

ДИСКУССИИ

И.Ф. Михайлов

СОЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ И ПРОИСХОЖДЕНИЕ МОРАЛЬНЫХ НОРМ

Михайлов Игорь Феликсович – кандидат философских наук, старший научный сотрудник. Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: ifmikhailov@gmail.com

В предлагаемой авторской концепции общество и человек понимаются как вложенная вычислительная система мультиагентной архитектуры. «Вычислительность» как понятие расширяется за пределы машины Тьюринга и рассматривается как способность рассчитывать и предсказывать статистические распределения, в максимальной степени соответствующие характеристикам внешней среды, чтобы достичь оптимального значения предполагаемой контрольной переменной. Общество индуцирует в агентах правила их поведения подобно тому, как магнитное поле индуцирует ток в движущемся проводнике. Индукция осуществляется путем отбора эмпирически найденных индивидами паттернов поведения и сертификации тех из них, которые наиболее полезны для достижения целевых показателей системы. Сертифицированные правила воспринимаются индивидами как, с одной стороны, имманентные им, а с другой – как в принципе допускающие их нарушение, поскольку как отдельные агенты, так и общество в целом суть статистические машины. Объективная вероятность неисполнения требования правила воспринимается как моральная свобода, а гормональное подкрепление следования общему правилу – как ощущение морального долга. Вычислительный подход к морали преодолевает индуктивизм натуралистических и утилитаристских теорий и в то же время научно объясняет «трансцендентальность» нравственного закона, не прибегая к идеалистической метафизике, как это делает деонтизм кантовского толка.

Ключевые слова: натурализм, утилитаризм, деонтизм, мораль, вычисление, мозг, нейросеть

Для цитирования: Михайлов И.Ф. Социальные вычисления и происхождение моральных норм // Философский журнал / Philosophy Journal. 2022. Т. 15. № 1. С. 51–68.

Скрывая от нас все бессознательные заключения, к которым он приходит, наш мозг создает у нас иллюзию непосредственного контакта с материальным миром. В то же самое время он создает у нас иллюзию, что наш внутренний мир обособлен и принадлежит только нам. Эти две иллюзии дают нам ощущение, что в мире, в котором мы живем, мы действуем как независимые деятели. Вместе с тем мы можем делиться опытом восприятия окружающего мира с другими людьми. За многие тысячелетия эта способность делиться опытом создала человеческую культуру, которая, в свою очередь, может влиять на работу нашего мозга.

Крис Фрит¹

Вычислительный подход к морали

Мораль может быть понята как совокупность социальных ограничений, необходимость которых не следует непосредственно из природных факторов. Например, запреты на близкородственные связи или на употребление алкоголя в некоторых обществах не являются в строгом смысле слова моральными, потому что их легко можно объяснить природными факторами. А, например, максима «не обижай слабых» значительно хуже поддается натуралистическому объяснению, поскольку с точки зрения физической эффективности наиболее энергетически эффективный путь к обладанию неким ресурсом состоит в том, чтобы забрать этот ресурс у более слабой особи, если он имеется у нее в готовом виде, – вместо того, чтобы пытаться произвести его собственным трудом или добыть где-то в природе. Эта и подобные максимы составляют собственно моральную сферу – столь трудный для научного объяснения «нравственный закон внутри нас».

Среди многочисленных философских попыток справиться с этой проблемой можно выделить три основных подхода: натурализм, утилитаризм (который почитается некоторыми за подвид консеквенциализма) и деонтизм. Натурализм пытается объяснить всё естественными причинами, утилитаризм вводит некую телеологию, опираясь на последствия поступков, а деонтизм основывается на представлении о трансцендентальном моральном долге. При этом первые два подхода сталкиваются с невозможностью объяснить всеобщность моральной регуляции: у разных особей могут быть разные причины поведения и разные представления о его последствиях. И тогда вопрос – каким образом получается так, что всё сообщество руководствуется более или менее одной моральной системой, – повисает в воздухе. Деонтизм как бы разрешает эту проблему, но за счет опоры на догматическую – и, как правило, идеалистическую – метафизику.

В русскоязычной философской литературе также существует солидная традиция философии морали вообще и теории происхождения моральных норм в частности. По понятным историческим причинам, эта проблематика

¹ Фрит К. Мозг и душа: как нервная деятельность формирует наш внутренний мир. М., 2010. С. 37–38.

