

Agreement in the Bayesian Brain

Adam Brandenburger^{1,2} Pierfrancesco La Mura³ Kai Steverson¹

¹ Interdisciplinary Institute for the Study of Decision Making (IISDM), New York University, U.S.A.

² Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, U.S.A.

³ HHL Leipzig Graduate School of Management, Leipzig, Germany

Version 12/18/17

Agreement in a Decentralized Network

The brain is a highly distributed system (Felleman and Van Essen, 1991; Scannell, Blakemore, and Young, 1995)

How does a decentralized network reach a ‘shared view’ --- that is, agreement --- about its external environment?

Possible connections to neuroscience:

Is **binocular rivalry** a situation where (different regions of) the brain cannot reach agreement?

Is **feature binding** a consequence (or cause?) of agreement across different regions?

Plan of the Talk

Extensive literature in game theory on how a Bayesian network comes to a shared view, i.e., to agreement (Aumann 1976 et seq.)

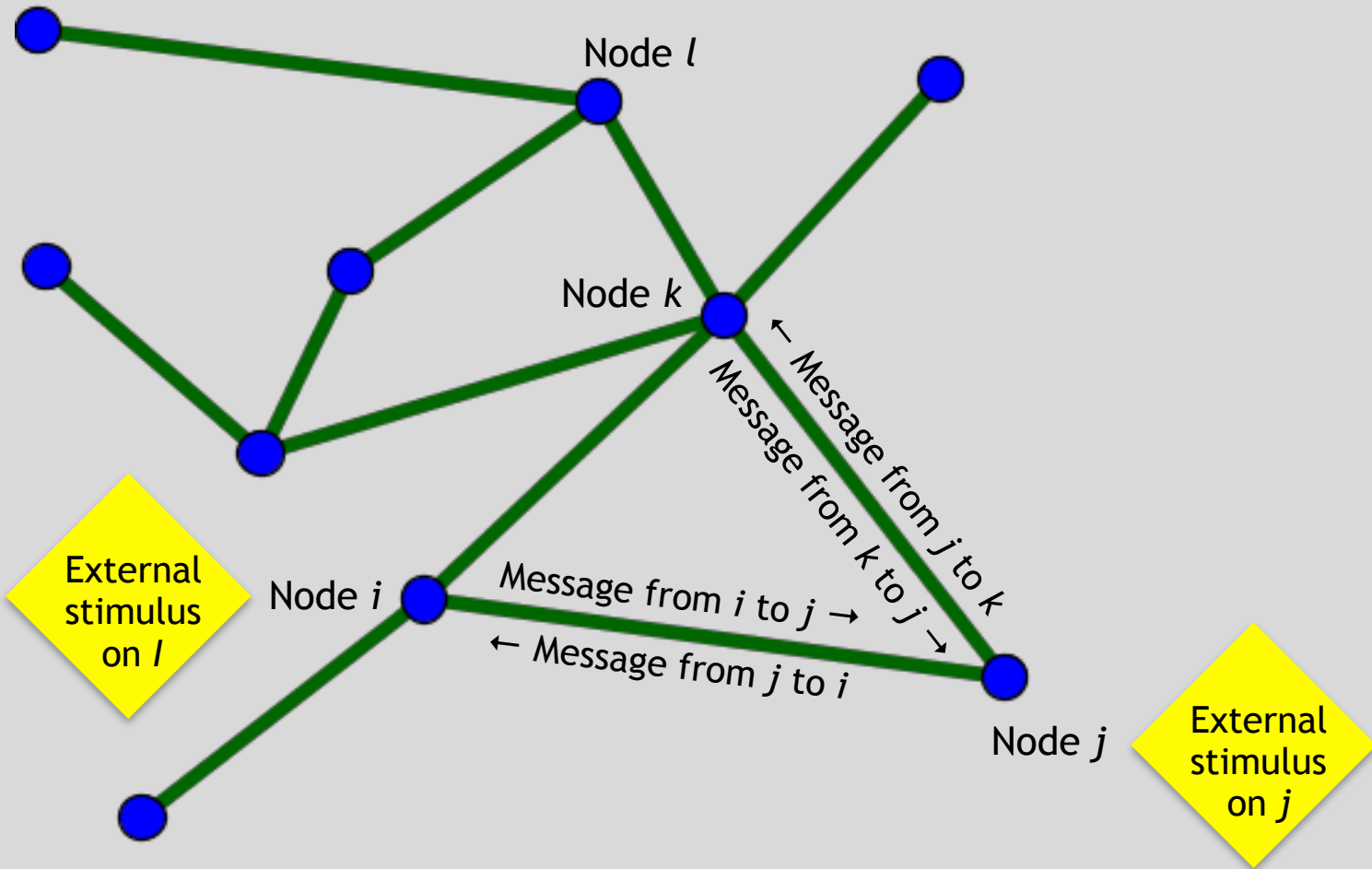
Today's talk is largely a guided tour of this literature

The literature suggests a connection between the Bayesian brain hypothesis (Pouget et al., 2013) and the problem of agreement

We will end with a discussion of possible connections between this literature and neuroscience:

- binocular rivalry
- feature binding
- global clocks (highly speculative)

The General Picture



Bayesian Networks: Information and Signals

Time goes from 0, 1, 2, ... (global clock)

There is some finite number of Bayesian nodes

The unknown true state of the world s is drawn from a set S

All nodes share a common prior on S

At time 0, each node receives a private signal about the true state

Nodes update probabilities about the true state using Bayes' rule

Message Functions

In every time period, each node sends a “message,” which might be the

- probability of a particular state or set of states (event)
- most likely state
- expectation of a random variable of interest
- other

The network structure determines which nodes send messages to which other nodes

In every time period, after receiving messages, each node updates its probabilities via Bayes’ rule

The literature has assumed all nodes employ the same message function

- using different but isomorphic message functions is equivalent
- sending messages can constitute only part of a node’s action set

