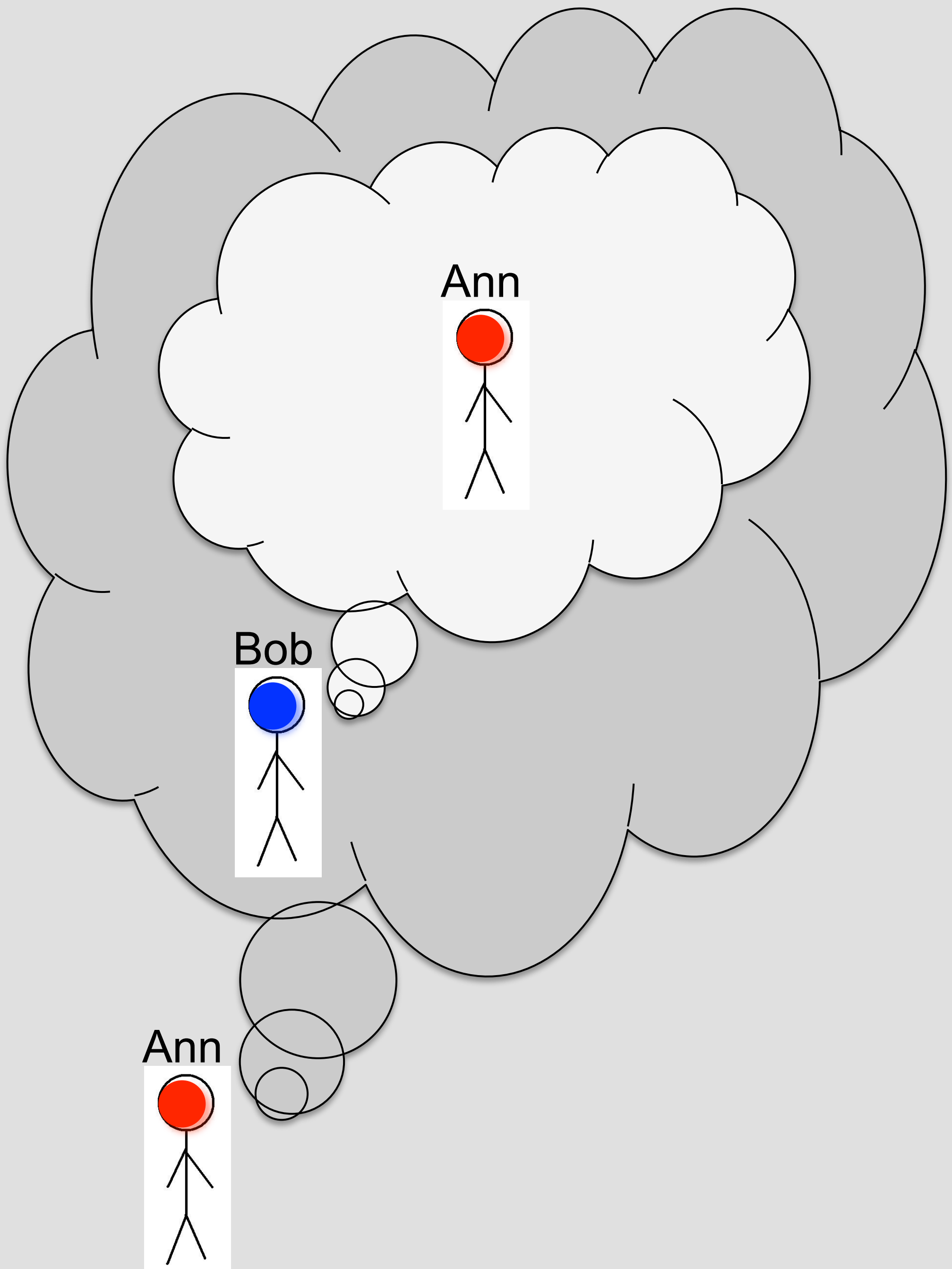
Thinking About Thinking and Its Cognitive Limits

Adam Brandenburger
Stern School of Business
Polytechnic School of Engineering
NYU Shanghai
New York University

Xiaomin Li
Behavioral & Social Neuroscience Program
California Institute of Technology


Ann
Bob
Ann

n players $S_1, S_2, \ldots, S_n$ are playing a given game of strategy, $\mathcal{G}$. How must one of the participants, S_m, play in order to achieve a most advantageous result?

