Agreeing to Disagree

Wessel Bruinsma

1 November 2018

Image from `relativelyinteresting.com/win-argument-according-science/`.

Alice

Image from relativelyinteresting.com/win-argument-according-science/.

Alice disagrees

Image from relativelyinteresting.com/win-argument-according-science/.

I. A Model of Knowledge

I. A Model of Knowledge

II. The Exciting Bit

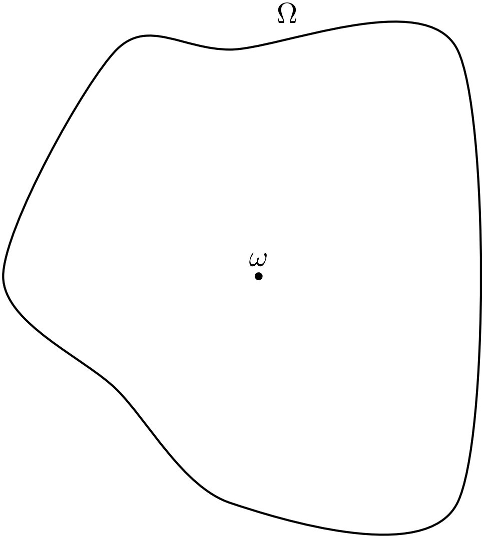
III. Questioning our Assumptions

A Model of Knowledge

Alice's World

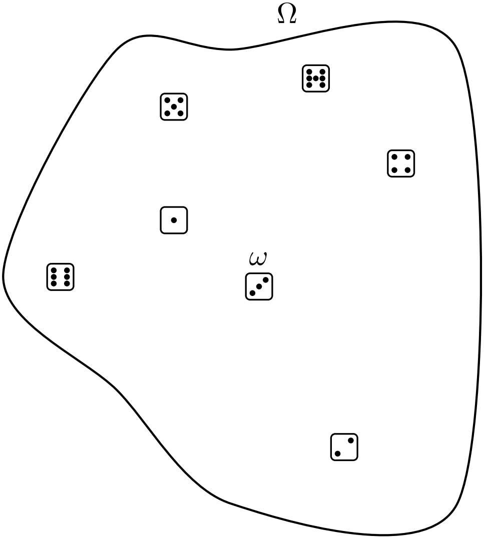
Ω:

states of Alice's world.



Ω:

states of Alice's world.

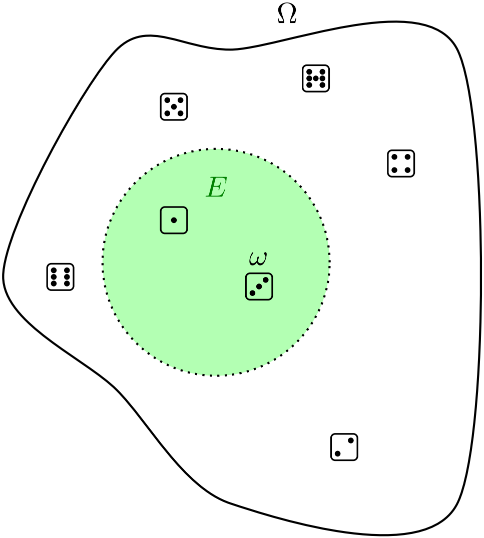


Ω:

states of Alice's world.

$E \subseteq \Omega$:

event.

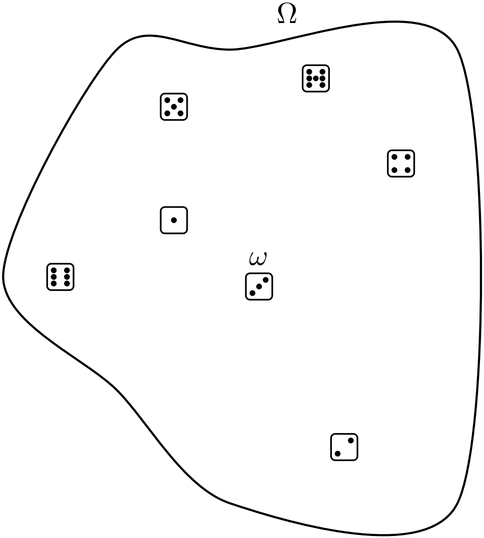


Ω:

states of Alice's world.

$E \subseteq \Omega$:

event.