Agreement

We say that two nodes “agree” if they send the same message

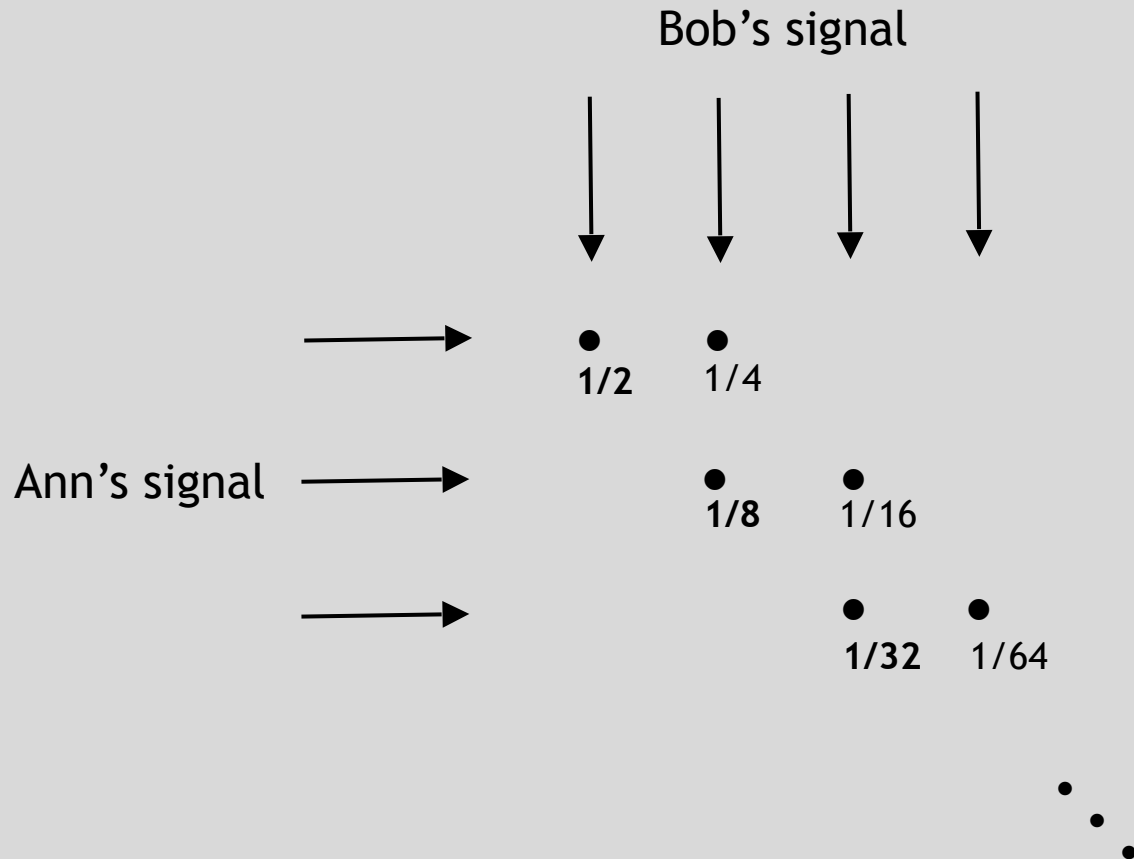
Agreement means that nodes have come to a degree of consensus about the external world

Agreement is not the same as information pooling!

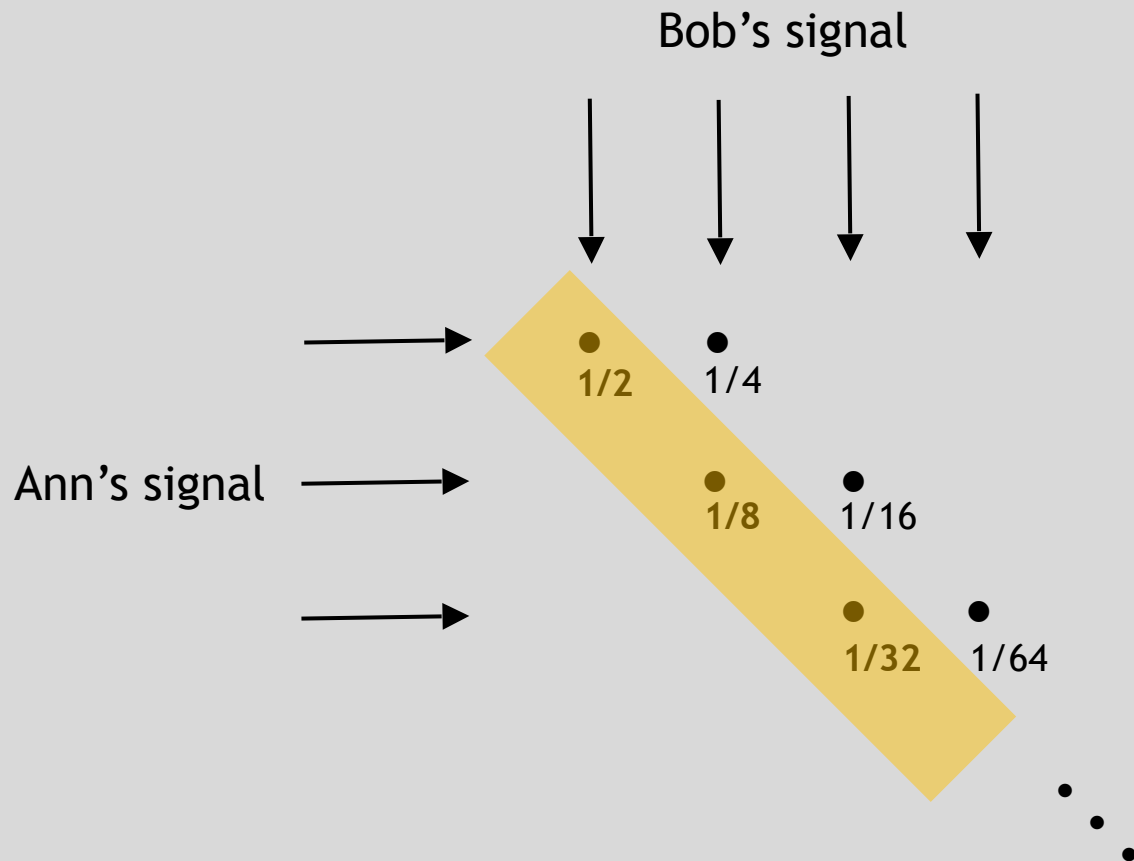
Two nodes might send the same message but have different information about the true state of the world

