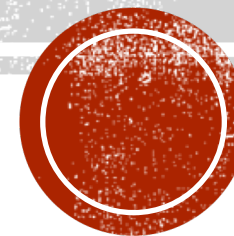


# НЕЙРОГЕРМЕНЕВТИКА КАРЛА ФРИСТОНА

Игорь Ф. Михайлов

Междисциплинарные проблемы когнитивных наук

Москва, Институт философии РАН, 15.10.2019



# КАРЛ ФРИСТОН



Британский нейробиолог и специалист по визуализации мозга, член Британской Медицинской Академии (1999), член Королевского научного общества (2006), научный директор Лаборатории функциональных изображений при Университетском колледже Лондона, профессор Лондонского университета.

# НЕЙРОГЕРМЕНЕВТИКА

Критерии оценки и подстройки  
моей интерпретации вашего  
поведения — это те же критерии,  
которые лежат в основе действий  
и восприятия в целом; а именно,  
**минимизация ошибки  
предсказания или  
(вариационной) свободной  
энергии.**

Представление о  
коммуникации,  
основанное на  
**генеративной модели  
или нарративе,**  
которой разделяют  
агенты,  
обменивающиеся  
сенсорными сигналами.

Мы пытаемся смоделировать  
причины наших ощущений и  
скорректировать эти модели,  
чтобы **максимизировать  
свидетельства байесовской  
модели или, что  
эквивалентно,  
минимизировать удивление.**

# НЕЙРОГЕРМЕНЕВТИКА

Критерии оценки и подстройки моей интерпретации вашего поведения — это те же критерии, которые лежат в основе действий и восприятия в целом; а именно, **минимизация ошибки предсказания или (вариационной) свободной энергии.**

Представление о коммуникации, основанное на **генеративной модели или нарративе**, которой разделяют агенты, обменивающиеся сенсорными сигналами.

Мы пытаемся смоделировать причины наших ощущений и скорректировать эти модели, чтобы **максимизировать свидетельства байесовской модели или, что эквивалентно, минимизировать удивление.**

