

# Сознание – логически непротиворечивая прогностическая модель реальности

Е.Е. Витяев<sup>1,2</sup>

<sup>1</sup>Институт математики им. С. Л. Соболева СО РАН

<sup>2</sup>Новосибирский Государственный Университет

В работах П.К. Анохина показано, что принцип опережающего отражения действительности является основой всего живого. Опережающее отражение действительности основано на причинности внешнего мира. В данной работе показывается, что отражение мозгом причинности внешнего мира может быть организовано в логически непротиворечивую прогностическую модель реальности, которая непрерывно во времени и в пространстве проверяет себя на адекватность этой реальности. Эта модель и есть сознание. При этом сознание отражает не все причинные связи, а только те, которые связаны с достижением организмом определенных целей (удовлетворением потребностей). В соответствии с информационной теорией эмоций П.В. Симонова [12-13], субъективные эмоциональные ощущения возникают только тогда, когда есть вероятностный прогноз достижения цели(ей) и не удовлетворенная потребность, соответствующая этим целям. Поэтому сознание всегда окрашено некоторыми субъективными ощущениями (qualia), которые отражают достигаемые организмом цели вместе с оценки вероятности достижения этих целей. При этом автоматизмы, не содержащие вероятностные прогнозы достижения целей (промежуточных целей), уходят из сознания.

Хороший пример целевой направленности сознания приводит У. Найсером: "Мы записали на видеомagnetофон две "игры" (например, футбол и хоккей – Е.В.), а затем с помощью зеркала осуществили полное визуальное наложение двух передач – как если бы на телевизионном экране одновременно демонстрировались два канала ... Испытуемых просили наблюдать за одной игрой и игнорировать другую, нажимая на ключ при каждом целевом событии (например, при каждом ударе по мячу, шайбе – Е.В.) в наблюдаемой игре. ... При темпе 40 целевых событий в минуту было одинаково легко следить за игрой независимо от того, демонстрировалась она вместе с другой или отдельно. Количество ошибок составляло примерно 3% ... Естественность этой задачи и отсутствие интерференции со стороны второго эпизода просто удивительны. Испытуемый *не видит* (выделение – Е.В.) irrelevantную игру ... Циклическая модель восприятия позволяет легко объяснить эти результаты" [11 с.103-105].

Предлагаемая нами в работе оригинальная формализация причинных связей и циклических причинных связей включает в себя циклическую модель восприятия [5-6]. Эта циклическая модель непрерывно во времени и в пространстве про-

веряет себя на адекватность реальности, как это описано в книге С.Д. Смирнова «Психология образа»: «Все это позволяет нарисовать следующую картину хода познавательной деятельности на уровне восприятия. Индивид всегда имеет некоторый образ или модель окружения, которая непрерывна во времени и пространстве и носит прогностический характер, т.е. в ней экстраполируются и воспроизводятся на языке чувственных модальностей ожидаемые результаты воздействия источника стимула на наши органы чувств» [15].

Сознание, как логически непротиворечивая прогностическая модель реальности, не только непрерывно во времени и пространстве прогнозирует, какие стимулы будут восприняты в следующий момент, но и непрерывно проверяет правильность этих прогнозов. Совпадение их с реально поступающими стимулами создает ощущение присутствия во внешнем мире.

В нашей формализации причинные связи обнаруживаются нейронами, формальная модель которых [28] удовлетворяет правилу Хебба [18]. Циклические причинные связи обнаруживаются клеточными ансамблями, которые, как мы покажем, обнаруживают «естественную» классификацию объектов внешнего мира [3] и обладают свойством интегрированной информации по G.Tononi [20].

Важным свойством сознания, которое проявляется не во всех структурах мозга, как это показано G.Tononi [20], является его интеграционный характер. Наиболее емкой фразой, характеризующей этот характер, является: «различия делающие различие» ("differences that make a difference") [20]. Она означает, что совокупность различий приводит к качественному отличию или иначе качественно отличающиеся объекты различны по целой совокупности различных свойств. Чтобы уловить это свойство сознания G.Tononi вводит понятие интегрированной информации, когда информация от системы причинных связей свойств объекта превосходит информацию от совокупности свойств самих по себе. Тогда объекты различаются не по отдельным свойствам, а по совокупности (паттерну) различающихся свойств. В нашей формализации циклические причинные связи автоматически формируют паттерны различающихся свойств, которые причинно взаимосвязаны. В нашем эксперименте на закодированных цифрах [5-6] рассматриваются все возможные причинные связи между различными признаками цифр, которые верны для всех цифр, но при этом выделение и идентификация отдельных цифр происходит