Information pooling is not the question we study and would be overly demanding to ask of many networks

Example of a Network of Two Nodes

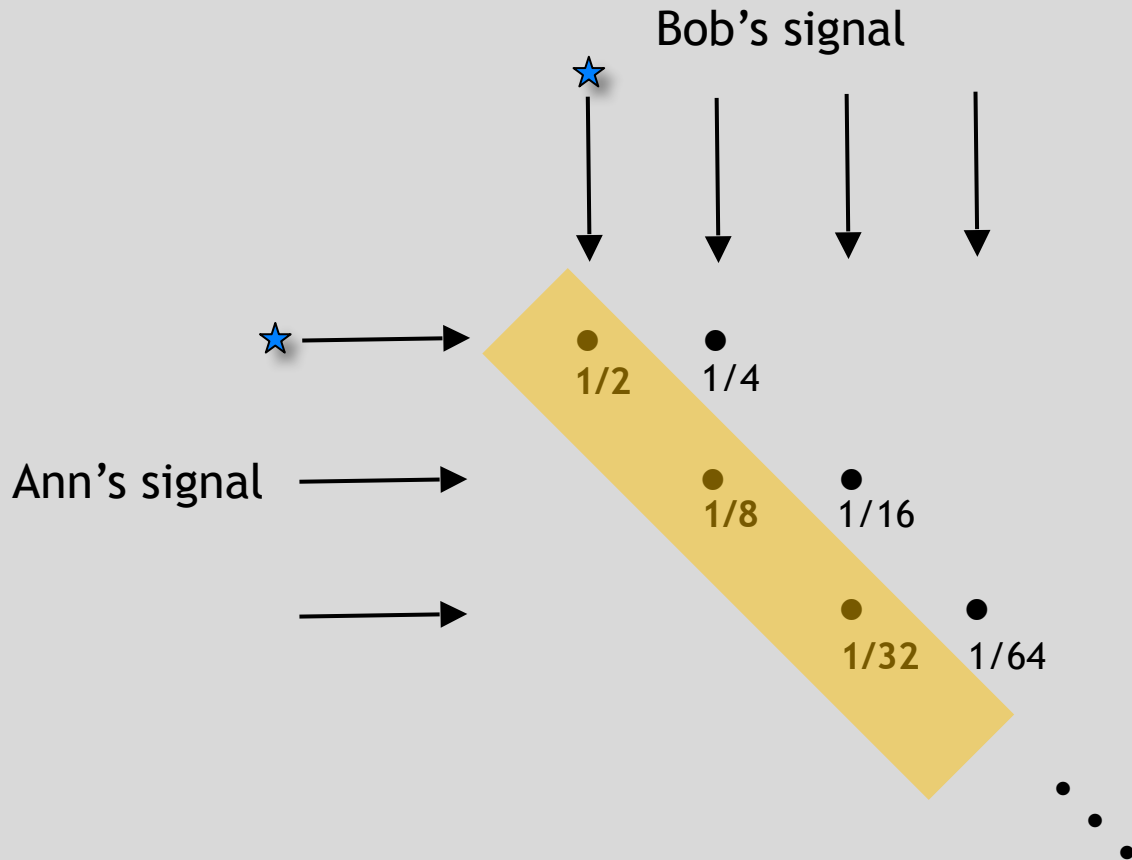


Example of a Network of Two Nodes



Each period Ann and Bob announce their probabilities of the highlighted set
(This example does exhibit information pooling but a later example will not)

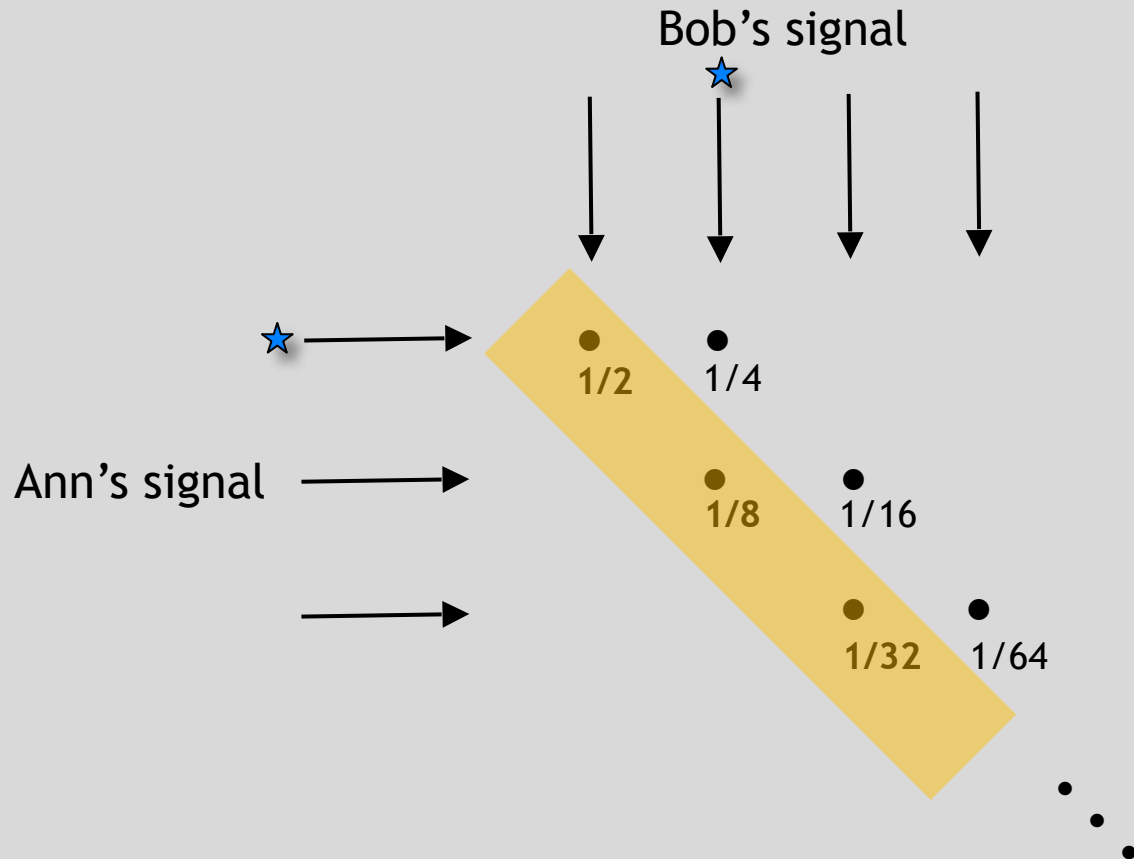
Example of a Network of Two Nodes (Scenario 1)