рассматривается у нас в основном в культурно-историческом аспекте, с акцентом на реальное происхождение моральных норм скорее, чем на априорный концептуальный анализ нормативности как таковой. Так, Р.Г. Апресян считает, что:

В теоретической абстракции номогенез может быть представлен как акт рационального решения, результат нормотворчества. Однако насколько можно судить по доступным историческим данным, в реальности формирование стандартов происходило стихийно, на основе осмысления опыта эффективного поведения и взаимодействия, практического и рефлексивного обобщения этого опыта, его объективации, универсализации и постепенного закрепления в культуре².

В приведенном фрагменте можно выделить две мысли: (1) априорные теории морали могут обладать концептуальной ясностью, но не соответствовать реальному историческому процессу; и (2) в реальности отбор и закрепление моральных норм происходит на основе осмысления опыта *эффективного* поведения. Как я постараюсь показать далее, с этой идеей вполне можно работать при условии возможно более строгого понимания эффективности. В общем случае более эффективным является нечто – процесс, устройство, организация, сотрудник и т.д., – если при меньших затратах оно обеспечивает больший возврат. Возникает естественный вопрос: затратах чего? На мой взгляд, всякая теория выигрывает в том числе от универсальности своей методологии. Поскольку подход, который я далее собираюсь предложить, может быть применен в принципе к любым неравновесным системам, а не только к социальным, под затратами, по всей видимости, стоит понимать энергетические затраты как наиболее универсальный аспект естественных взаимодействий. Но этот критерий на данном этапе принимается скорее как гипотетический – я не исключаю, что будущая теория окажется более *эффективной*, основываясь на иных величинах.

Предлагаемый мною вычислительный подход к морали, на мой взгляд, свободен от недостатков вышеуказанных подходов, что я и постараюсь показать в дальнейшем изложении. В рамках этого видения *мораль* понимается как *система правил и алгоритмов, которую общество как обучающаяся вычислительная система вырабатывает ради сохранения гомеостаза и/или энергетического оптимума*. Вычислительная цель этой системы намечена в данном случае условно и приблизительно и в ходе уточнения теории может быть заменена на другую. Более точное ее определение – предмет эмпирических исследований. Например, в рамках популярной ныне теории предиктивного процессинга предполагается, что все сложные неравновесные системы – биологические, когнитивные или социальные – преследуют цель минимизации «свободной энергии», т.е. разницы между данными, предсказанными внутренней генеративной моделью системы, и реальными входящими данными. Но в любом случае существует некая контрольная переменная, на достижение оптимального значения которой направлена вся деятельность системы. Очевидно, что вычислительный подход к морали является холистическим: поведение частей объясняется задачами целого.

² Апресян Р.Г. Нормативная регуляция поведения // Человек. 2018. № 1. С. 5.

Дискуссии вокруг вычислений

В исходном определении общество предстает как обучающаяся вычислительная система. Но что значит «вычислительная»? Вычислительная система осуществляет вычислительные процессы, а последние характеризуются алгоритмичностью, субстратонезависимостью и контрфактуальностью. Алгоритмичность предполагает, что весь процесс можно представить как последовательное выполнение правил вида «если..., то...». Контрфактуальность тесно связана с этим свойством и означает, что процесс способен воспринять и обработать любые альтернативные данные из некоего допустимого множества. Далее я постараюсь показать, что контрфактуальность лежит в основе того, что мы привычно называем субъективностью, и именно социальность делает ее возможной в том конкретном воплощении, с которым мы сталкиваемся в человеческом мире. Субстратонезависимость – это синоним множественной реализации, возможности осуществить один и тот же по целям и содержанию процесс в различных материальных средах, обладающих необходимыми структурными характеристиками.

Однако в этом пункте часто приходится сталкиваться с возражениями такого рода, что употребление термина «вычислительный» является слишком сильной антропоморфизацией, поскольку в природных процессах только при очень развитом воображении можно увидеть целенаправленное получение некоторых численных значений. Поэтому одни авторы иронически спрашивают, какими именно разделами математики пользуется природа, осуществляя натуральные вычисления³, а другие, признавая недостаточность традиционного физического дискурса для описания сложных процессов в природе, вполне серьезно предлагают избегать термина «вычислительный» и использовать вместо него термин «алгоритмический»⁴. В основе этих разночтений лежат два обстоятельства: специфика русского языка и линейное понимание алгоритмичности.