только по паттернам свойств цифр, циклически взаимно предсказывающих друг-друга причинными связями. Более того, сами паттерны свойств в нашей формализации автоматически формируются причинными связями, как циклически взаимно предсказывающийся набор свойств, формирующий неподвижную точку. Не все области мозга способны формировать богатые циклические причинные связи для идентификации объектов внешнего мира, а также для непрерывной во времени и пространстве проверки правильности сделанных предсказаний, что создает ощущение реального существования этих объектов во внешнем мире.

Свойство сознания по восприятию объектов паттернами свойств, в которых «различия делают различие», на самом деле опирается на высокую коррелированность свойств «естественных» объектов внешнего мира. Это свойство подтверждено естествоиспытателями, которые строили «естественные» классификации, а также исследованиями по формированию «естественных» понятий в когнитивных науках. Для формализации «естественных» понятий Bob Rehder [22-23] выдвинул теорию причинных моделей, в которой отношение объекта к категории основывается уже не на множестве признаков и близости по признакам, а на основании сходства порождающего причинного механизма: «объект классифицируется как член некоторой категории в той степени, в которой его свойства, вероятно, были сгенерированы причинными законами данной категории» [22-23].

Сознание, как и «естественная» классификация объектов внешнего мира иерархична. Как говорит Дж. Гибсон, внешний мир имеет свойство «встроенности»: «... ущелья встроены в горы, деревья встроены в ущелья, листья встроены в деревья, клетки встроены в листья. При любом масштабе можно обнаружить, что одни формы содержат в себе другие» [8]. На каждом уровне детальности формируются свои «естественные» классы и образы по одним и тем же законам. Наиболее общим образом является «образ мира», который формируется, начиная с первых дней жизни. Субъект никогда не воспринимает отдельные объекты, а всегда воспринимается некоторая целостная картина. Отдельные объекты выделяются в «видимом поле» [8,16] в процессе деятельности с этими объектами как с предметами. Предметная деятельность преобразует «видимое поле» в «видимый мир» и «образ мира» (Леонтьев А.Н. [10]). Это не тривиальный процесс и требует многослойной нейронной обработки, как показано в экспериментах с Deep Learning. Наша формализация в виде циклических причинных связей может осуществлять такую же иерархическую обработку видимого поля в виде иерархии «естественных» классов в соответствии со встроенностью объектов реальности и, как показано в наших экспериментах [5], делает это принципиально точнее, чем Deep Learning.

Восприятие отдельного объекта начинается не с объекта, а с «образа мира» (Смирнов С.Д. [15]): «Образ мира не складывается из образов отдель-

ных явлений и предметов, а с самого начала развивается и функционирует как некоторое целое. Это значит, что любой образ есть не что иное, как элемент образа мира, и сущность его не в нем самом, а в том месте, в той функции, которую он выполняет в целостном отражении реальности. Эта характеристика образа мира определяется взаимосвязями и взаимозависимостями между элементами самой объективной реальности. ...". С информационной точки зрения, чем выше иерархия некоторого «естественного» класса, тем он устойчивей и инвариантен по отношению к всевозможным вариациям соответствующего образа. В нашей формализации неподвижная точка взаимопредсказаний свойств более высоких по иерархии классов описывается причинными связями с более высокими оценками вероятности. Поскольку предсказаний в нашей формализации всегда осуществляются по более высоким закономерностям, то более высокие по иерархии классы начинают доминировать над нижележащими классами и играть ведущую роль.

У. Найсер отмечал, что восприятие как процесс не завершено, пока перцептивный цикл не завершится: «Тахистоскопические эксперименты попросту не относятся к нормальным перцептивным навыкам, и термин «восприятие» в полном смысле слова нельзя отнести к тому, что в них происходит» [11]. Эксперименты с маскировкой стимулов также показывают, что пока причинные связи не зациклились и не привели к устойчивому циклическому возбуждению (неподвижной точке) предвосхищений и проверке совпадения их с реальными стимулами, мы не можем осознать эту реальность.