Period 0: Ann says probability $2/3$ and Bob says probability 1

Period 1: Ann can infer Bob's signal, and they both say probability 1 (agreement)

Example of a Network of Two Nodes (Scenario 2)

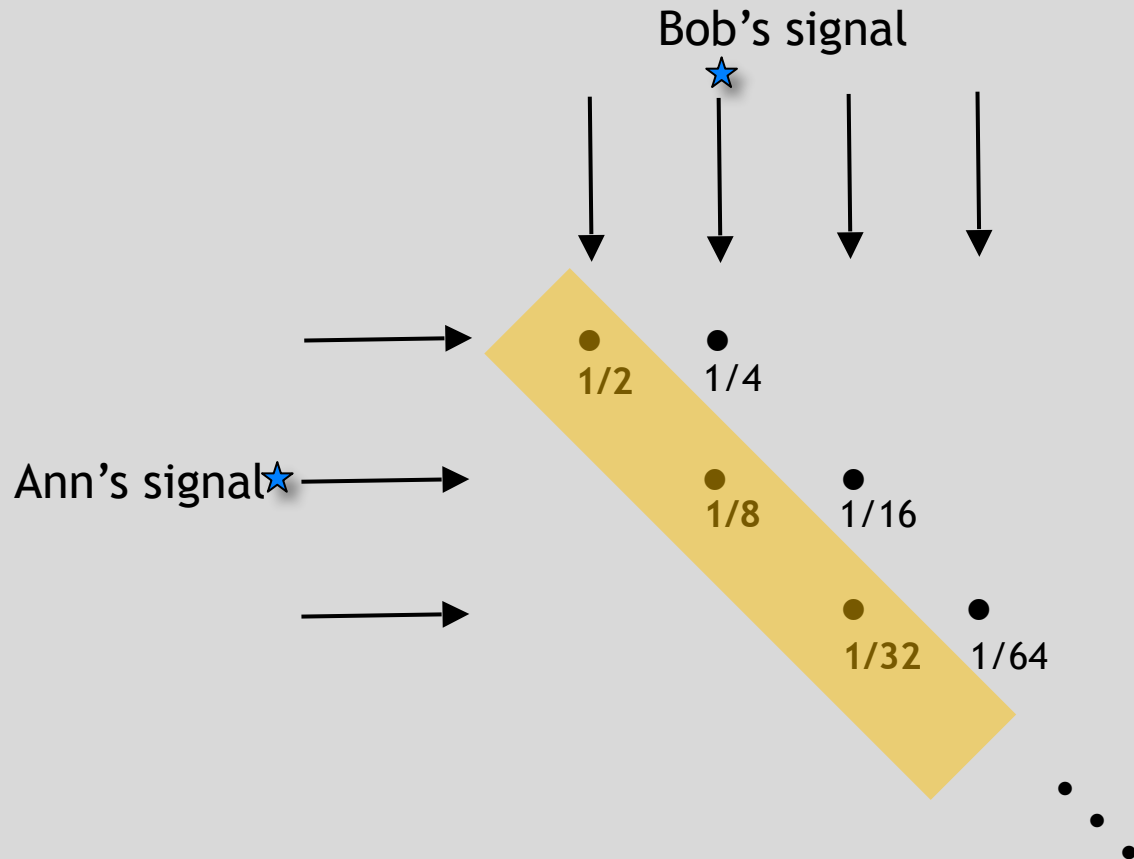


Period 0: Ann says probability $2/3$ and Bob says probability $1/3$

Period 1: Ann can rule out Scenario 1 and therefore says probability 0

Period 2: Both Ann and Bob say probability 0 (agreement)

Example of a Network of Two Nodes (Scenario 3)



Period 0: Ann says probability $2/3$ and Bob says probability $1/3$

Period 1: Ann says probability $2/3$ and Bob says probability $1/3$

Period 2: Bob can rule out Scenario 2 and therefore says probability 1

The Union Consistency Principle

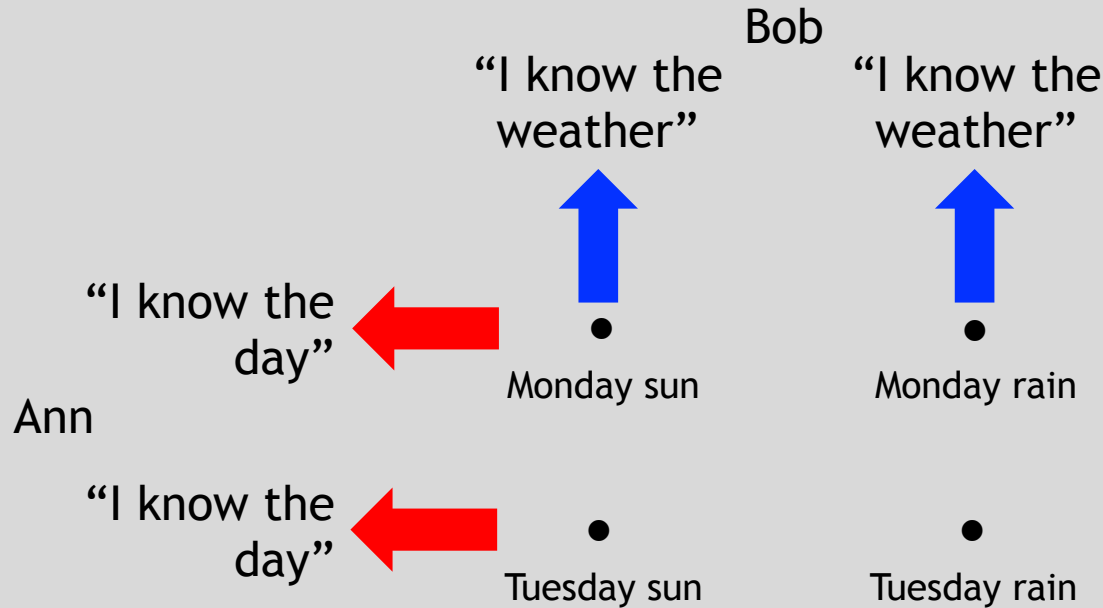
A message function satisfies the union consistency principle if
signal s induces message m and signal t induces message m
implies
knowing the signal is s OR t induces message m