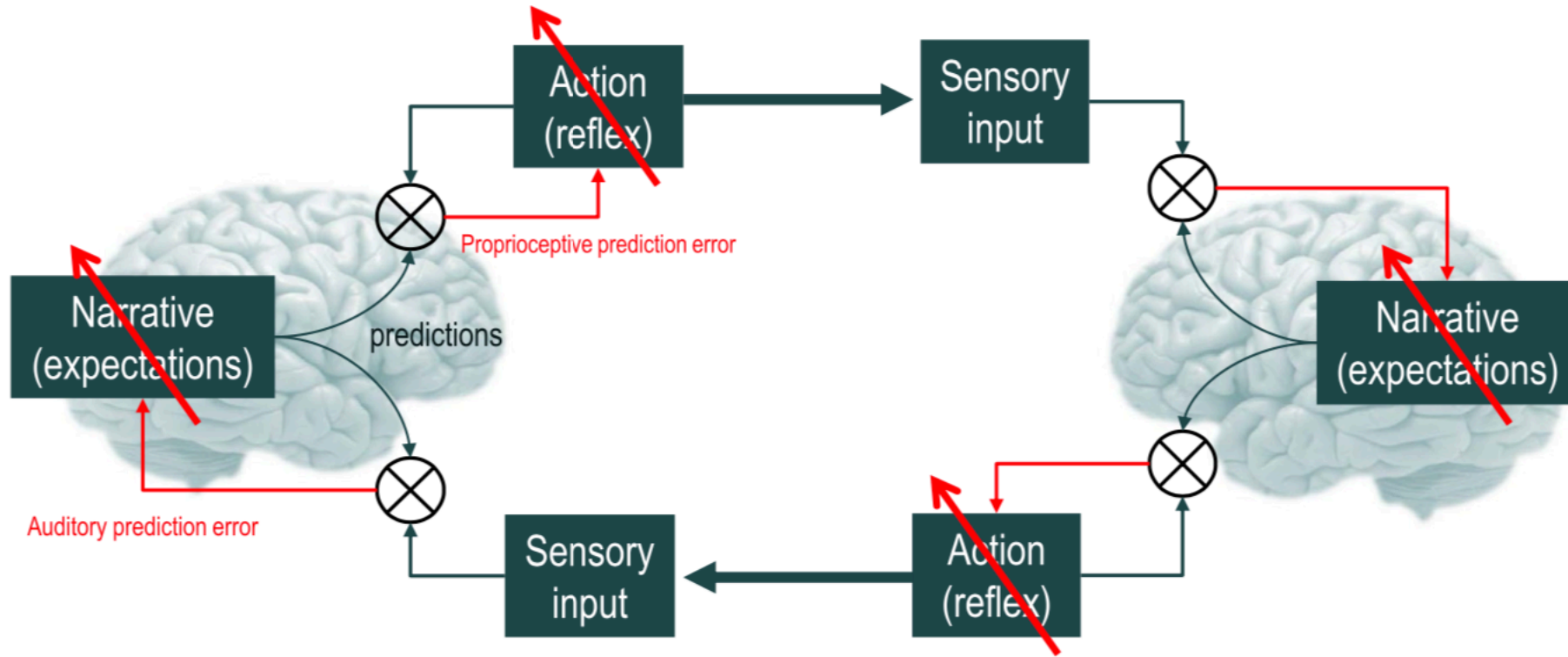
# НЕЙРОГЕРМЕНЕВТИКА

Критерии оценки и подстройки моей интерпретации вашего поведения — это те же критерии, которые лежат в основе действий и восприятия в целом; а именно, **минимизация ошибки предсказания или (вариационной) свободной энергии.**

Представление о коммуникации, основанное на **генеративной модели или нарративе**, которой разделяют агенты, обменивающиеся сенсорными сигналами.

Мы пытаемся смоделировать причины наших ощущений и скорректировать эти модели, чтобы **максимизировать свидетельства байесовской модели или, что эквивалентно, минимизировать удивление.**

# НЕЙРОГЕРМЕНЕВТИКА



# НЕЙРОГЕРМЕНЕВТИКА

Герменевтический круг можно закрыть, просто обновив генеративные модели и их прогнозы, чтобы минимизировать ошибки предсказания. Важно отметить, что эти ошибки предсказания могут быть вычислены, даже не зная истинного состояния другого; что, тем самым, решает проблему герменевтики (Friston and Frith, 2015a, 129–130).

Ср. Витгенштейн: ребёнок **понял**, что такое натуральный ряд чисел, просто когда научился его правильно продолжить с любого числа.

# НЕЙРОГЕРМЕНЕВТИКА

Герменевтический круг можно закрыть, просто обновив генеративные модели и их прогнозы, чтобы минимизировать ошибки предсказания. Важно отметить, что эти ошибки предсказания могут быть вычислены, даже не зная истинного состояния другого; что, тем самым, решает проблему герменевтики (Friston and Frith, 2015a, 129–130).

Ср. Витгенштейн: ребёнок **понял**, что такое натуральный ряд чисел, просто когда научился его правильно продолжить с любого числа.



# ПРЕДИКТИВНОЕ КОДИРОВАНИЕ (PREDICTIVE CODING)

- Восходит к идее **Г. фон Гельмгольца** о «бессознательных выводах».
- Формируются в раннем жизненном опыте и служат основой многих перцептивных явлений.
- Мы настраиваем свои органы чувств для различения воздействий с максимальной точностью.
- Восприятие как результат встречи идущего из внешнего мира и того, что уже есть у индивида.



# ПРЕДИКТИВНОЕ КОДИРОВАНИЕ (PREDICTIVE CODING)

- Clark 2013
- Hohwy 2013
- Friston & Kiebel 2009





Многослойная иерархическая архитектура когнитивного аппарата.



Верхние слои формируют байесовские модели (предполагаемой) реальности и транслируют их вниз.



Нижние слои сравнивают эти модели с полученными снизу (извне) данными, вычисляют ошибки предсказания и транслируют их вверх.



«Верхние» модели уточняются, и цикл повторяется.

## ПРЕДИКТИВНОЕ КОДИРОВАНИЕ (PREDICTIVE CODING)

# ПРЕДИКТИВНОЕ КОДИРОВАНИЕ (PREDICTIVE CODING)

# ПРЕДИКТИВНОЕ КОДИРОВАНИЕ (PREDICTIVE CODING)

Предиктивное кодирование представляет собой биологически правдоподобную схему для обновления представлений о состояниях мира с использованием сенсорных образцов.

Эти нисходящие прогнозы сравниваются с репрезентациями на более низком уровне, чтобы сформировать ошибку предсказания (обычно связанную с активностью поверхностных пирамидальных клеток) (Friston and Frith, 2015b, 392).