Одна из главных ролей сознания – ликвидация противоречий в осознании поступающих стимулов. С информационной точки зрения это старая и до сих пор не решенная проблема философии науки – проблема статистической двусмысленности: при обнаружении причинных связей на данных, как правило, обнаруживается совокупность правил, приводящих к противоречиям. Для разрешения проблема статистической двусмысленности К.Гемпель предложил использовать максимально специфические правила. Нами предложена формализация максимально специфических правил, для которых не возникает противоречий [27]. Предложена также формальная модель нейрона, удовлетворяющая правилу Хебба и обнаруживающая максимально специфические причинные (условные) связи как замыкания условных связей на уровне нейрона. В нашей формализации причинные связи предсказывают не только наличие свойства (стимула), но и его отсутствие. Таким образом, моделируется не только возбуждение, но и торможение нейронной сети.

Для формализации неподвижных точек предсказания также необходимо было доказать, что в них не возникает противоречий (не происходит одновременное возбуждение и торможение нейрона). Это осуществлено в работах, где неподвижные точки предсказаний формализованы в виде

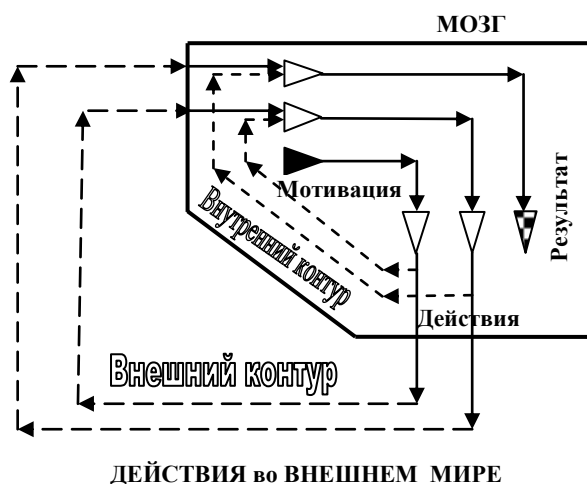


Рис. 1. Формирование акцептора результатов

вероятностных формальных понятий [7,31-32]. Такие неподвижные точки взаимно предсказывают не только наличие стимулов, но и их отсутствие, которое осуществляет вытормаживание не совместимых с данными предсказаниями стимулов. В частности, это моделирует восприятие иллюзий типа фигура-фон, когда есть два не совместимых образа в виде двух различных неподвижных точек. Осознание поступающей стимуляции требует разрешение противоречия между возникающими неподвижными точками, входящими в стимуляцию фигура-фон. В разрешении противоречий главную роль играют наиболее сильные неподвижные точки наиболее инвариантных классов. В любом случае, пока не сформируется непротиворечивое согласование всех неподвижных точек воспринимаемых в данный момент образов, включая «образ мира», формируя целостную картину воспринимаемой реальности, осознание воспринятого не возникает.

Прогнозирование стимулов и сличение их с реальностью осуществляется не только клеточными ансамблями Хебба, но и через деятельность. Как показано в теории функциональных систем, предвосхищение результатов деятельности реализуется специальными коллатеральными ответвлениями от произведенных действий, которые поступают на «вход» мозга, конвергируя с афферентацией от входных стимулов [17]. Фактически это означает выработку условных (причинных) связей между осуществлением действий (эффекторным возбуждением) и последующим восприятием результатов действий, представленных их афферентными стимулами (см. рис. 1). Поэтому предвосхищение некоторого стимула в результате некоторого (перцептивного) действия по внутреннему контуру мозга сопоставляется с реально поступающими стимулами, полученными в результате осуществления этого действия по внешнему контуру. Такое предвосхищение включает процесс прогнозирования достижения целей. Этот процесс, в соответствии с принципом опережающего отражения действительности, детально описан в теории функциональных систем и ее формализациях [1,17,4,29-30]. В нашей формализации

зации опережающее отражение действительности в полном соответствии с теорией функциональных систем [30] осуществляется причинными связями, представленными на рис. 1 и формальными нейронами [28], предсказания по которым непротиворечивы. Осознание всех причинных связей, представленных на рис. 1 вплоть до предвосхищения достижения результата в виде «апетита» осуществляется по «внутреннему контуру» работы мозга (рис. 1) и включается в сознание.