Многие авторы (если не большинство) настойчиво сводят вычисление к манипуляциям с символами – цифровыми или лексическими⁵. Однако такой подход, по сути, аналогичен подмене анализа, например, музыки как таковой анализом языка ее нотной записи – подмене, которую, на мой взгляд, совершил Нельсон Гудмен в своей знаменитой работе⁶. То есть одна из возможных репрезентаций сущности выдается за саму сущность. Число есть символическая репрезентация количества, но количество – это реальное свойство материальных объектов. Русское слово «вычисление» заставляет нас думать о символических операциях с числами, тогда как латынь и испытывавший ее влияние английский предлагают как минимум два частично синонимичных термина: *calculatio* (*calculation*) и *computatio* (*computation*). В нестрогих разговорных контекстах они могут быть взаимозаменяемы. Однако в целях формирования строгого научного лексикона стоит приглядеться к их этимологии. Этимология в общем и целом – не лучший

³ Бажанов В.А. Вычисляющая природа – реальность или метафора? // *Философия науки и техники*. 2021. Т. 26. № 2. С. 38–43.

⁴ Шалак В.И. Алгоритмические явления в природе: модель объяснения // *Вопросы философии*. 2020. № 11. С. 120–124.

⁵ *Mitkowski M.* From Computer Metaphor to Computational Modeling: The Evolution of Computationalism // *Minds and Machines*. 2018. Vol. 28. P. 515–541.

⁶ *Goodman N.* *Languages of Art*. Indianapolis, 1976.

советчик в деле определения термина, но в данном случае она ведет по правильному пути.

Calculatio имеет корнем латинское *calculi*, что буквально обозначает морскую гальку – главное средство манипуляций с числами у пифагорейцев («кальций» – однокоренное слово). Видимо, с этим связано достаточно узкое значение слова «калькулятор» в русском языке. Напротив, латинское *putare* означает «думать» или «полагать». Приставка *com-* означает «вместе». Поэтому *computatio* может обозначать не только арифметические, но также логические и вообще любые правилосообразные операции. И, закономерным образом, слово «компьютер» в русском языке обозначает устройство, значительно более универсальное, чем калькулятор. Однако, когда мы произносим «вычисление», большая часть слушателей связывает обозначаемые этим словом операции с тем, что делает калькулятор – именно потому, что в корне слова присутствует «число». К сожалению, в русском языке нет адекватного перевода латинского *computatio*. Поэтому остается только договориться, что в контексте данной статьи термин вычисление будет употребляться именно в этом – более универсальном – значении.

Другое возражение против использования термина «вычисление» в контексте природных или социальных процессов связано со специфическим пониманием их алгоритмичности. Так, В.И. Шалак считает, что «И. Михайлов увидел смысл социальности в информационном обмене между людьми. Именно поэтому вместо общего понятия алгоритма он обратился к более узким понятиям *вычисления* и *репрезентации*. С нашей точки зрения, это методологическая ошибка»⁷. Он ссылается на мою статью⁸, где я действительно утверждал, что класс вычислительных процессов более узок, чем класс процессов алгоритмических. Алгоритмический процесс – это процесс, который можно представить как следующий правилам из некоторого множества. Вычислительный процесс – это алгоритмический процесс, в котором объектом управления являются репрезентации. И далее у меня имеется уточнение, которое Владимир Иванович, к сожалению, проигнорировал: «Репрезентация есть отображение (mapping) формальных свойств одного процесса в формальные свойства другого», на основании чего, согласно моему видению, «...сами репрезентации “отвязываются” от какого бы то ни было психологизма и получают чисто формальное определение»⁹. В.И. Шалак полагает, что именно информационный подход, который он мне приписывает, заставляет меня прибегать к «вычислительной» терминологии, поскольку, по его мнению, «символьные преобразования составляют лишь узкий класс алгоритмических явлений» и «в природе большинство таких явлений характеризуется переходами не между символьными репрезентациями, а между *реальными предметами и явлениями*»¹⁰. Это, как и предшествующие рассуждения в его статье о вычислениях как операциях над «символьными репрезентациями» чисел, показывает, что он понимает отношение репрезентации как семантическое отношение между символом и его значением, т.е. вполне в духе когнитивного классицизма. Адепты последнего рассматривают

⁷ Шалак В.И. Алгоритмическая модель социальных процессов // Философские проблемы информационных технологий и киберпространства. 2021. № 1. С. 51.