# ПРЕДИКТИВНОЕ КОДИРОВАНИЕ (PREDICTIVE CODING)

Предиктивное кодирование представляет собой биологически правдоподобную схему для обновления представлений о состояниях мира с использованием сенсорных образцов.

Эти нисходящие прогнозы сравниваются с репрезентациями на более низком уровне, чтобы сформировать ошибку предсказания (обычно связанную с активностью поверхностных пирамидальных клеток) (Friston and Frith, 2015b, 392).

# ВЫСОКОТОЧНАЯ ПЕРЕДАЧА СООБЩЕНИЙ

Ошибки прогнозирования могут быть затем взвешены по их точности, так что только точная информация накапливается и усваивается на высоком или глубоком иерархическом уровне. Что касается всех ожиданий, ожидаемая точность максимизирует свидетельства байесовской модели.

Трансляция ошибок взвешенного по точности предсказания основывается на регулировании силы входа на синаптическом уровне ... Эта нейромодуляторное управление силой входного сигнала соответствует (байесово-оптимальному) кодированию точности с точки зрения возбудимости популяций нейронов, сообщающих об ошибках предсказания (Friston and Frith, 2015b, 392).

# ВЫСОКОТОЧНАЯ ПЕРЕДАЧА СООБЩЕНИЙ

Ошибки прогнозирования могут быть затем взвешены по их точности, так что только точная информация накапливается и усваивается на высоком или глубоком иерархическом уровне. Что касается всех ожиданий, ожидаемая точность максимизирует свидетельства байесовской модели.

Трансляция ошибок взвешенного по точности предсказания основывается на регулировании силы входа на синаптическом уровне ... Эта нейромодуляторное управление силой входного сигнала соответствует (байесово-оптимальному) кодированию точности с точки зрения возбудимости популяций нейронов, сообщающих об ошибках предсказания (Friston and Frith, 2015b, 392).



Ещё один способ минимизировать ошибки предсказания; а именно путем повторной выборки сенсорных данных, чтобы они соответствовали прогнозам; другими словами, изменяя сенсорные входы, изменяя мир через действие.

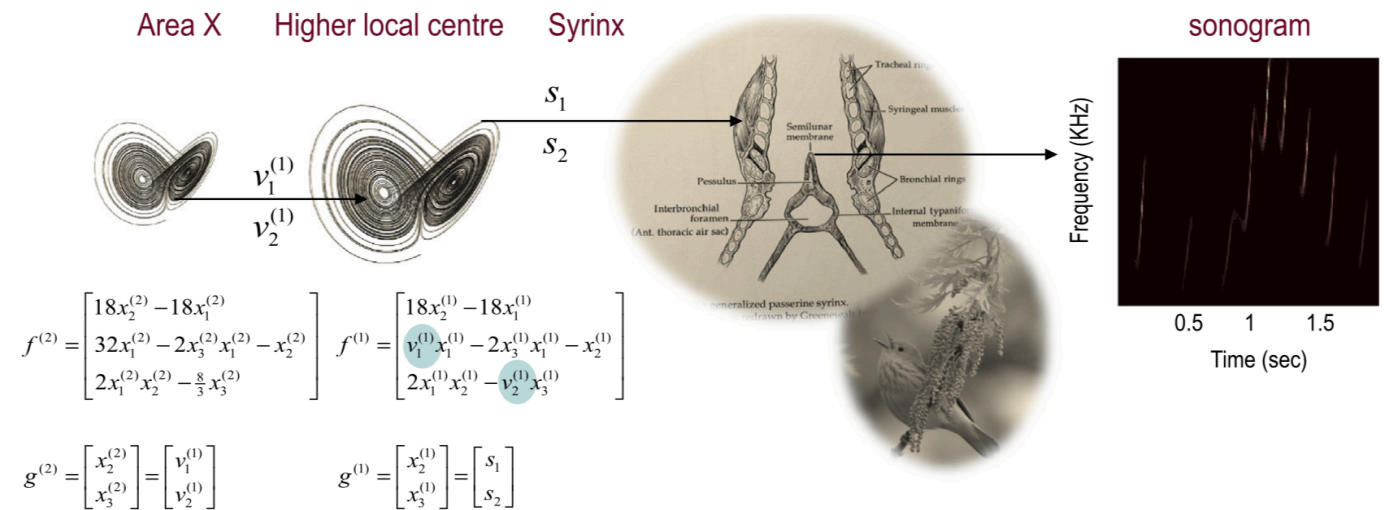
Одни и те же механизмы вывода лежат в основе, казалось бы, различных функций (действия, познания и восприятия).

