

Альтернативная формализация сознания как интегрированной информации по G.Tononi's

Е.Е. Витяев

vityaev@math.nsc.ru

Институт математики им. С. Л. Соболева СО РАН (Новосибирск, Россия)

Новосибирский Государственный Университет (Новосибирск, Россия)

Цель работы – дать более точную формализацию сознания как теории интегрированной информации по G.Tononi (2010), основываясь на исследованиях в области «естественных» классификаций (Забродин, 1981).

Tononi определяет сознание как первичное понятие, которое обладает следующими свойствами: composition, information, integration, exclusion (Tononi, 2012). Эти свойства определяются через феноменологические аксиомы:

- composition – сознание структурно: каждый опыт многогранен (он включает элементы комбинируемые различными способами);
- information – сознание дифференцировано: каждый опыт высоко информативен и отличается особым образом от многих других опытов;
- integration – сознание целостно: каждый опыт есть интегрированная единица, которая не может быть редуцирована к взаимно независимым компонентам;
- exclusion – сознание эксклюзивно: каждый опыт конкретен и имеет определенные пространственно-временные границы;
- experience (опыт) – максимально интегрированная концептуальная структура;
- complex – максимальная, не редуцируемая на составные части, единица.

Чтобы формально определить эти понятия Tononi вводит понятие интегрированной информации и сознания как интегрированной информации: «Интегрированная информация количественно определяет редукцию неопределенности, полученную системой, приходящей в определенное состояние за счет **причинных** взаимодействий между своими частям» (Tononi, 2010). Формально интегрированная информация определяется как энтропия системы по отношению к суммарной энтропии её частей. Единицами интегрированной информации являются комплексы (complex). Совокупность всей интегрированной информации дает опыт (experience, quale) – максимально интегрированную концептуальную структуру, включающую множество концептов.

Из этих определений не совсем ясно, почему интегрированная информация определяет сознание. Далее будет дана другая формализация интегрированной информации, которая прояснит, в каком смысле она связана с сознанием.

Сформулируем закон природы, обнаруженный в области «естественных» классификаций (таких как таблица Менделеева или система видов Дарвина). Для «естественных» классов обнаружена большая избыточность описывающих их признаков, проявляющаяся в свойстве «таксономической насыщенности» (Кожара В.Л. 1989): для объектов, описываемых N существенными признаками, всегда можно подобрать такое $n \ll N$, что классификация, построенная по n случайно выбранным из N признакам, практически одна и та же. Это означает, что совокупность произвольно выбранных n признаков практически однозначно определяют остальные $N-n$ признаков. Легко подсчитать, что в этом случае существует степенное число причинных связей, устанавливающих взаимосвязь между N признаками. Этот эффект объясняется тем, что «естественные» классификации являются генетическими – естественные объекты всегда имеет некоторый генезис своего происхождения. Если рассмотреть генезис с точки зрения принципа физического детерминизма (знание состояния некоторой физической системы позволяет, в принципе, предсказать все последующие её состояния), то оказывается, что комплекс параметров начального состояния объекта в процессе его генезиса определяет весь комплекс параметров результирующего естественного объекта. Отсюда следует, что **причинность комплексна** и N существенных признаков, получающихся в процессе генезиса объекта, связаны между собой степенным числом причинных связей.

В теории G.Tononi интегрированная информация определяется как внутреннее свойство системы. Соответственно сознание тоже является внутренней характеристикой системы. Рассмотрим интегрированную информацию и сознание, не как внутренние свойства системы, а как способность системы *отражать* комплексы причинных связей внешних естественных объектов. При этом причинные связи должны образовываться в результате обучения (индуктивного вывода). Обнаружение причинных связей – не простая проблема. Индуктивный вывод причинных связей, во-первых, сталкивается с проблемой статистической двусмысленности, во-вторых, вывод предсказаний по обнаруженным причинным связям сталкивается с проблемой падения вероятностей предсказания в процессе вывода (см. Витяев, 2006). Нами эти проблемы решены введением специального семантического вероятностного вывода (Vityaev E.E. 2006), причем так, что получается формальная модель нейрона, удовлетворяющая правилу Хебба и обнаруживающая причинные связи (Vityaev E.E. 2013). Таким образом, смысл деятельности нейрона в процессе отражения – обнаружение причинных связей.

Феноменологические аксиомы G.Tononi не говорят о том, как происходит процесс отражения, они говорят о том, что, если информация будет интегрироваться в системе определенным образом, то она будет проявлять сознание. Эту, в целом, правильную интуицию G.Tononi, мы преобразуем в процесс отражения на нейронном уровне.

Нами показано и подтверждено компьютерным экспериментом, что комплексы могут быть обнаружены в процессе отражения как неподвижные точки предсказаний по причинным связям (Витяев и Неупокоев 2012). На нейронном уровне это означает, что когда нейроны клеточных ансамблей начинают синхронно возбуждать друг друга взаимно предсказывающими (возбуждающими) причинными связями, то тем самым воспринимается некоторый естественный объект.

То, что G.Tononi называет опытом (quale) в наших определениях будет «естественной» классификацией воспринимаемой действительности, удовлетворяющей всем требованиям (Забродин 1981) «естественных классификаций». Тогда composition – композиция воспринимаемых «естественных классов»; information – совокупность всех обнаруживаемых причинных связей; integration – неподвижные точки предсказаний по причинным связям, обнаруживающие образы действительности как «естественные» классы; exclusion – вытормаживание альтернативных образов причинными связями с отрицанием; сознание – «образ мира», создаваемый «естественной» классификаций внешнего мира, который активен – он непрерывно во времени по всей совокупности причинных связей и получающихся из них неподвижных точек предсказания («естественных» образов) предвосхищает свойства внешнего мира и проверяет правильность сделанных предсказаний.

Эта работа поддержана Российским Фондом Фундаментальных Исследований № 11-07-00560-а, интеграционными проектами СО РАН № 3, 87, 136.

Tononi G. 2010. Information integration: its relevance to brain function and consciousness. Archives Italiennes de Biologie 148, 299-322.

Забродин В.Ю. 1981. О критериях естественной классификации. НТИ, сер.2, №8.

Tononi. G. 2012. Integrated information theory of consciousness: an updated account. Archives Italiennes de Biologie 150, 56-90.

Кожара В.Л. 1989. Анализ информативно насыщенных таксономических структур как способ выявления географических закономерностей. Дисс. Канд. Геогр. Н., Москва.

Витяев Е.Е. 2006. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. Новосибирский гос. ун-т. Новосибирск.

Vityaev E.E. 2006. The logic of prediction // Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference, eds. S.S. Goncharov, R. Downey, H. Ono, World Scientific, Singapore, 263-276.

Vityaev E.E. 2013. A formal model of neuron that provides consistent predictions // Biologically Inspired Cognitive Architectures 2012. Proceedings of the Third Annual Meeting of the BICA Society (Eds. A. Chella, et al.), Advances in Intelligent Systems and Computing, v.196, Springer, 339-344.

Витяев Е.Е., Неупокоев Н.В. 2012. Формальная модель восприятия и образа как неподвижной точки предвосхищений. Нейроинформатика 6, № 1, 28-41.