⁸ Михайлов И.Ф. Вычислительный подход в социальном познании // Философия науки и техники. 2021. Т. 26. № 1. С. 23–37.

⁹ Там же. С. 30.

¹⁰ Шалак В.И. Алгоритмическая модель социальных процессов. С. 50.

когнитивные акты как операции над «внутренними» символами в соответствии с некоторыми синтаксическими правилами¹¹. Эта позиция заставляла и заставляет ее сторонников искать «транспорты» (*vehicles*) символических репрезентаций в мозге, что довольно проблематично в связи с имеющимся эмпирическим материалом. Но важнее то, что эта семантическая интерпретация ведет когнитивных теоретиков прямо к «парадоксу гомункула», о чем подробнее скажем далее. Именно поэтому некоторые философы и когнитивные ученые уже пытаются разрабатывать концепции вычислений, которые позволили бы интерпретировать соответствующим образом некоторые естественные – в т.ч. нейробиологические – процессы.

Еще раз: я понимаю вычисления как алгоритмические процессы, оперирующие репрезентациями, а отношение репрезентации – как функциональную зависимость формальных свойств одного процесса от формальных свойств другого. В этой интерпретации нет никаких семантических отношений и синтаксических правил, никаких символов – только функциональные отношения. Почему я предпочитаю называть такие процессы вычислительными? Потому что, как правило – если не всегда, – они осуществляются с целью адаптации поведения сложных неравновесных систем к меняющимся условиям среды, репрезентированным во входных данных, которые для этого нужно получить и обработать. Система функционально взаимосвязанных процессов может действовать подобно искусственному компьютеру: преобразовывать входные данные в выходные в соответствии с некоторыми алгоритмами. При этом задействованные взаимозависимости и преобразования (*computatio*) вовсе не обязательно должны описываться какими-либо операциями с числами (*calculatio*).

Чтобы еще более наглядно объяснить суть нашего расхождения, обратимся к более или менее удачному примеру В.И. Шалака: «Реализуя алгоритм приобретения хлеба в магазине, мы ставим цель получить вполне реальную буханку хлеба, а не ее символьную репрезентацию в виде кассового чека»¹². В рамках его позиции подразумевается, что операции по покупке реального хлеба являются алгоритмическими, но не вычислительными, тогда как операции с чеками, оставшимися после покупки, могут быть вычислительными: например, при анализе расходов семейного бюджета. С моей же точки зрения, ситуация выглядит следующим образом. Если для покупки хлеба я сконструировал примитивного робота, который может реализовать простой линейный алгоритм (переместиться в магазин – протянуть продавцу деньги – назвать желаемый вид хлеба – получить покупку – получить сдачу – вернуться домой), то можно сказать, что такой робот совершает алгоритмический, но не вычислительный процесс – но и то только отвлекаясь от того обстоятельства, что хотя бы в навигационной системе такого робота магазин, хлеб, деньги и продавец должны быть как-то *репрезентированы* и им, роботом, опознаны. Но если алгоритмы моего робота окажутся достаточно сложны для того, чтобы он знал, *что* нужно делать, если магазин окажется закрыт, если хлеб не завезли, если у продавца не окажется сдачи и т.п., чтобы при любых неожиданных входных данных он вернулся домой с наилучшим в данных обстоятельствах результатом, – при этих условиях

¹¹ Pylyshyn Z.W. Computation and cognition: Toward a foundation for cognitive science. Cambridge (MA), 1986. P. xiii.

¹² Шалак В.И. Алгоритмическая модель социальных процессов. С. 50.

то, что он делает, может быть обозначено латинским глаголом 'computare' («решать, рассуждать, делать выводы» и т.п.). И тогда сам процесс его прохождения по всем этапам решения этой задачи со всей их контрафактуальностью может быть обозначен латинским 'computatio' или английским 'computation'. И у меня нет иного выбора, как перевести это понятие на русский словом «вычисление».