Другими словами, действие и восприятие являются аспектами одного и того же основного императива; а именно — минимизировать ошибки иерархического предсказания путем выборочного тестирования сенсорных входов (Friston and Frith, 2015b, 393)..

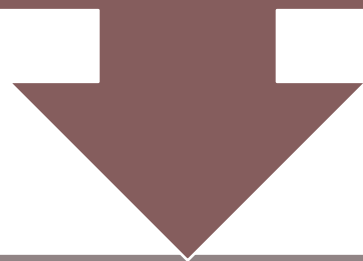
# АКТИВНЫЙ ВЫВОД (ACTIVE INFERENCE)

# АТТРАКТОРЫ

- Модели, основанные на аттракторах, могут генерировать и, следовательно, кодировать структурированные последовательности событий, когда состояния протекают по разным частям многообразия.
- Иерархически развернутые аттракторы позволяют мозгу предсказывать или представлять различные категории последовательностей (Friston and Frith, 2015b, 394).



Проще говоря, мы не можем говорить и слушать одновременно.



Эвристически это предполагает, что активный вывод представлен в одном из двух режимов; или забота о ощущениях или действия во время периодов сенсорного ослабления (Friston and Frith, 2015b, 398).

## СЕНСОРНОЕ ОСЛАБЛЕНИЕ И ДЕЙСТВИЕ

# 01

Действие и восприятие зависят от тонкого баланса между точностью проприоцептивных и экстероцептивных ошибок предсказания, которые управляют перцептивным синтезом ощущений (вызванных другими) или реализацией сенсорных предсказаний (вызванных самим собой).

# 02

Интересно порассуждать о психопатологии, которая может сопровождать собой сенсорного ослабления (Friston and Frith, 2015b, 398).

# 03

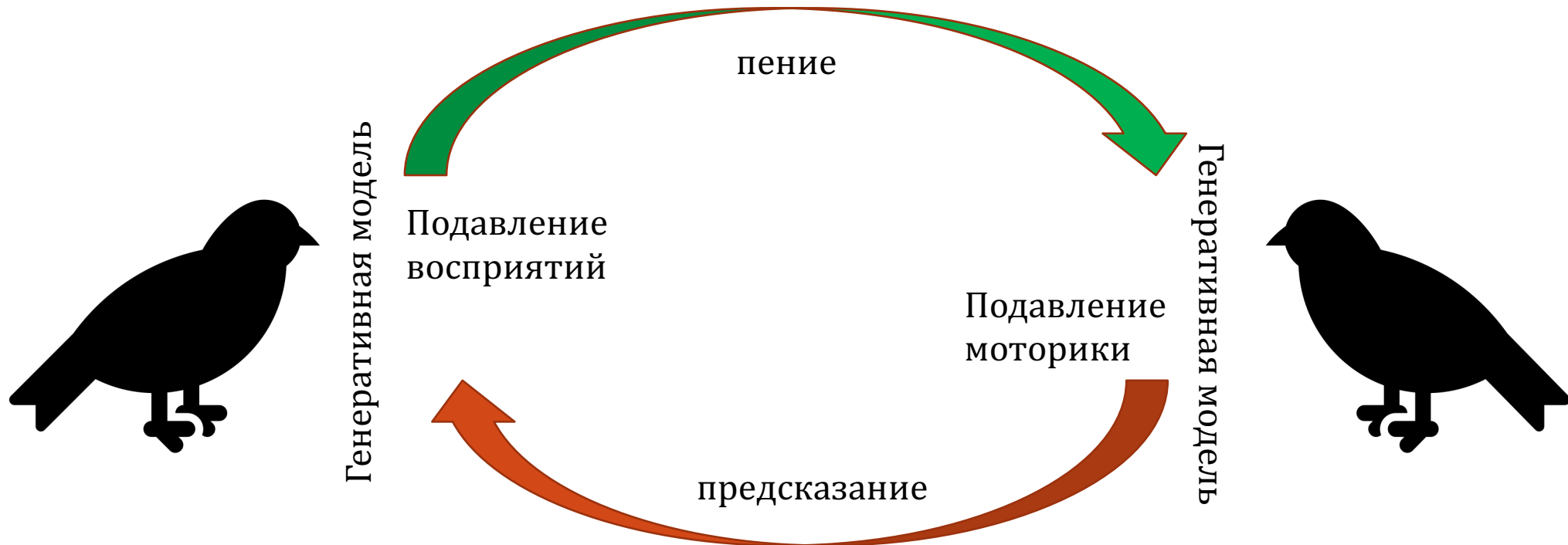
Недостаток сенсорного ослабления — в частности, относительной силы сенсорной и предыдущей (экстрасенсорной) точности — был предложен в качестве основы аутизма (р. 401).