**Заключение.** Суммируя вышесказанное, можно выделить следующие основные функции сознания:

1. Обеспечение логически непротиворечивого и прогностического представления реальности. Это становится возможным за счет хорошей структурированности самого внешнего мира – его причинности, «естественной» классификации и «встроенности». Мозг улавливает эту структурированность клеточными ансамблями и внутренним контуром работы мозга в процессе деятельности, которые в нашей формализации представлены формальной работой нейрона, обнаруживающей причинные связи, а также неподвижными точками предсказаний, образующими логически непротиворечивые модели реальности, включающие как стимулы, которые возможны в данной модели, так и стимулы, которые в данной модели не возможны. Это создает конкуренцию между моделями реальности, учитывая предметность воспринимаемого образа, заставляя сознание искать наиболее непротиворечивую модель («образ мира») реальности («видимом поле»).

2. В создаваемый в данный момент «образ мира», сознание включает, прежде всего, потребности и вероятности их удовлетворения, что вызывает соответствующие эмоции и субъективные состояния (qualia). В вероятностный прогноз не входят автоматизмы действий, поэтому они уходят из сознания.

3. Другой (по отношению к непротиворечивости) основной функцией сознания является постоянная и непрерывная во времени и пространстве проверка совпадения, создаваемой модели реальности («образа мира») и самой реальности. Это осуществляется постоянной проверкой во времени и пространстве совпадения прогнозов, осуществляемых моделью реальности с самой реальностью, что создает ощущение внешнего мира. Поэтому ощущения концентрируются в местах прогноза и контакта прогнозируемых стимулов с реальными. В нашей формализации модель нейрона и неподвижные точки (включая неподвижные точки, основанные на деятельности) осуществляют прогноз. Неподвижная точка замыкает прогнозы внутри себя только в том случае, если все они подтвердились, если нет, то в неподвижной точке возникает противоречие, которое вызывает ориентировочно-исследовательскую реакцию и дальнейшую работу сознания по ликвидации противоречия. Такая работа сознания не может вмешиваться в работу неподвижных точек, поскольку неподвижные точки являются точной моделью

реальности, но может изменить поступающую стимуляцию, варьируя условия восприятия.