The previous examples obey this principle

- probability of a particular state or set of states (event)
- most likely state
- expectation of a random variable of interest

Theorem (Cave 1983): Suppose the union consistency principle holds. Then if two nodes share messages back and forth, they will eventually agree.

Failure of Union Consistency



The two nodes will repeat their respective announcements and never agree

The union consistency principle is violated

***n*-Node Networks with a Global Message**

There are $n + 1$ nodes and a random variable of interest X

In every time period, each node sends its expected value of X to an aggregator node

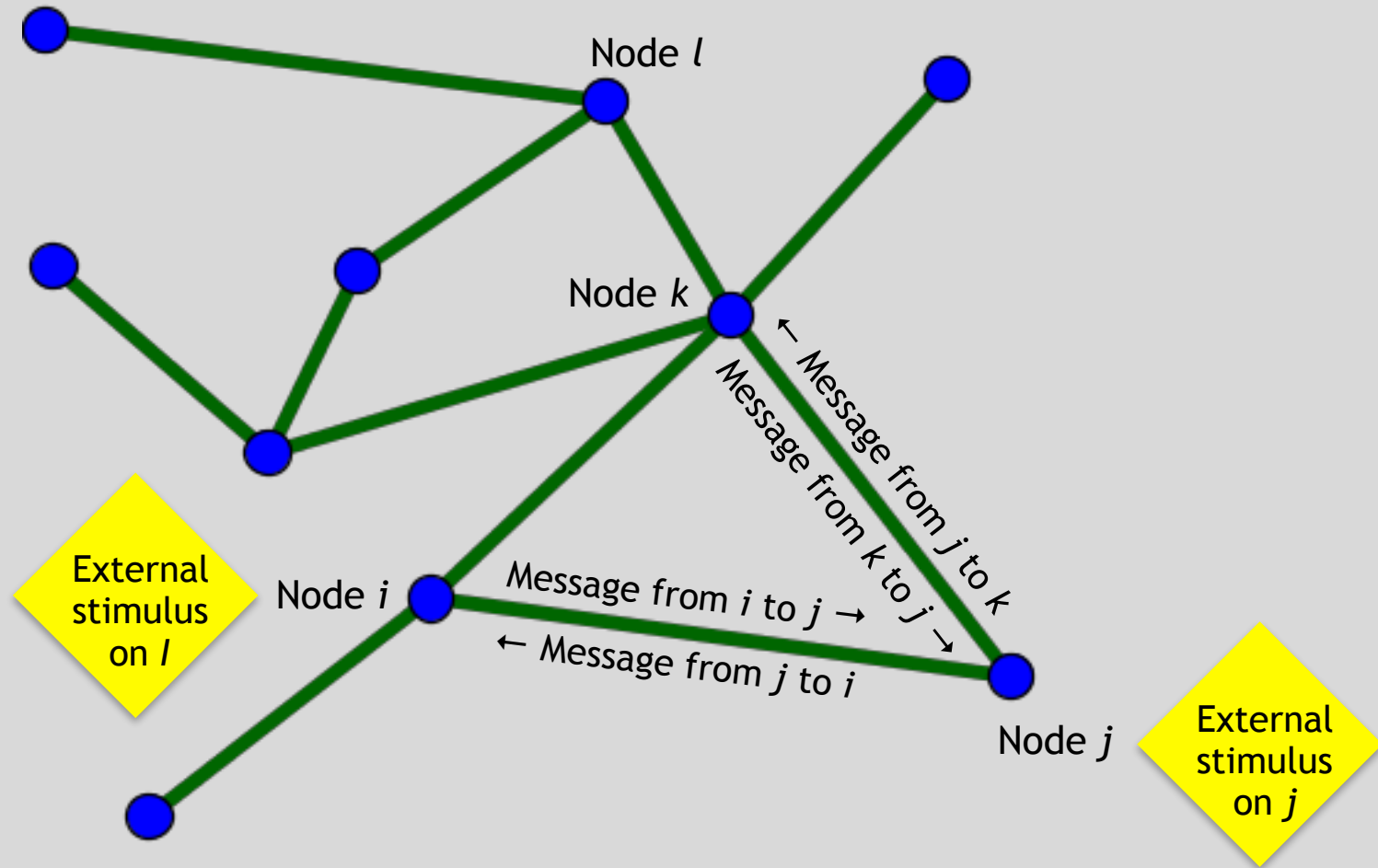
The aggregator node computes the average expectation and communicates this number in a global message to all nodes

Theorem (Nielsen et al. 1990): The n nodes will eventually agree, i.e., they will eventually compute the same expectation.