## СЕНСОРНОЕ ОСЛАБЛЕНИЕ И ДЕЙСТВИЕ

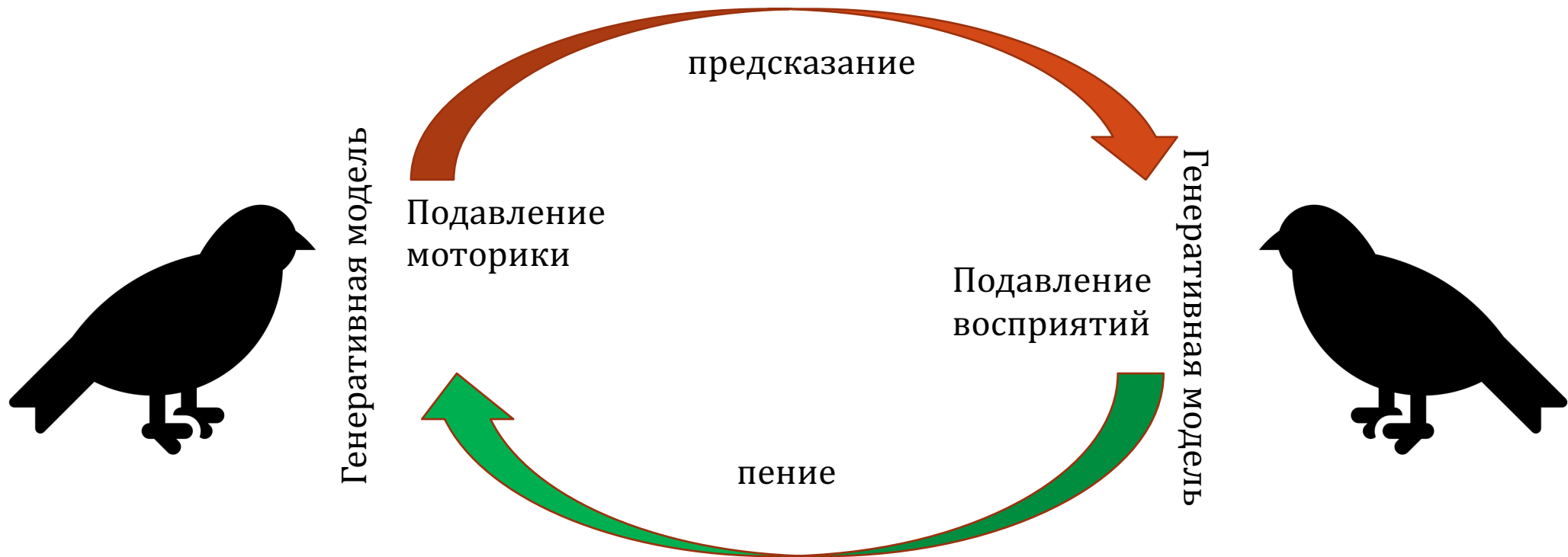
# СВОБОДНАЯ ЭНЕРГИЯ И ЭНТРОПИЯ

- Минимизация свободной энергии посредством активного вывода неявно сводит к минимуму неожиданность (или ошибки предсказания), что эквивалентно максимизации очевидности байесовской модели.
- Долгосрочное среднее значение неожиданности называется энтропией, что означает, что минимизация неожиданности, минимизирует (теоретико-информационную) энтропию.
- Мера объема аттрактора обеспечивается его мерно-теоретической энтропией... Хотя она формально отличается от информационно-теоретической энтропии, они обе отражают объем аттракторного множества динамических состояний, занимаемых рассматриваемыми системами. Это означает, что минимизация свободной энергии (информационно-теоретической энтропии) уменьшает объем (мерно-теоретической энтропию) случайного динамического аттрактора (многообразия синхронизации); тем самым порождая обобщённую синхронию.