### Литература

1. Анохин К.В., Бурцев М.С., Зарайская И.Ю., Лукашев А.О., Редько В.Г. Проект «Мозг анимата»: разработка мо-дели адаптивного поведения на основе теории функциональных систем // Восьмая национальная конференция по искусственному интеллекту с международным участием. Труды конференции. М.: Физматлит, 2002. Т.2. С.781-789.
2. Бернштейн Н.А. Биомеханика и физиология движений. // Избранные психологические труды, Москва-Воронеж, 1997, с.605.
3. Витяев Е.Е., Мартынович В.В. Формализация "естественной" классификации и систематики через неподвижные точки предсказаний // СИБИРСКИЕ ЭЛЕКТРОННЫЕ МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ (Siberian Electronic Mathematical Reports), Том 12, Институтом математики им. С. Л. Соболева СО РАН, 2015, стр. 1006-1031.
4. Витяев Е.Е., Принципы работы мозга, содержащиеся в теории функциональных систем П.К. Анохина и теории эмоций П.В. Симонова // Нейроинформатика, 2008, том 3, № 1, стр. 25-78
5. Витяев Е.Е., Неупокоев Н.В. Математическая модель восприятия и образа. Информационные технологии в гуманитарных исследованиях, Вып.17, ИАЭТ СО РАН, Новосибирск, 2012, 63-72.
6. Витяев Е.Е., Неупокоев Н.В. Формальная модель восприятия и образа как неподвижной точки предвосхищений // Подходы к моделированию мышления. УРСС Эдиториал, Москва, 2014, стр. 155-172.
7. Витяев Е.Е., Демин А.В., Пономарёв Д.К. Вероятностное обобщение формальных понятий // Программирование, Т.38, №5, 2012, С. 219-230.
8. Гибсон Дж. Экологический подход к зрительному восприятию. М.: Прогресс, 1988. С. 462.
9. Демин А.В., Витяев Е.Е. Логическая модель адаптивной системы управления. Нейроинформатика, 2008, том 3, № 1, стр. 79-107
10. Леонтьев А.Н. Образ мира // Избранные психологические произведения. – М.: Педагогика, 1983. – С. 251-261.
11. Найсер У. Познание и реальность. “Прогресс”, М. 1981, с. 229.
12. Симонов П.В. Эмоциональный мозг. М.: Наука, 1981. с. 140.
13. Симонов П.В. Высшая нервная деятельность человека (мотивационно-эмоциональные аспекты). М.: Наука, 1975. с. 173.
14. Смирнов Е.С. Конструкция вида таксономической точки зрения // Зоол. Журн. Т. 17, №3, 1938, С. 387-418.
15. Смирнов С.Д. Психология образа: проблема активности психического отражения. МГУ, М., 1985, с.232.
16. Столин В.В. Исследование порождения зрительного пространственного образа. — В кн.: Восприятие и деятельность. М., 1976.
17. Судаков К.В. Общая Теория Функциональных Систем М.: Медицина, 1984. с. 222.
18. Hebb D.O. The organization of behavior. A neurophysiological theory. NY, 1949. 335 p.
19. Hempel, C. G. ‘Maximal Specificity and Lawlikeness in Probabilistic Explanation’, Philosophy of Science 35, 1968. – P. 16–33.
20. Masafumi Oizumi, Larissa Albantakis, Giulio Tononi. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0 // PLOS Computational Biology, May 2014, V.10. Issue 5.
21. Mill, J.S. System of Logic, Ratiocinative and Inductive. L., 1843.
22. Bob Rehder. Categorization as causal reasoning // Cognitive Science 27 (2003) 709–748.
23. Bob Rehder, Jay B. Martin. Towards A Generative Model of Causal Cycles // 33rd Annual Meeting of the Cognitive Science Society 2011, (CogSci 2011), Boston, Massachusetts, USA, 20-23 July 2011, V.1 pp. 2944-2949.
24. Rosch, E., Mervis, C.B. Family resemblances. Studies in the internal structure of categories // Cognitive Psychology, 7, 1975, P. 573–605.
25. Rosch, E., Principles of Categorization // Rosch, E. & Lloyd, B.B. (eds), Cognition and Categorization, Lawrence Erlbaum Associates, Publishers, (Hillsdale), 1978. P. 27–48
26. B. H. Ross, E. G. Taylor, E. L. Middleton, and T. J. Nokes. Concept and Category Learning in Humans // H. L. Roediger, III (Ed.), Cognitive Psychology of Memory. Vol. [2] of Learning and Memory: A Comprehensive Reference, 4 vols. (J.Byrne Editor), Oxford: Elsevier, 2008, P. 535-556.
27. Evgenii Vityaev. The logic of prediction // Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference (August 16-19, 2005, Novosibirsk, Russia), edited by S.S. Goncharov, R. Downey, H. Ono, World Scientific, Singapore, 2006, P. 263-276.
28. Vityaev E.E. A formal model of neuron that provides consistent predictions // Biologically Inspired Cognitive Architectures 2012. Proceedings of the Third Annual Meeting of the BICA Society (A. Chella, R.Pirrone, R. Sorbello, K.R. Johansdottir, Eds). In Advances in Intelligent Systems and Computing, v.196, Springer: Heidelberg, New York, Dor-drecht, London. 2013, P. 339-344.
29. Evgenii Vityaev. Unified formalization of "natural" classification, "natural" concepts, and consciousness as integrated information by Giulio Tononi // The Sixth international conference on Biologically Inspired Cognitive Architectures (BICA 2015, November 6-8, Lyon, France), Procedia Computer Science, v.71, Elsevier, 2015. pp 169-177.
30. Evgenii E. Vityaev Purposefulness as a Principle of Brain Activity // Anticipation: Learning from the Past, (ed.) M. Nadin. Cognitive Systems Monographs, V.25, Chapter No.: 13. Springer, 2015, pp. 231-254.
31. E.E. Vityaev, A.V. Demin, D. K. Ponomaryov. Probabilistic Generalization of Formal Concepts // Programming and Computer Software, 2012, Vol. 38, No. 5. P. 219–230.
32. E.E. Vityaev, V.V. Martinovich. Probabilistic Formal Concepts with Negation // A. Voronkov, I. Virbitskaite (Eds.): PCI 2014, LNCS 8974, P. 385-399.