Определение натуральных вычислений

Исходя из сказанного, мы можем назвать существенные признаки вычислительного процесса. Это:

- (1) алгоритмичность,
- (2) функциональная взаимосвязь между (под)процессами, понимаемая как репрезентация,
- (3) контрафактуальность (способность изменять параметры процесса при изменении входных данных, но не прерывать его), и
- (4) ориентированность на результат.

Относительно (4) можно сказать следующее. Если верно сказанное выше – а именно что натуральные вычислительные процессы всегда или как правило реализуются сложными неравновесными системами для адаптации к непредсказуемым внешним или внутренним условиям, – то должен быть некий показатель (множество контрольных переменных), которые кодируют результат адаптации и на желаемые значения которых ориентируются алгоритмы всех (под)процессов системы. Так, в нашем с В.И. Шалаком примере с хлебом умный робот на всех логических развилках своего поиска будет принимать разные решения в зависимости от характера поставленной ему глобальной задачи:

- вернуться с любым хлебом как можно скорее, или
- принести определенный хлеб, чего бы это ни стоило, или
- не найдя хлеба, заменить его дрожжевым тестом, или
- не найдя хлеба, заменить его крекерами и т.п.

Например, в двух последних случаях он с большей вероятностью выберет супермаркет, где продаются эти взаимозаменяемые в рамках его задания товары, а не специализированную хлебную лавку, даже если она ближе расположена.

Одним словом, у системы должна иметься репрезентация желаемого состояния и алгоритмы подстройки составных частей и подпроцессов для достижения показателей, закодированных в этой репрезентации. Как мне представляется, эта модель достаточна для описания всего многообразия *натуральных вычислений* – т.е. вычислительных процессов в природе¹³. Мне известны по крайней мере две теории, рассматривающие свой предмет как вычислительные системы такого рода: это теория функциональных систем П.К. Анохина и теория предиктивного процессинга Карла Фристана.

Чтобы определить место социальных вычислений в общем множестве естественных вычислительных процессов, обратимся к классификации типов

¹³ К которой я отнес бы и общество, непонятно почему «качественно» отделяемое от нее некоторыми авторами.

вычислений, предложенной Брюсом Макленнаном¹⁵. Автор классифицирует их по двум показателям: тип процессинга и пространство состояний. Каждый из них может принимать одно из двух значений: дискретный или континуальный. Соответственно, у нас образуется четыре возможные комбинации. Континуальный процессинг над континуальным же пространством состояний характеризует некоторые природные процессы и некоторые аналоговые технические устройства вроде регулятора Ватта. Дискретный процессинг над дискретным пространством состояний – это классические тьюринговы вычисления. Континуальный процессинг над дискретным пространством состояний может описывать квантовые вычисления, хотя, насколько я понимаю, это пока проблемная область. И, наконец, дискретный процессинг над континуальным пространством состояний может охватывать сразу несколько областей: нейрофизиологические вычисления, в частности – процессы в мозге, нейросетевые вычисления – обработка, осуществляемая в искусственных нейросетях, и собственно социальные вычисления.

Важный аспект в понимании того, что такое вычисление, добавляет широко цитируемая концепция Дэвида Марра, специалиста в области нейрофизиологии зрения, одного из основателей вычислительной нейронауки. В книге 1982 г., переизданной в 2010 г.¹⁶, он утверждает, что любая вычислительная система может быть описана на трех уровнях:

- (1) теория вычисления, описывающая его цель, приемлемость и стратегию,
- (2) репрезентация и алгоритм, где определяются формы представления входных и выходных данных, а также алгоритм преобразования одного в другое, и
- (3) физическая реализация алгоритма и репрезентаций.

Важно, что эти уровни относительно независимы друг от друга: теория и телеология вычисления могут быть реализованы различными алгоритмами, а, в свою очередь, любой алгоритм может осуществляться в различных физических средах, обладающих достаточным числом возможных состояний.

Томмазо Поджио, друг и соавтор покойного Марра, имея за плечами тридцатилетний опыт развития когнитивных технологий, предлагает следующий апдейт марровской схемы:

- (i) эволюция,
- (ii) обучение и развитие,
- (iii) вычисления,
- (iv) алгоритмы,
- (v) биологический субстрат (*wetware*), физический субстрат (*hardware*), схемы и компоненты¹⁷.