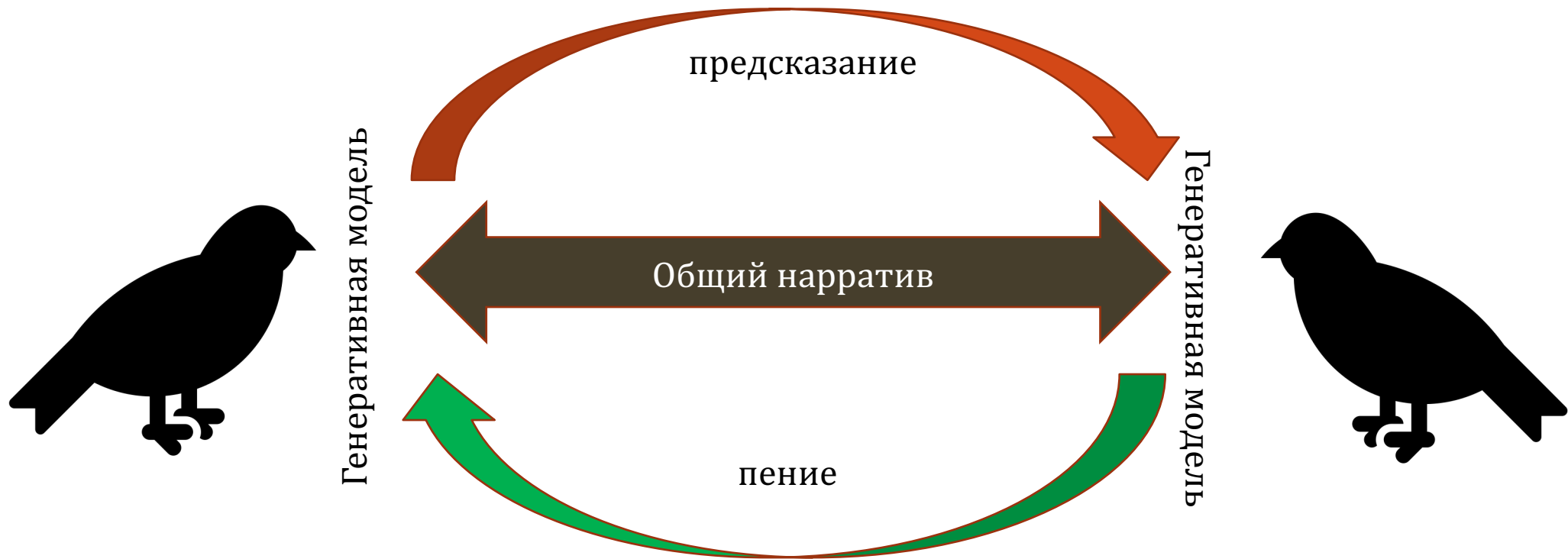
# ПТИЧЬЕ ПЕНИЕ (СИМУЛЯЦИЯ)



# ПТИЧЬЕ ПЕНИЕ (СИМУЛЯЦИЯ)



# ПТИЧЬЕ ПЕНИЕ (СИМУЛЯЦИЯ)





# ПТИЧЬЕ ПЕНИЕ (СИМУЛЯЦИЯ)

Каждая птица **слушала в течение двух секунд** (с низкой проприоцептивной точностью и высокой экстероцептивной точностью), а затем **пела в течение двух секунд** (с высокой проприоцептивной точностью и ослабленной слуховой точностью). Другими словами, птицы не просто повторяют то, что они слышали, они следуют за **нарративом, воплощенным динамическими аттракторами** (центральными генераторами паттернов) в их генеративных моделях, которые были синхронизированы посредством сенсорного обмена. Как отмечалось выше, это означает, что обе птицы могут петь из одного нотного листа, сохраняя последовательную и иерархическую структуру в своем общем повествовании. Именно это явление — благодаря просто обобщенной (в данном случае идентичной) синхронизации внутренних состояний — мы связываем с **коммуникацией** (Friston and Frith, 2015b, 400).

# ПТИЧЬЕ ПЕНИЕ (СИМУЛЯЦИЯ)

- Основная идея заключается в том, что **внутренние или генеративные модели, используемые для определения собственного поведения, могут быть использованы для вывода убеждений (например, намерений) другого** - при условии, что обе стороны имеют достаточно сходные генеративные модели.
  - (Ссылка на **зеркальные нейроны**.)
  - Эта перспектива создаёт репрезентации набора интенциональных актов и нарративов почти **юнгианского характера** — предполагая коллективное повествование, которое разделяют общающиеся агенты (включая себя).
- (Friston and Frith, 2015b, 401).