[I]t is inherent in the concept of "strategy" that all information about the actions of the participants and the outcomes of "draws" [i.e., moves by Nature] a player is able to obtain or infer is already incorporated in the "strategy." Consequently, each player must choose his strategy in complete ignorance of the choices of the rest of the players and of the results of the "draws."

In minimax theory --- whether for the two-player zerosum case or the *n*-player non-zerosum case --- players avoid a predictive approach

Von Neumann, J., "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100, 1928, 295- 320; Bargman, S., English translation: "On the Theory of Games of Strategy," in Tucker, A., and R.D. Luce (eds.), *Contributions to the Theory of Games, Volume IV*, Princeton University Press, 1955. 13-42; Von Neumann, J., and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944

http://upload.wikimedia.org/wikipedia/commons/0/04/Nassau_Hall_Princeton.JPG
http://www.pnas.org/site/classics/classics5.xhtml
http://www.awesomestories.com/asset/view/John-Nash-Photo-as-a-Young-Man

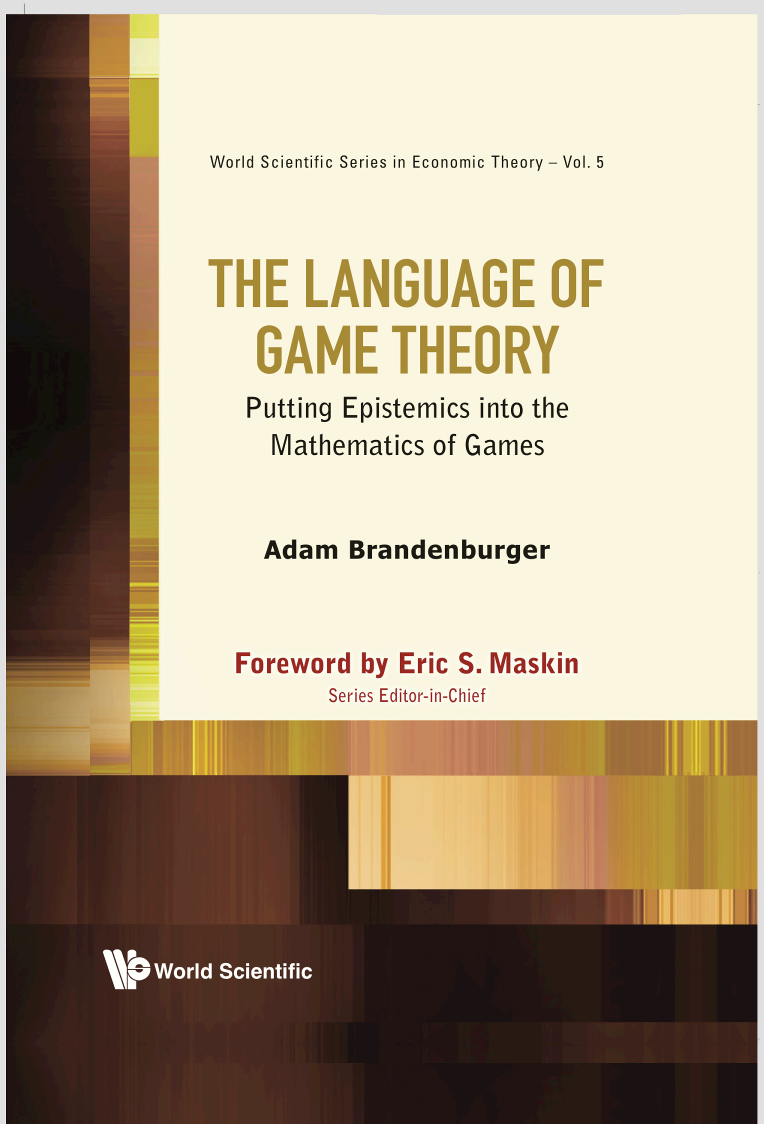
John von Neumann: "Each player must choose his strategy in 'complete ignorance'"