This process bears a resemblance to neural normalization with feedback (recurrence)

(The aggregator node can also compute an idiosyncratic generalized (Kolmogorov) average)

n -Node Networks with Bi-Directional Local Messages



n -Node Networks with Bi-Directional Local Messages

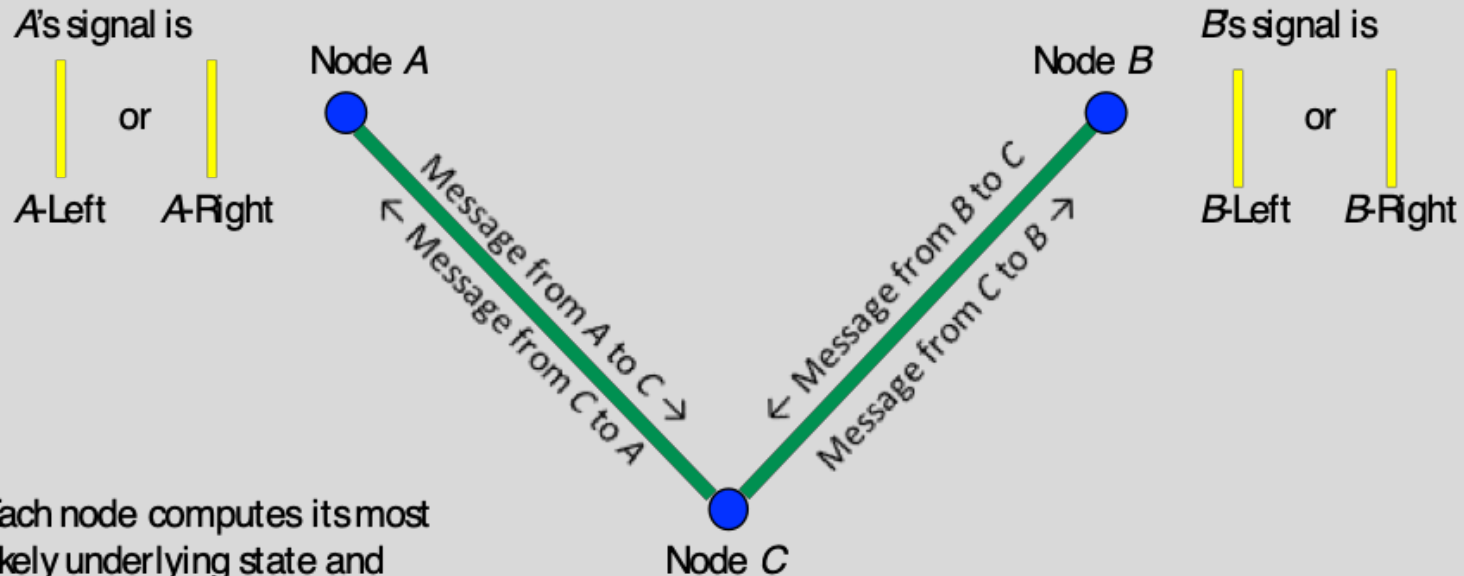
Bi-directionality means that if node i communicates with node j , then node j communicates with node i

We assume the network is connected (no isolated components)

Theorem (Krauscki 1996): Suppose the (common) message function satisfies union consistency. Then if the nodes keep sending messages, they will eventually agree.

But are brain networks necessarily bi-directional?

A Numerical Example

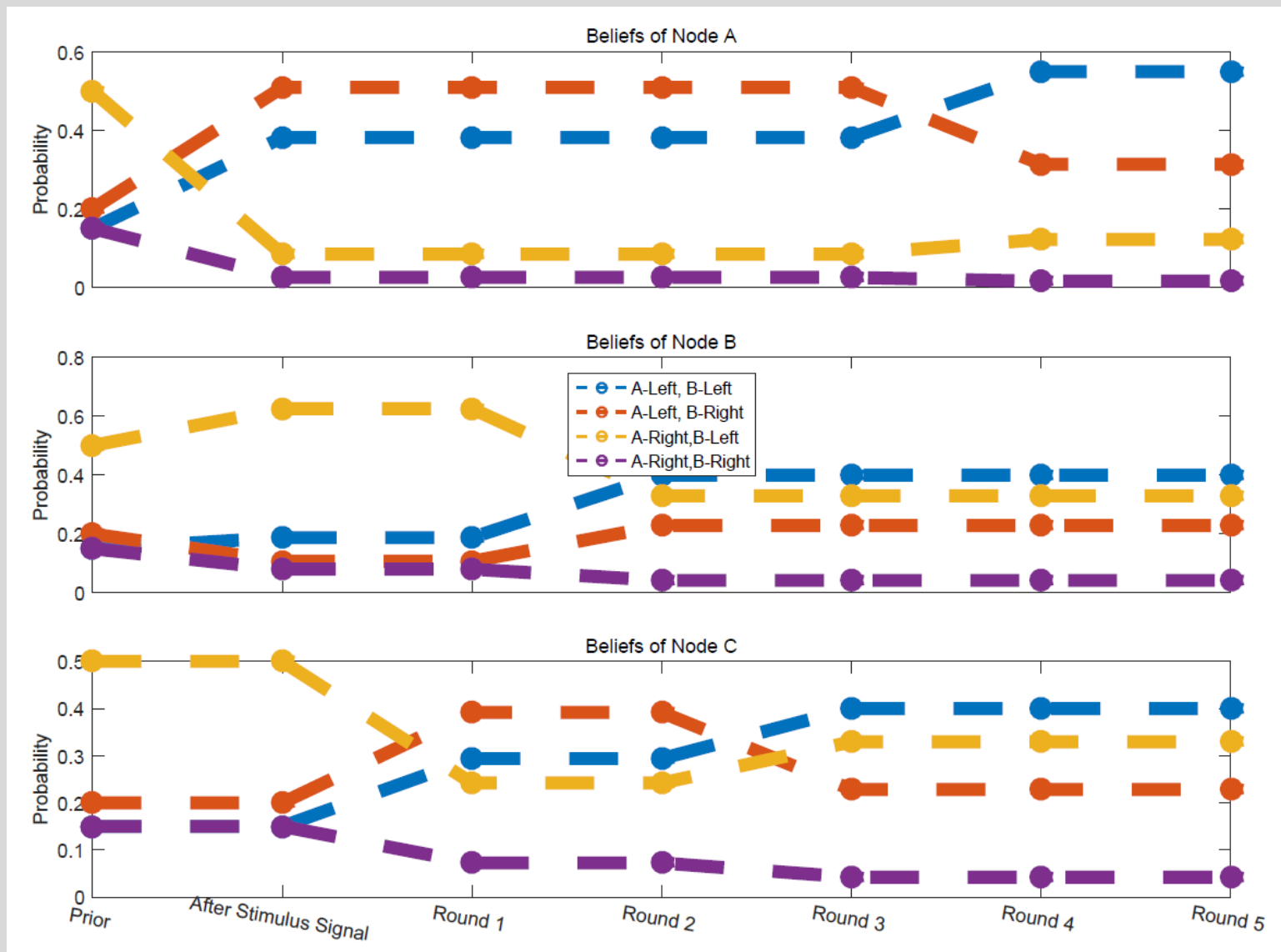


Each node computes its most likely underlying state and sends this as its message