Очевидно, что уровню (1) в схеме Марра у Поджио соответствует уровень (iii). Над ним надстраиваются уровни, необходимость выделения которых стала очевидной в результате развития эволюционного, генетического программирования, а также технологий глубокого обучения, чье философское значение, на мой взгляд, не всеми осознается и сейчас. На самом деле,

¹⁵ MacLennan B.J. Natural computation and non-Turing models of computation // Theoretical Computer Science. 2004. Vol. 317. No. 1. P. 115–145.

¹⁶ Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Cambridge (Mass.), 2010. P. 24–25.

¹⁷ Poggio T. The Levels of Understanding Framework, Revised // Perception. 2012. Vol. 41. No. 9. P. 1021.

эти технологии существенно меняют классические представления о том, что такое интеллект вообще, теснее увязывая его с закономерностями биологической эволюции и показывая иллюзорность некоторых постулированных до сих пор «качественных отличий». И, что тоже немаловажно, к «физическому» (v) основанию этой, уже пятиуровневой, схемы добавляется биологический субстрат по той же причине: XXI в. в истории искусственного интеллекта (ИИ) начался с бурного развития и, наверное, можно сказать, торжества технологий, «вдохновленных биологией» (*bio-inspired*): эволюционные и генетические алгоритмы, а также искусственные нейросети, без которых современная цивилизация уже не мыслима. А освоение этих технологий, в свою очередь, помогло увидеть множество вычислительных по своей природе явлений и процессов в живой материи. Дэвид Марр был одним из первопроходцев и, наверное, можно сказать, провидцев вычислительной революции в естественных науках. На ее примере можно сказать, что распространение этих подходов и методов на социальное познание вполне ожидаемо.

Онтология натуральных вычислений

Когнитивные процессы, по всей видимости, вырастают из биологических, являясь некоторой надстройкой над ними, и, как некоторые из них¹⁸, представляют собой распределенные статистические вычисления. Если это верно, то и мозг, и общество суть естественные распределенные вычислительные устройства, оперирующие вероятностными предсказаниями и нацеленные на достижение максимальной энергетической эффективности.

В рамках этой концепции общество может быть представлено как *метасеть* – сеть, состоящая из сетей. Нейронные сети индивидов характеризуются тем, что между их нейронами существуют жесткие физические связи. Более того, пластичные свойства этих связей лежат в основе когнитивных способностей – например, памяти. Напротив, между элементами социальной сети устойчивых физических связей нет, поэтому такая сеть более адаптивна за счет большего количества возможных состояний, создаваемых перекомбинациями ее узлов. Люди, у которых слово «компьютер» ассоциируется исключительно с устройством, стоящим у них на столе и выполняющим программы, написанные другими людьми с определенными целями, в этом месте обычно задаются вопросом: насколько «компьютерная метафора» релевантна исследованию этой сложной материи? Для того, чтобы

¹⁸ Существенным отличием вычислительных биологических процессов от невычислительных является их относительная субстратонезависимость. Так, пищеварение у определенного живого вида может происходить с участием только определенных веществ и ферментов, а, напротив, транскрипция РНК и синтез протеинов могли бы, абстрактно говоря, быть реализованы в иных материальных субстратах, структурные особенности которых позволили бы реализовать те же функциональные взаимосвязи. Хрестоматийный пример: у одних видов вирусов гены состоят из ДНК, у других – из РНК, при том, что молекулы этих кислот существенно различны по составу. Именно поэтому, как справедливо указывает В.И. Шалак, я рассматриваю не только общество, но и всю живую природу как сферу «информационных обменов», т.е. функциональных взаимодействий, основанных на структурных, а не субстанциальных свойствах.

дать более или менее компетентный ответ, нужно совершить небольшой экскурс в историю этой метафоры и ее роли в когнитивных науках.

Идея компьютера появилась в контексте исследования математических проблем, в частности проблемы вычислимости функций, поставленной Дэвидом Гильбертом. В середине 1930-х гг. практически одновременно было предложено два ее решения: λ -исчисление Алонзо Чёрча и α -машина Алана Тьюринга¹⁹. Поскольку почти сразу же была доказана эквивалентность обеих моделей вычислений, в дальнейшем мы будем говорить о чуть более популярной из них – машине Тьюринга (МТ)²⁰.