Ω:

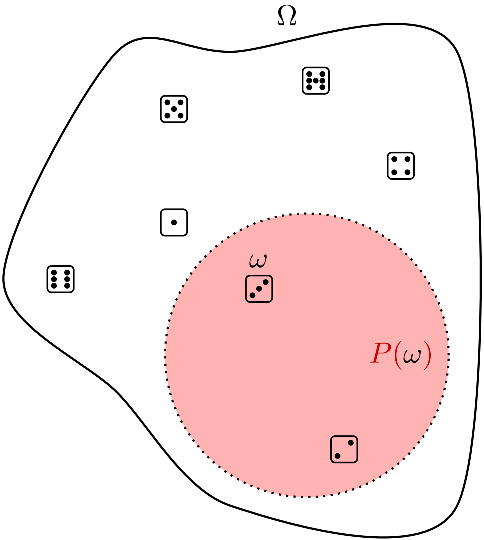
states of Alice's world.

$E \subseteq \Omega$:

event.

$P(\omega) \subseteq \Omega$:

Alice's knowledge.



Alice's World

Ω:

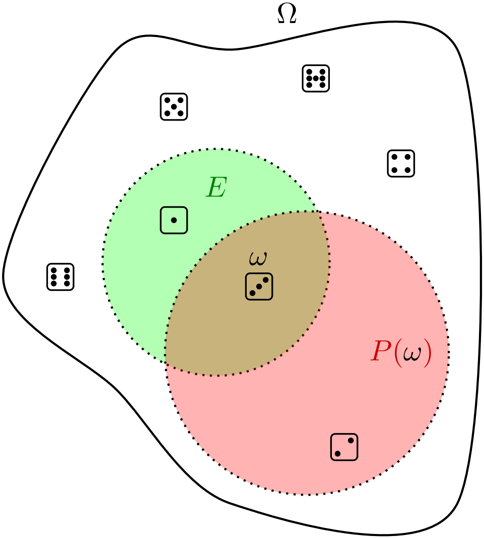
states of Alice's world.

$E \subseteq \Omega$:

event.

$P(\omega) \subseteq \Omega$:

Alice's knowledge.



Ω:

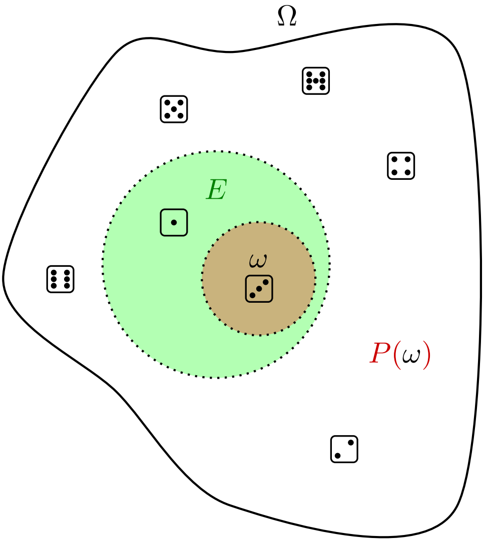
states of Alice's world.

$E \subseteq \Omega$:

event.

$P(\omega) \subseteq \Omega$:

Alice's knowledge.



Alice's World (2)

$P(\omega) \subseteq E$:

at ω, Alice *knows* E.

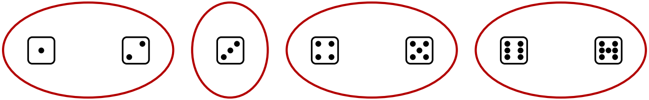
$P(\omega) \subseteq E$:

at ω, Alice *knows* E.

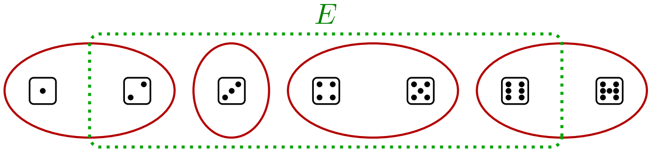
Alice's knowledge function:

$K(E) = \{\omega : \text{Alice knows } E\}$.