# НОВЫЙ АССОЦИАЦИОНИЗМ

- Ryder & Favorov 2001
- Ассоциационизм был призван свести разум к нерациональному. Именно в этом смысле теория Юма была механистической.
- Критика (тогдашняя — Кант — и современная — напр., Фодор): ассоциации — слишком примитивный механизм, чтобы объяснить рациональный вывод.
- В отличие от ассоциаций, юмова теория абстракций не механистична. Она не объясняет принцип выбора сознанием «общих признаков» для абстрагирования.



# НОВЫЙ АССОЦИАЦИОНИЗМ

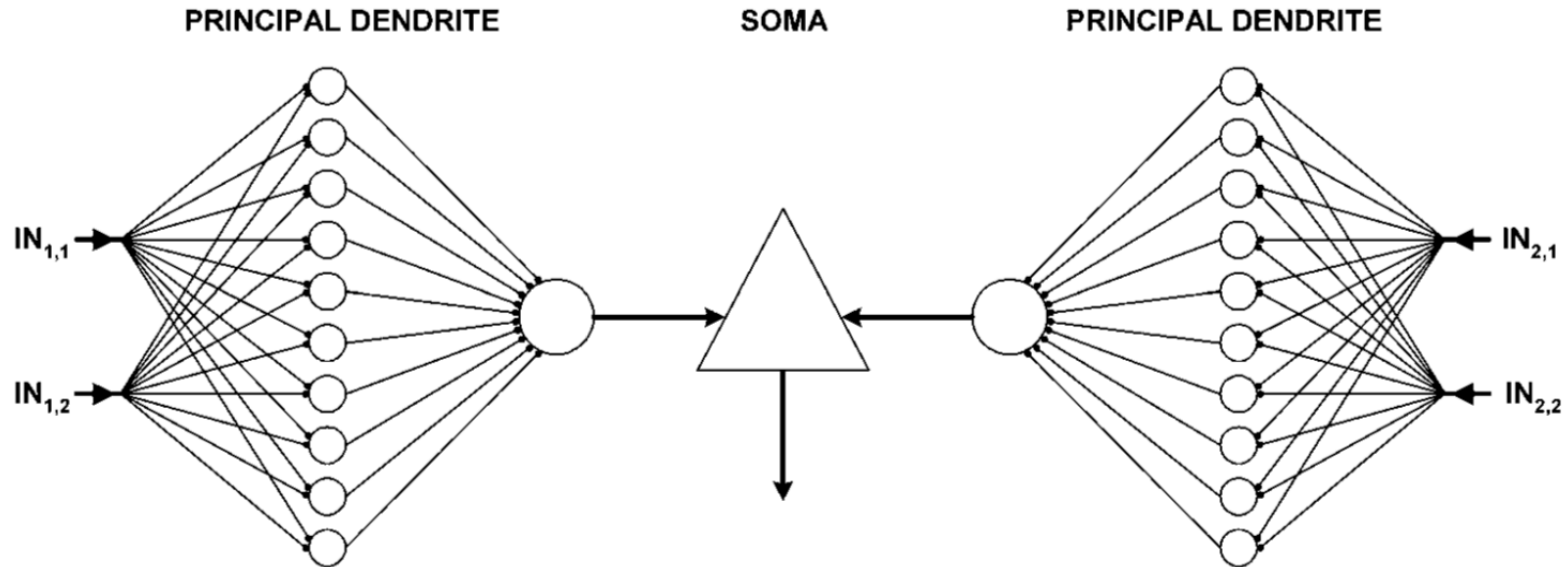


Figure 1. The SINBAD model of a pyramidal cell with two principal dendrites connected to the soma (shown as a triangle). Each principal dendrite is modeled as an error backpropagation network with one output unit, a single layer of ten hidden units, and two input channels.

- Ryder & Favorov 2001
- 'SINBAD' model: a Set of INteracting BAcKpropagating Dendrites
- «дендриты»: 2 сети, связанные через тело «клетки»
- алгоритм обратного распространения ошибок
- выходные данные «дендритов» складываются вместе для получения выходных данных «клетки»
- выход ячейки также используется в качестве обучающего сигнала для каждого дендрита
- Процесс подстройки весов соединений будет продолжаться до тех пор, пока каждый дендрит не научится предсказывать — на основе своих собственных входных данных — ответы другого дендрита на свои входные данные.

