \in \dot{\mathcal{P}}$ and if $\dot{\mathcal{P}} \left(\ddot{P} \right)$ is convex, then $I \left(\hat{P} || \ddot{P} \right) = \inf_{\dot{P} \in \dot{\mathcal{P}}(\ddot{P})} I \left(\dot{P} || \ddot{P} \right)$.
 - F. Topsøe, *Kybernetika* 15 (1979), 8-27; P. Harremoës and F. Topsøe, *Entropy* 3 (2001), 191-226; F. Topsøe, in *Entropy, Search, Complexity* (Springer, 2007), 179-207

Decisions under the entropy-maximizing Bayesian model of sufficient evidence



Decisions under entropy-based pooling of Bayesian model of sufficient evidence

Bickel, *International Journal of Approximate Reasoning* **66**, 53-72



The three players

- **Combiner** is you, the statistician who combines the candidate distributions
 - Goal 1: minimize error; Goal 2: beat Chooser
- **Chooser** is the imaginary statistician who instead chooses a single candidate distribution
 - Goal 1: minimize error; Goal 2: beat Combiner
- **Nature** picks the true distribution among those that are plausible in order to help Chooser



Information game

- Information divergence and inferential gain:

$$D(P||Q) = \int dP(\xi) \log \frac{dP(\xi)}{dQ(\xi)}$$

$$D(P' || P'' \rightarrow Q) = D(P' || P'') - D(P' || Q)$$

- Utility paid to Statistician 2 (Combiner or Chooser):

$$U(\dot{P}, P_1, P_2) = \left(-D(\dot{P} || P_2), D(\dot{P} || P_1 \rightarrow P_2) \right)$$

- Reduction to game of Combiner v. Nature-Chooser Coalition:

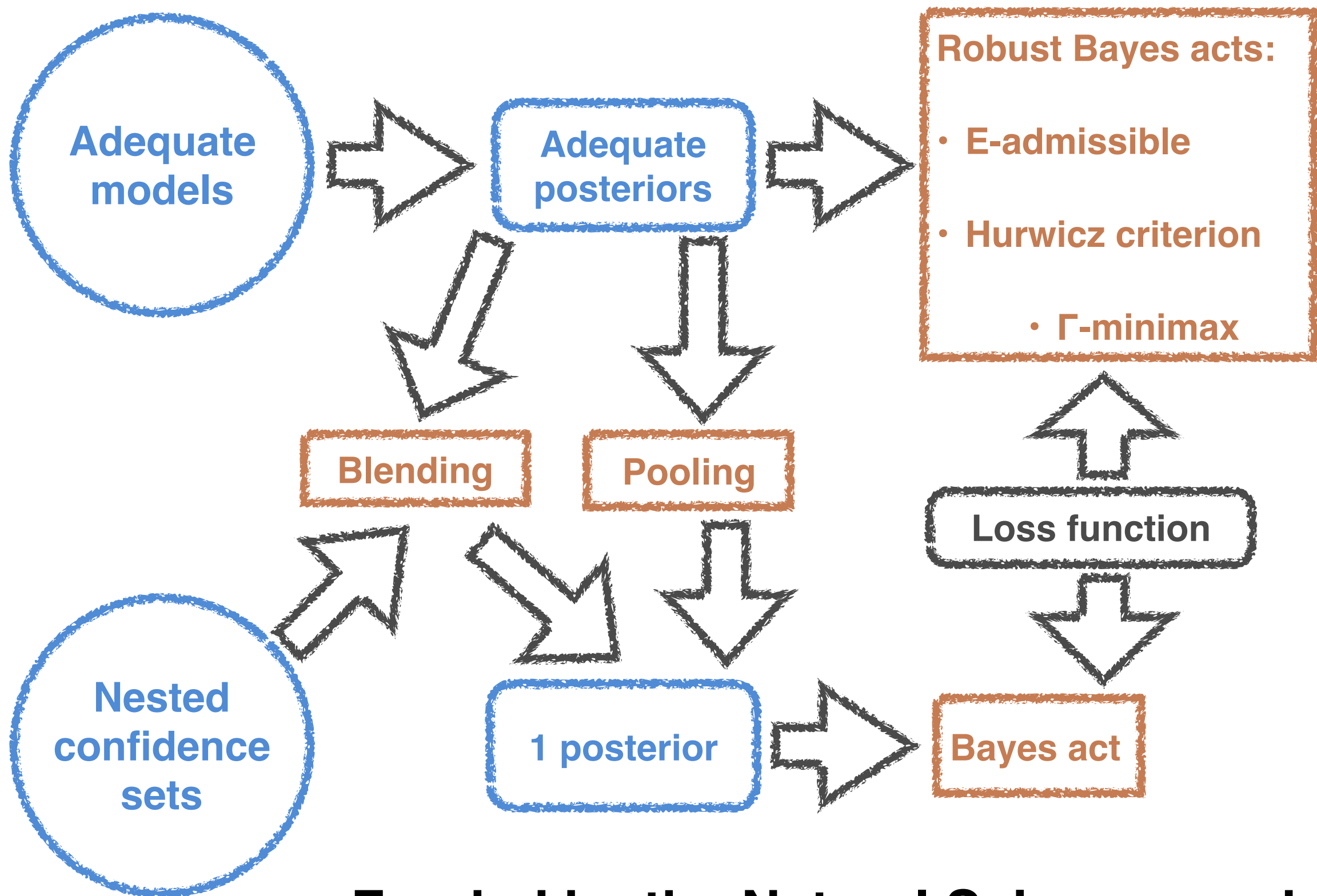
$$P^+ = \arg \inf_Q \sup_{P'} D(P' || Q)$$



Q



P'



**Funded by the Natural Sciences and
Engineering Research Council of Canada**