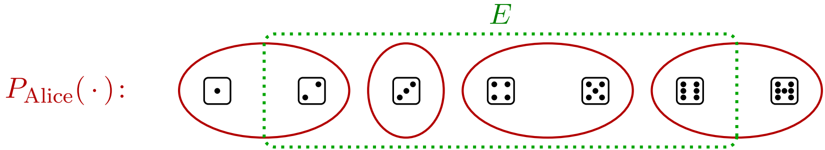
$P_{\text{Alice}}(\cdot):$



$P_{\text{Alice}}(\,\cdot\,):$

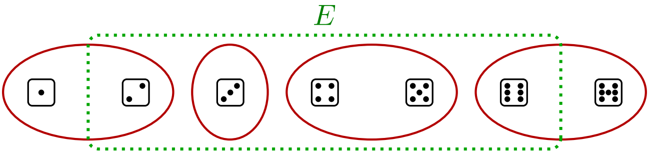


The Rare Die



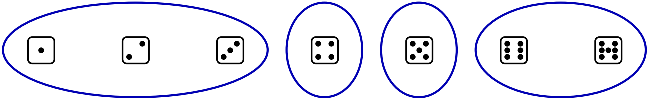
$$K_{\text{Alice}}(E) = \{⚂, ⚃, ⚄\}.$$

$P_{\mathrm{Alice}}(\cdot):$

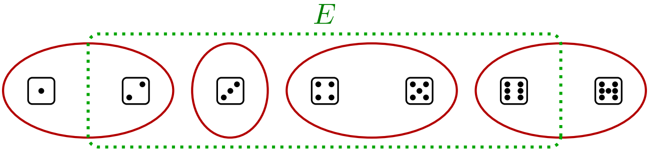


$K_{\mathrm{Alice}}(E) = \{$⚂, ⚃, ⚄$\}.$

$P_{\mathrm{Bob}}(\cdot):$

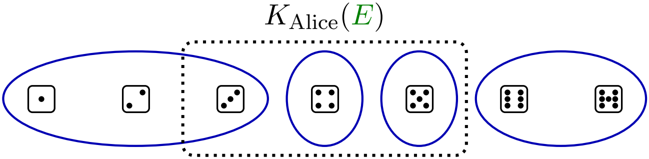


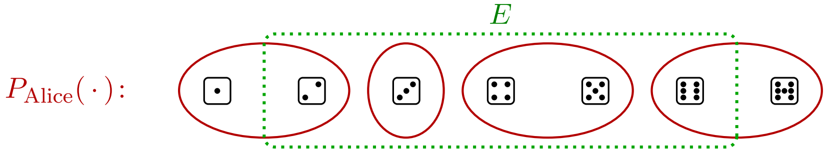
$P_{\text{Alice}}(\,\cdot\,)$:



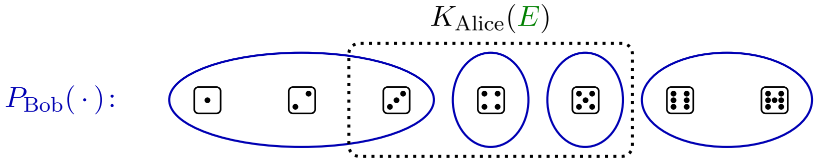
$$K_{\text{Alice}}(E) = \{⚂, ⚃, ⚄\}.$$

$P_{\text{Bob}}(\,\cdot\,)$:





$$K_{\text{Alice}}(E) = \{⚂, ⚃, ⚄\}.$$



$$K_{\text{Bob}}(K_{\text{Alice}}(E)) = \{⚃, ⚄\}.$$

$\omega \in K_{\text{Alice}}(E)$:

Alice knows E.

$\omega \in K_{\text{Alice}}(E)$:

Alice knows E.

$\omega \in K_{\text{Bob}}(K_{\text{Alice}}(E))$:

Bob knows that Alice knows E.

$\omega \in K_{\text{Alice}}(E)$:

Alice knows E.

$\omega \in K_{\text{Bob}}(K_{\text{Alice}}(E))$:

Bob knows that Alice knows E.

$\omega \in K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E)))$:

Alice knows that Bob knows that Alice knows E.

$\omega \in K_{\text{Alice}}(E)$:

Alice knows E.

$\omega \in K_{\text{Bob}}(K_{\text{Alice}}(E))$:

Bob knows that Alice knows E.

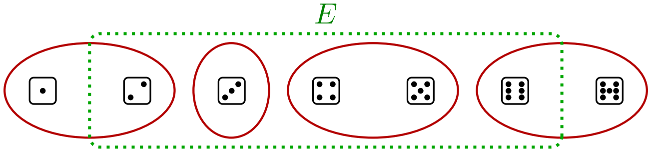
$\omega \in K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E)))$:

Alice knows that Bob knows that Alice knows E.

$$\vdots$$

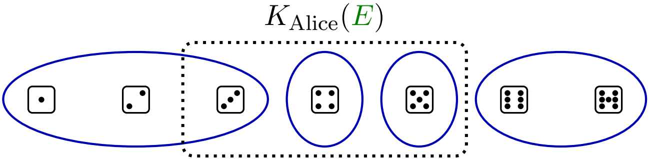
At ω, E is **common knowledge** between Alice and Bob.