# НОВЫЙ АССОЦИОНИЗМ

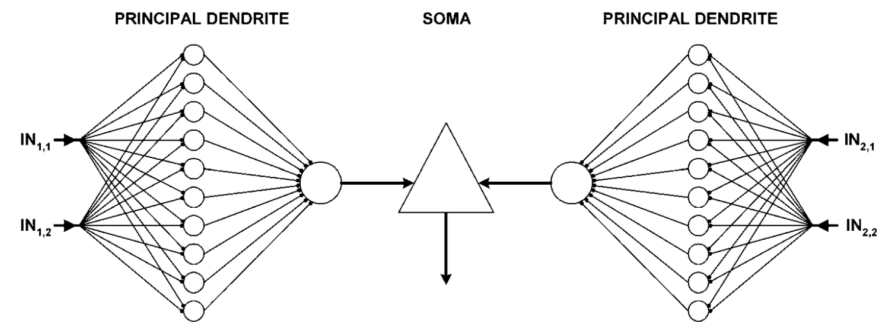
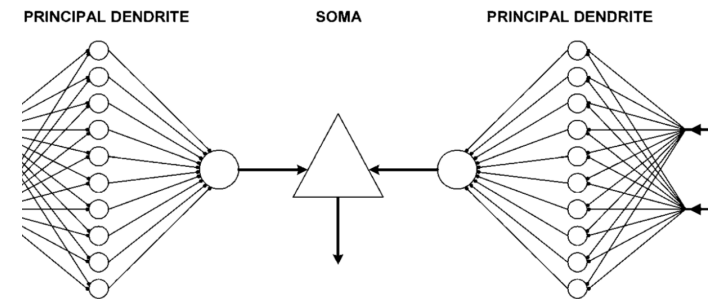


Figure 1. The SINBAD model of a pyramidal cell with two principal dendrites connected to the soma (shown as a triangle). Each principal dendrite is modeled as an error backpropagation network with one output unit, a single layer of ten hidden units, and two input channels.

- (1) обучая друг друга, дендриты в клетке могут настраиваться на разные, но коррелированные, выявляющие порядок функции условий окружающей среды,
- и (2) клетка в целом может научиться распознавать источник этих коррелированных функций как упорядоченное свойство её сенсорной среды.
- Первая способность реализует ассоциацию функций по условиям, в отличие от простой ассоциации Юма.
- Вторая способность (которая на самом деле является лишь обратной стороной первой) реализует процесс получения репрезентации, который заменяет абстракцию Юма.

# НОВЫЙ АССОЦИАЦИОНИЗМ

- Деятельность, которая начинается на периферии (как сенсорные данные), распространяется по сети таким образом, что реализует не только простую индукцию, но и дедуктивные рассуждения и вывод к лучшему объяснению. Это показывает, как простой биологически реалистичный механизм может раскрыть сложность порядка в природе и использовать эти знания для жизненно важной задачи предсказания [Ryder & Favorov 2001, 191].
- Теория Райдера–Фаворова, появившись десятилетием ранее **Predictive Coding**, предлагает механистическое объяснение нейродинамической реализации «генеративных моделей» в смысле Фристана,



re 1. The SINBAD model of a pyramidal cell with two principal dendrites connect soma (shown as a triangle). Each principal dendrite is modeled as an error backpropagation network with one output unit, a single layer of ten hidden units, and two input channels.

# ЛИТЕРАТУРА

Спасибо за внимание!  
ifmikhailov@gmail.com

- K. Friston, C. Frith (2015a). Active inference, communication and hermeneutics // Cortex 68, 129–143
- K. Friston, C. Frith (2015b). A Duet for one // Consciousness and Cognition 36, 390–405
- Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 1211e1221.
- Helmholtz, H. (1866/1962). Concerning the perceptions in general. In Treatise on physiological optics (J. Southall, Trans., 3rd ed., Vol. III). Dover, New York.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.
- Hohwy, J. (2013). The Predictive Mind. Oxford: Oxford University Press.
- Ryder, D., Favorov, O. L. (2001) The New Associationism: A Neural Explanation for the Predictive Powers of Cerebral Cortex // Brain and Mind 2: 161–194.