John Nash: "[A] rational prediction should be unique"

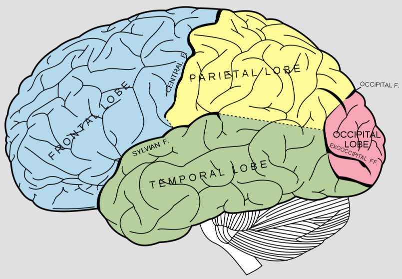
We proceed by investigating the question: what would be a "rational" prediction of the behavior to be expected of rational playing the game in question? By using the principles that a rational prediction should be unique, that the players should be able to deduce and make use of it, and that such knowledge on the part of each player of what to expect the others to do should not lead him to act out of conformity with the prediction, one is led to the concept of a solution defined before.

In equilibrium theory, players are assumed to have access to the actual strategies chosen by the other players

Nash, J., "Non-Cooperative Games," doctoral dissertation, Princeton University, 1950; available at http://rbsc.princeton.edu/topics/nash-john-1928-2015; Nash, J., "Non-Cooperative Games," *Annals of Mathematics*, 54, 1951, 286-295

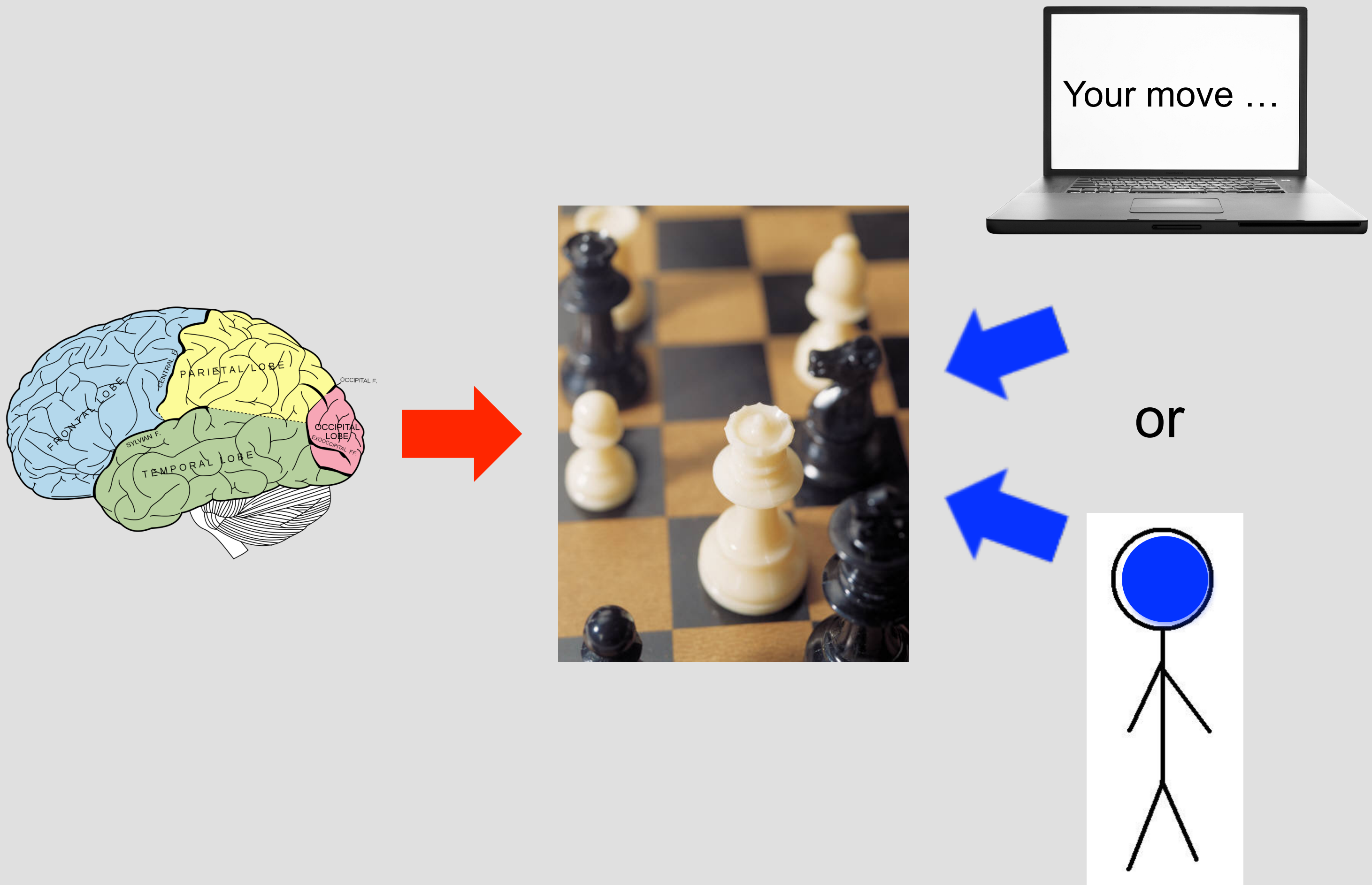

World Scientific Series in Economic Theory – Vol. 5
THE LANGUAGE OF
GAME THEORY
Putting Epistemics into the
Mathematics of Games
Adam Brandenburger
Foreword by Eric S. Maskin
Series Editor-in-Chief
World Scientific

Theory of Mind in tasks:



Gallagher, H. et al.,"Reading the Mind in Cartoons and Stories: An fMRI Study of 'Theory of Mind' in Verbal and Nonverbal Tasks," *Neuropsychologia*, 38, 2000, 11-21; http://upload.wikimedia.org/wikipedia/commons/thumb/1/1a/Gray728.svg/1024px-Gray728.svg.png

Theory of Mind in games:



McCabe, K. et al.,"A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange," *PNAS*, 98, 2001, 11832-11835; Gallagher, H., et al., "Imaging the Intentional Stance in a Competitive Game," *NeuroImage*, 16, 2002, 814-821; Rilling, J., et al., "The Neural Correlates of Theory of Mind Within Interpersonal Interactions," *NeuroImage*, 22, 2004, 1694-1703; http://upload.wikimedia.org/wikipedia/commons/thumb/1/1a/Gray728.svg/1024px-Gray728.svg.png

Theory of Mind ability:

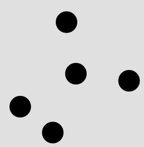
Stiller and Dunbar (2007)	Modal level of failure is 5 (−1?)
Arad and Rubinstein (2012)	Maximum level is 3
Kneeland (2015)	Maximum level is 4

A **cognitive limit** on thinking about thinking comes into effect at a small finite number of levels

Stiller, J., and R. Dunbar, "Perspective-Taking and Memory Capacity Predict Social Network Size," *Social Networks*, 29, 2007, 93-104; Arad, A., and A. Rubinstein, "The 11-20 Money Request Game: A Level-*k* Reasoning Study," *American Economic Review*, 102, 2012, 3561-3573; Kneeland, T., "Identifying Higher-Order Rationality," 2015, forthcoming in *Econometrica* , available at http://terri.microeconomics.ca

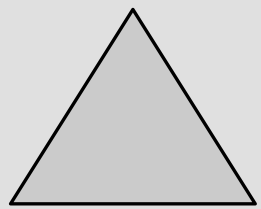
Models of thinking about thinking:

In game theory, uncertainty about the **structure** of the game (Harsanyi 1967-68) or about the **strategies** in the game (epistemic game theory) has been modeled via sequences of spaces



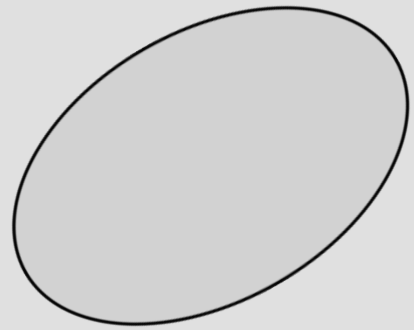
S

0th-order space



$\wp(S)$

1st-order space



$\wp(\wp(S))$

2nd-order space

...