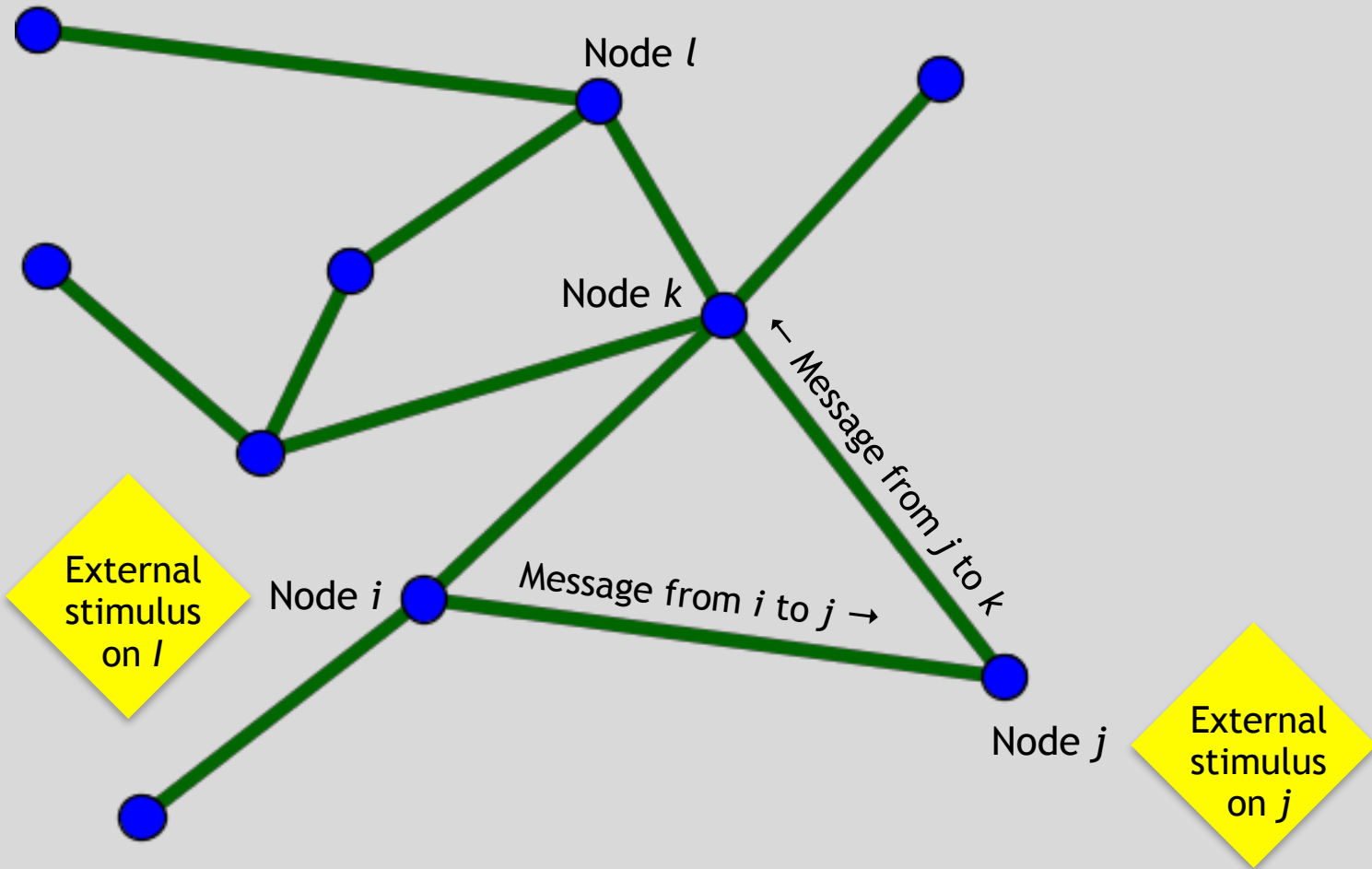
Underlying state	Prior probability	Conditional probability of signal				
		s_1^A	s_2^A	s_3^A	s_1^B	s_2^B
A-Left:B-Left	0.15	0.15	0.70	0.15	0.70	0.30
A-Left:B-Right	0.20	0.15	0.70	0.15	0.30	0.70
A-Right:B-Left	0.50	0.01	0.20	0.79	0.70	0.30
A-Right:B-Right	0.15	0.01	0.20	0.79	0.30	0.70

Numerical Example cont'd

Suppose the underlying state is A-Left:B-Left, node A receives signal s_1 , node B receives signal s_1