$P_{\text{Alice}}(\cdot)$:



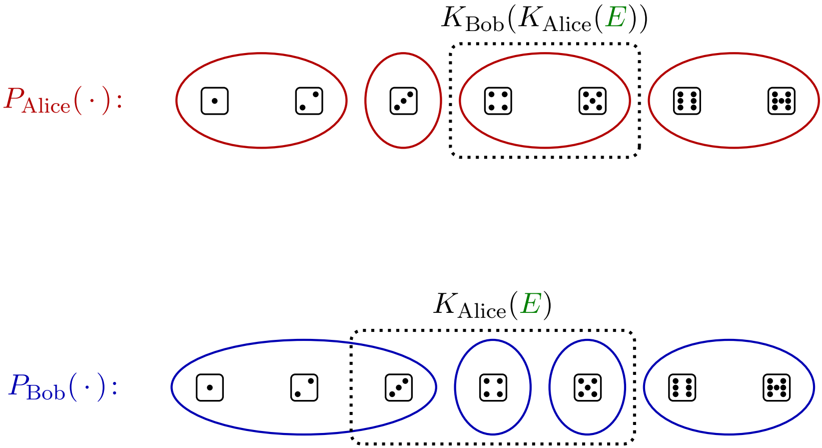
$$K_{\text{Alice}}(E) = \{⚂, ⚃, ⚄\}.$$

$P_{\text{Bob}}(\cdot)$:



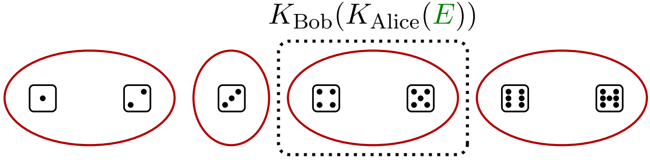
$$K_{\text{Bob}}(K_{\text{Alice}}(E)) = \{⚃, ⚄\}.$$

The Rare Die (2)



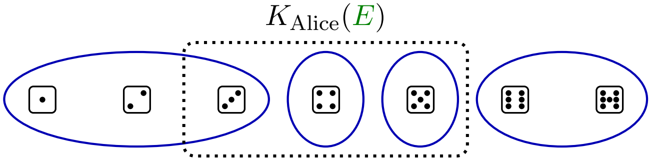
$K_{\mathrm{Bob}}(K_{\mathrm{Alice}}(E)) = \{$⚃, ⚄$\}$.

$P_{\text{Alice}}(\,\cdot\,)$:



$$K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E))) = \{⚃, ⚄\}.$$

$P_{\text{Bob}}(\,\cdot\,)$:

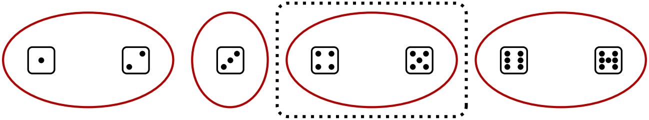


$$K_{\text{Bob}}(K_{\text{Alice}}(E)) = \{⚃, ⚄\}.$$

The Rare Die (2)

$K_{\text{Bob}}(K_{\text{Alice}}(E))$

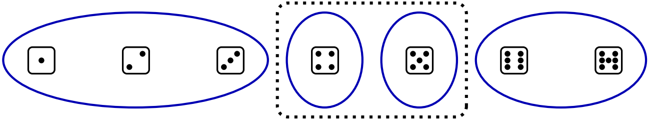
$P_{\text{Alice}}(\cdot)$:



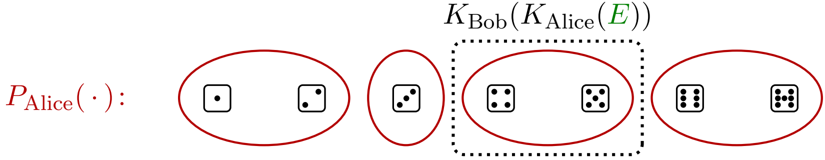
$$K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E))) = \{⚃, ⚄\}.$$

$K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E)))$

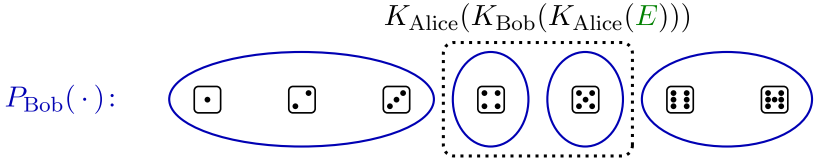
$P_{\text{Bob}}(\cdot)$:



The Rare Die (2)



$$K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E))) = \{⚃, ⚄\}.$$



$$K_{\text{Bob}}(K_{\text{Alice}}(K_{\text{Bob}}(K_{\text{Alice}}(E)))) = \{⚃, ⚄\}.$$

The Exciting Bit

- Alice and Bob have a common prior μ.