Under standard regularity assumptions, the sizes of the spaces are:

$|S|$ $\quad 2^{\aleph_0} \quad 2^{\aleph_0} \quad \ldots$

Harsanyi, J., "Games with Incomplete Information Played by 'Bayesian' Players, I-III," *Management Science*, 14, 1967-8, 159-182, 320-334, 486-502

Neural evidence in Bhatt and Camerer (2005): Players were asked to make choices and to state first- and second-order beliefs in games

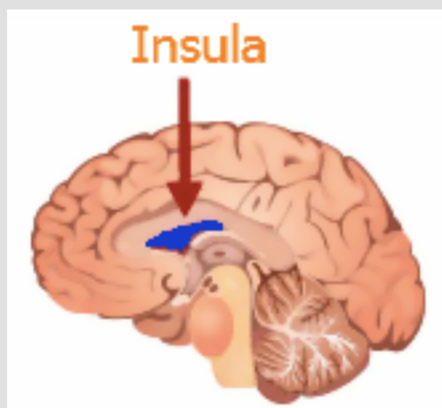
$$\hat{s}_a \quad (1)$$

$$\Box_a \, \hat{s}_b \quad (2)$$

$$\Box_a \, \Box_b \, \hat{r}_a \quad (3)$$

Brain activity in event (3) was found to have more similarities with activity in event (1) than with activity in event (2)

Greater activity was found in the anterior insula in event (3) than in event (2)



Bhatt, M., and C. Camerer, "Self-Referential Thinking and Equilibrium as States of Mind in Games: fMRI Evidence," *Games and Economic Behavior*, 52, 2005, 424-459; http://www.easynotecards.com/print_cards/29867

An anchoring and adjusting process:

Ann selects a candidate strategy choice (she **anchors**)

She then examines her view as to whether or not Bob thinks she intends to make this choice (she **adjusts**)

There is precedent for anchoring and adjusting processes in the Theory of Mind literature, e.g., in gauging another individual's preferences (Epley, Keysar, Van Boven, and Gilovich 2004; Tamir and Mitchell 2010)

In some sense, this process can be thought of as an (internal) **equilibrium** vs. **disequilibrium** process

Epley, N., B. Keysar, L. Van Boven, and T. Gilovich, "Perspective Taking as Egocentric Anchoring and Adjustment," *Journal of Personality and Social Psychology*, 87, 2004, 327-339; Tamir, D., and J. Mitchell, "Neural Correlates of Anchoring-and-Adjustment During Mentalizing," *PNAS*, 107, 2010, 10827-10832

Say Ann is **rational** if $\hat{s}_a$ and $\Box_a \hat{r}_b$ imply that s_a maximizes Ann's (expected) payoff when she assigns probability 1 to Bob's choosing r_b. Define rationality for Bob similarly (with Ann and Bob interchanged).

Epistemic conditions for Nash equilibrium:

$\hat{s}_a$	$\Box_a \hat{s}_b$
$\hat{s}_b$	$\Box_b \hat{s}_a$
$\Box_a \hat{s}_b$	$\Box_a \Box_b \hat{s}_a$
$\Box_b \hat{s}_a$	$\Box_b \Box_a \hat{s}_b$
Ann is rational	$\Box_a$ [Bob is rational]
Bob is rational	$\Box_b$ [Ann is rational]

Aumann, R., and A. Brandenburger, "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1995, 1161-1180

Axioms:

$\hat{s}_a \to \Box_a \hat{s}_a$, $\Box_a \hat{s}_b \to \Box_a \Box_a \hat{s}_b$, and [Ann is rational] $\to \Box_a$ [Ann is rational]

Epistemic conditions subjectivized:

$\hat{s}_a$	➡	$\hat{s}_a$
$\hat{s}_b$		$\Box_a \hat{s}_b$
$\Box_a \hat{s}_b$		$\Box_a \Box_b \hat{s}_a$
$\Box_b \hat{s}_a$		Ann is rational
Ann is rational		$\Box_a$ [Bob is rational]
Bob is rational		

Disequilibrium:

Change Ann’s second-order belief relative to her choice via negations, and then push negation in

$\neg \Box_a \Box_b \hat{s}_a$		$\Diamond_a \Diamond_b \neg \hat{s}_a$
$\Box_a \neg \Box_b \hat{s}_a$		$\Box_a \Diamond_b \neg \hat{s}_a$
$\Box_a \Box_b \neg \hat{s}_a$		$\Box_a \Box_b \neg \hat{s}_a$
$\neg \Box_a \neg \Box_b \hat{s}_a$	$\Rightarrow$	$\Diamond_a \Box_b \hat{s}_a$
$\Box_a \neg \Box_b \neg \hat{s}_a$		$\Box_a \Diamond_b \hat{s}_a$
$\neg \Box_a \Box_b \neg \hat{s}_a$		$\Diamond_a \Diamond_b \hat{s}_a$
$\neg \Box_a \neg \Box_b \neg \hat{s}_a$		$\Diamond_a \Box_b \neg \hat{s}_a$

Disequilibrium cont'd:

Disqualify conditions consistent with equilibrium

$\Diamond_a \, \Diamond_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Box_a \, \Diamond_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Box_a \, \Box_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Diamond_a \, \Box_b \, \hat{\boldsymbol{s}}_a$

$\Box_a \, \Diamond_b \, \hat{\boldsymbol{s}}_a$

$\Diamond_a \, \Diamond_b \, \hat{\boldsymbol{s}}_a$

$\Diamond_a \, \Box_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Diamond_a \, \Diamond_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Box_a \, \Diamond_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Box_a \, \Box_b \, \neg \, \hat{\boldsymbol{s}}_a$

$\Diamond_a \, \Box_b \, \neg \, \hat{\boldsymbol{s}}_a$

Following the analogous process with the second set of epistemic conditions for Nash equilibrium:

$\Box_a \, \hat{s}_b$

$\Box_b \, \hat{s}_a$

$\Box_a \, \Box_b \, \hat{s}_a$

$\Box_b \, \Box_a \, \hat{s}_b$

$\Box_a$ [Bob is rational]

$\Box_b$ [Ann is rational]

⇒

$\Box_a \, \hat{s}_b$

$\Box_a \, \Box_b \, \hat{s}_a$

$\Box_a \, \Box_b \, \Box_a \, \hat{s}_b$

$\Box_a$ [Bob is rational]

$\Box_a \, \Box_b$ [Ann is rational]

⇒

$\Diamond_a \, \Diamond_b \, \Diamond_a \, \neg \, \hat{s}_b$

$\Box_a \, \Diamond_b \, \Diamond_a \, \neg \, \hat{s}_b$

$\Box_a \, \Box_b \, \Diamond_a \, \neg \, \hat{s}_b$

$\Box_a \, \Box_b \, \Box_a \, \neg \, \hat{s}_b$

$\Diamond_a \, \Box_b \, \Box_a \, \neg \, \hat{s}_b$

$\Diamond_a \, \Box_b \, \Diamond_a \, \neg \, \hat{s}_b$

$\Box_a \, \Diamond_b \, \Box_a \, \neg \, \hat{s}_b$

$\Diamond_a \, \Diamond_b \, \Box_b \, \neg \, \hat{s}_b$

Complexity of thinking about thinking:

Ann (tentatively) fixes a strategy-pair (s_a, s_b)

In forming an *m*th-order belief, she considers $\mathbf{2^m + 1}$ cases

The process could be repeated for different candidate strategy pairs

The **exponential** increase in the number of cases at each level fits qualitatively with a small finite cognitive bound (we would not expect quantitative agreement)

To-do's:

Extend experiments to collect data on higher-order beliefs

Distinguish more carefully internal epistemic equilibrium from reasoning about levels of rationality

Study games with more than two players

Examine belief modalities in-between probabilistic and point-belief