n -Node Networks with Uni-Directional Local Messages



n -Node Networks with Uni-Directional Local Messages

We now allow communication to be uni-directional

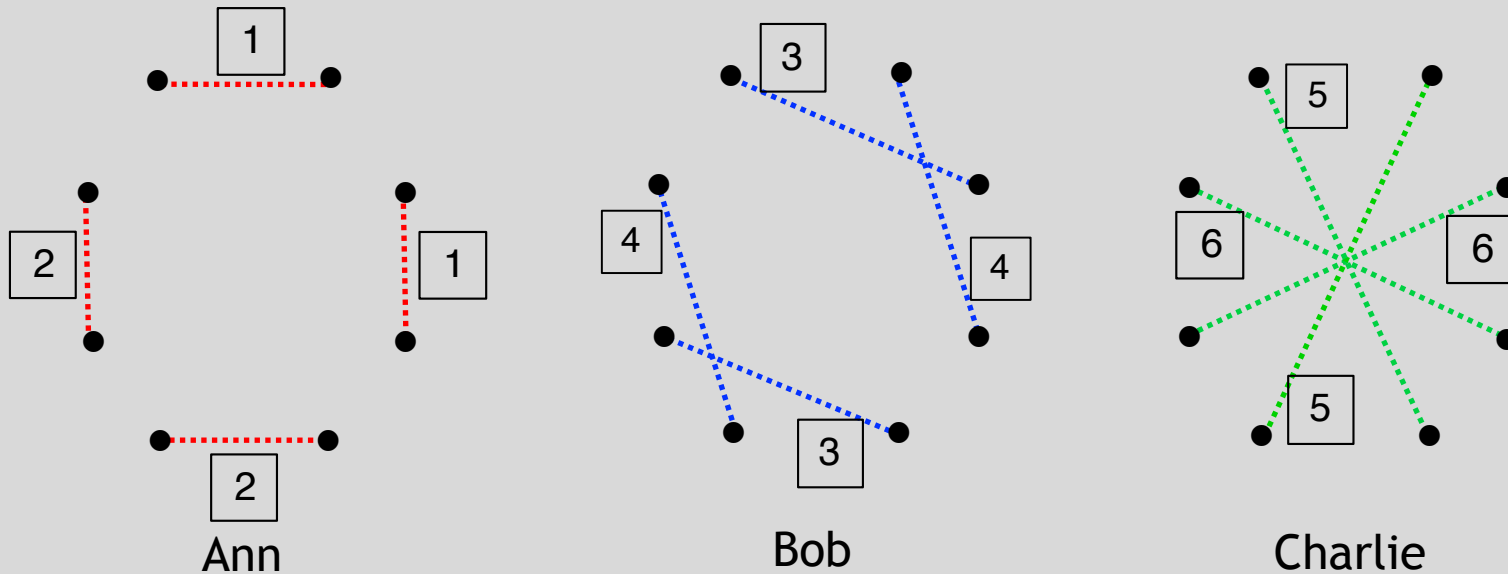
We assume the network is strongly connected, i.e., we can trace a path between any two nodes that respects the communication structure

Parikh and Krasucki (1990) gave an example which obeys union consistency, but agreement fails (see the picture on the next slide)

The example uses a message function with no obvious probabilistic interpretation

Open question: Would agreement result if more ‘natural’ message functions were employed?

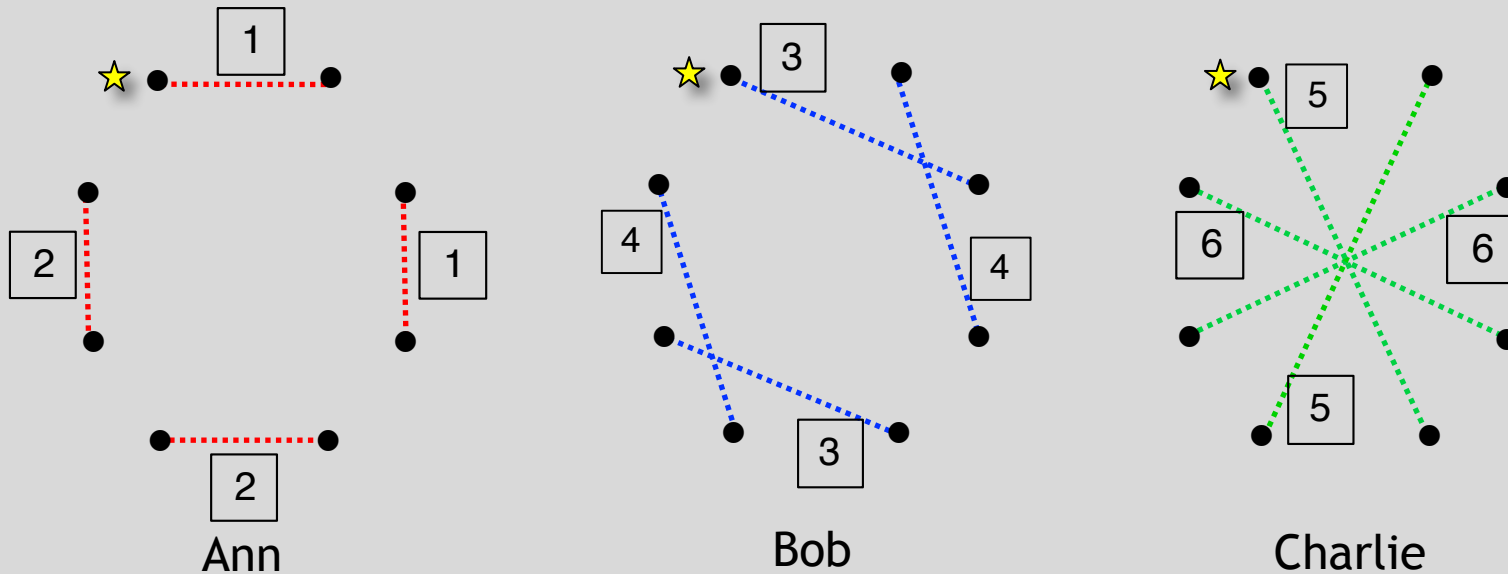
A Uni-Dimensional Network Exhibiting No Agreement



Ann sends messages to Bob, Bob sends messages to Charlie, and Charlie sends messages to Ann

Dashed lines indicate signals and boxes indicate messages

A Uni-Dimensional Network Exhibiting No Agreement cont'd



Suppose the starred state is realized

Then Bob learns nothing from Ann's message, Charlie learns nothing from Bob's message, and Ann learns nothing from Charlie's message

Agreement does not occur

Possible Connections to Neuroscience

Do different parts of the brain need to “agree,” i.e., to come to a shared view of the outside world?

- binocular rivalry
- feature binding
- (normalization with feedback)

The main lesson from game theory seems to be that a Bayesian network yields agreement under broad assumptions

Is this some support for the Bayesian brain hypothesis?

Limits of the results from game theory:

- union consistency condition
- bi-directional vs. uni-directional networks (maybe?)
- shared message function

Common Knowledge

So far we have talked about agreement in a network

The game theory literature also discusses “common knowledge”

(Common knowledge of a coin flip: I know the coin landed heads, I know you know it landed heads, I know you know I know it landed heads, and so on indefinitely)

If the network possesses a global clock, then agreement will be accompanied by common knowledge of agreement

But without a global clock, there can be agreement without common knowledge of agreement (Parikh and Krasucki 1990; Heifetz 1996)

Is there a neural counterpart to the game theorist’s concept of common knowledge?