- Alice and Bob have a common prior μ.
- Alice's belief about some event E: $\mu(E \mid P_{\text{Alice}}(\omega))$.

Aumann's Agreement Theorem

- Alice and Bob have a common prior μ.
- Alice's belief about some event E: $\mu(E \mid P_{\text{Alice}}(\omega))$.

Theorem (Aumann [Aum76])

If $\mu(E \mid P_{\text{Alice}}(\omega))$ and $\mu(E \mid P_{\text{Bob}}(\omega))$ are common knowledge between Alice and Bob, then these beliefs must be equal.

Aumann's Agreement Theorem

- Alice and Bob have a common prior μ.
- Alice's belief about some event E: $\mu(E \mid P_{\text{Alice}}(\omega))$.

Theorem (Aumann [Aum76])

If $\mu(E \mid P_{\text{Alice}}(\omega))$ and $\mu(E \mid P_{\text{Bob}}(\omega))$ are common knowledge between Alice and Bob, then these beliefs must be equal.

- Alice and Bob cannot agree to disagree.

Sketch of Proof

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Proposition

At ω, E is common knowledge between Alice and Bob iff there exists an event $\omega \in F \subseteq E$ that is self evident for both Alice and Bob.

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Proposition

At ω, E is common knowledge between Alice and Bob iff there exists an event $\omega \in F \subseteq E$ that is self evident for both Alice and Bob.

- $E =$ Alice and Bob have particular beliefs: q_{Alice} and q_{Bob}.

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Proposition

At ω, E is common knowledge between Alice and Bob iff there exists an event $\omega \in F \subseteq E$ that is self evident for both Alice and Bob.

- $F \subseteq E$ = Alice and Bob have particular beliefs: q_{Alice} and q_{Bob}.

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Proposition

At ω, E is common knowledge between Alice and Bob iff there exists an event $\omega \in F \subseteq E$ that is self evident for both Alice and Bob.

- $F \subseteq E =$ Alice and Bob have particular beliefs: q_{Alice} and q_{Bob}.
- $\mu(E \mid F) = q_{\text{Alice}}$.

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Proposition

At ω, E is common knowledge between Alice and Bob iff there exists an event $\omega \in F \subseteq E$ that is self evident for both Alice and Bob.

- $F \subseteq E$ = Alice and Bob have particular beliefs: q_{Alice} and q_{Bob}.
- $\mu(E \mid F) = q_{\text{Alice}}$.
- $\mu(E \mid F) = q_{\text{Bob}}$.

Sketch of Proof

- F is **self evident** if Alice knows it whenever it occurs.

Proposition

At ω, E is common knowledge between Alice and Bob iff there exists an event $\omega \in F \subseteq E$ that is self evident for both Alice and Bob.

- $F \subseteq E$ = Alice and Bob have particular beliefs: q_{Alice} and q_{Bob}.
- $\mu(E \mid F) = q_{\text{Alice}}$.
- $\mu(E \mid F) = q_{\text{Bob}}$.
- $q_{\text{Alice}} = q_{\text{Bob}}$. $\blacksquare$

Disagreement is Unpredictable

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

Disagreement is Unpredictable

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

- Consider future Bob: Bob$'$.

Disagreement is Unpredictable

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

- Consider future Bob: Bob$'$.
- Future Bob's estimate: $\mathbb{E}_{\text{Bob}'}(X)$.

Disagreement is Unpredictable

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

- Consider future Bob: Bob$'$.
- Future Bob's estimate: $\mathbb{E}_{\text{Bob}'}(X)$.
- Alice's estimate of future Bob's estimate: $\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X))$.

Disagreement is Unpredictable

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

- Consider future Bob: Bob′.
- Future Bob's estimate: $\mathbb{E}_{\text{Bob}'}(X)$.
- Alice's estimate of future Bob's estimate: $\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X))$.

Theorem ( )

$$\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X)) < \mathbb{E}_{\text{Alice}}(X)$$

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

- Consider future Bob: Bob$'$.
- Future Bob's estimate: $\mathbb{E}_{\text{Bob}'}(X)$.
- Alice's estimate of future Bob's estimate: $\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X))$.

Theorem (Hanson [Han02])

It cannot be that $\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X)) < \mathbb{E}_{\text{Alice}}(X)$ (or "$>$") is common knowledge between Alice and Bob.

Disagreement is Unpredictable

- Alice's estimate of some random variable X:

$$\mathbb{E}_{\text{Alice}}(X) = \mathbb{E}(X \mid P_{\text{Alice}}(\omega)).$$

- Consider future Bob: Bob′.
- Future Bob's estimate: $\mathbb{E}_{\text{Bob}'}(X)$.
- Alice's estimate of future Bob's estimate: $\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X))$.

Theorem (Hanson [Han02])

It cannot be that $\mathbb{E}_{\text{Alice}}(\mathbb{E}_{\text{Bob}'}(X)) < \mathbb{E}_{\text{Alice}}(X)$ (or "$>$") is common knowledge between Alice and Bob.

- Alice cannot anticipate the direction of Bob's disagreement.

Image from relativelyinteresting.com/win-argument-according-science/.

Questioning our Assumptions

Do we really have a common prior?

Principle of Monotonicity

$$F \subseteq E \implies K(F) \subseteq K(E).$$

Principle of Monotonicity

$$F \subseteq E \implies K(F) \subseteq K(E).$$

$$K(\text{know axioms}) \subseteq K(\text{know theorems}).$$

Principle of Substitution

$$F = E \implies K(F) = K(E).$$

$$F = E \implies K(F) = K(E).$$

$$K(\text{triangle is equilateral}) = K(\text{triangle is equiangular}).$$

$$\boxed{F = E} \implies K(F) = K(E).$$

$$K(\text{triangle is equilateral}) = K(\text{triangle is equiangular}).$$

$$F = E?$$

Extension Versus Intension

$$F = E?$$

- Extension: what an expression *designates*.
- Intension: the *idea* or *notion* conveyed.

$$F = E?$$

- Extension: what an expression *designates*.
- Intension: the *idea* or *notion* conveyed.

The state-space model of knowledge respects extensional equality, but disregards the intentional dimension.

"We publish this observation with some diffidence, since once one has the appropriate framework, it is mathematically trivial. Intuitively, though, it is not quite obvious..."

—Aumann, in his original paper [Aum76]

Alice disagrees

Image from relativelyinteresting.com/win-argument-according-science/.

- Common prior

Alice disagrees

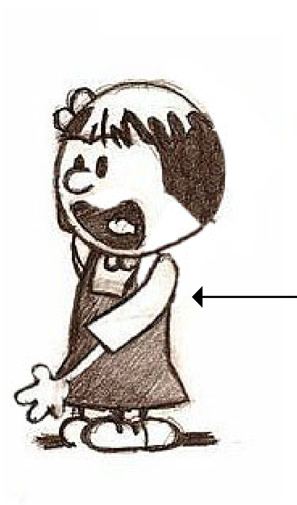
Image from relativelyinteresting.com/win-argument-according-science/.

- Common prior
- Accept model of knowledge

Alice disagrees

Image from relativelyinteresting.com/win-argument-according-science/.

- Common prior
- Accept model of knowledge



Alice disagrees

Image from relativelyinteresting.com/win-argument-according-science/.

Appendix

References

[Aar04] S. Aaronson. "The Complexity of Agreement". In: **arXiv preprint arXiv:cs/0406061** (June 2004). eprint: https://arxiv.org/abs/cs/0406061

[Aum76] Robert J. Aumann. "Agreeing to Disagree". In: **Ann. Statist.** 4.6 (Nov. 1976), pp. 1236–1239

[CH02] Tyler Cowen and Robin Hanson. "Are Disagreements Honest". In: **Journal of Economic Methodology** (2002)

[Han02] Robin Hanson. "Disagreement Is Unpredictable". In: **Economics Letters** 77.3 (2002), pp. 365–369

[Mor95] Stephen Morris. "The Common Prior Assumption in Economic Theory". In: **Economics and Philosophy** 11.2 (1995), pp. 227–253. DOI: 10.1017/S0266267100003382

[Mos07] Ivan Moscati. "Interactive and Common Knowledge of Information Partitions". In: (2007)