МТ легла в основу классической парадигмы когнитивной науки, которая в литературе также именуется символизмом или собственно когнитивизмом и в основе которой лежат труды и открытия Ноама Хомски, Зенона Пилишина, Джона Андерсона, Джерри Фодора и мн. др. Исследователи этой волны вполне серьезно полагали, что когнитивные процессы представляют собой серийную последовательную обработку строчек символов-репрезентаций в соответствии с некоторыми синтаксическими правилами. И в задачи новорожденной когнитивной науки входит идентификация физических реализаций этих символов в мозге и реальных правил их преобразований. Этапами формирования этой парадигмы можно считать победоносную, по мнению многих, полемику Хомского против Б.Ф. Скиннера в 1950-х гг., формулирование концепции физической символической системы А. Ньюэллом и Г. Саймоном как модели, эквивалентной сильному ИИ, в 1970-х; и разработку концепции «языка мысли» Джерри Фодором.

Модель, лежащая в основании когнитивного классицизма, имеет свои ограничения: это, в частности, неочевидная как минимум связь с эмпирическими данными и уязвимость перед парадоксом гомункула – в основании чьих-то когнитивных способностей мы предполагаем внутренние операции с семантически нагруженными символами, которые, в свою очередь, требуют чьих-то когнитивных способностей – таких как распознавание, понимание и т.п., – которые, как и всякие когнитивные способности, по определению должны также основываться на внутренних символических операциях, и так до бесконечности. Эти и другие обстоятельства стимулировали поиски альтернативных подходов, некоторые из которых предполагали полный отказ от «компьютерной метафоры», а другие основывались на других интерпретациях того, что есть вычисления: в частности, на признании того, что они не обязательно должны быть серийными, а могут быть параллельными или распределенными, не обязательно они должны быть и дискретными, а могут быть континуальными по предмету или по методам, не обязательно цифровыми, а могут быть аналоговыми, и т.п.

¹⁹ Turing A.M. On computable numbers, with an application to the entscheidungsproblem. A Correction // Proceedings of the London Mathematical Society. 1938. No. 1 (s2–43). P. 544–546.

²⁰ Интересно, что словом *computer* Тьюринг в указанной статье называет не машину, а человека – «вычислителя», – чьи действия машина эмулирует.

Распределенные вычисления

Дальнейшая история смены парадигм в когнитивной науке описана мною в специальной тематической статье²¹ и некоторых других публикациях. В рамках данной статьи нужно чуть подробнее остановиться на модели распределенных вычислений, которые, на мой взгляд, образуют основу и когнитивной, и социальной сферы. Эта модель предполагает, что устройство, осуществляющее такие вычисления, представляет собой систему взаимосвязанных процессоров, каждый из которых выполняет несложные алгоритмы с обратной связью, но в результате система в целом демонстрирует сложное адаптивное (интеллектуальное) поведение. Очевидно ее отличие от нашего настольного компьютера: там программа создается ради достижения конечной цели, пишется программистом, имеющим об этой цели ясное представление. Здесь же программы (алгоритмы) работают на уровне элементарных процессоров, организуя и облегчая их сетевое взаимодействие таким образом, что сеть сама себя оптимизирует так или иначе в зависимости от получаемых ею малопредсказуемых данных.

Далее я предполагаю рассмотреть две модели таких распределенных вычислений, имеющие достаточно широкое распространение в практических воплощениях ИИ. Это *искусственные нейросети (ИНС)*, в которых физически или виртуально связанные между собой искусственные нейроны запрограммированы на то, чтобы менять характеристики межнейронных связей в соответствии с проходящей через них информацией. В результате сеть становится более «обученной» и «интеллектуальной», готовой к выполнению задач в определенной предметной области – фактически, ее вероятностной моделью, способной принимать автономные решения. И это *мультиагентные системы (МАС)*, состоящие из физических или виртуальных агентов, между которыми нет устойчивых связей, но которые способны распознавать друг друга и взаимодействовать на основе вшитых в их процессоры несложных правил. Оказавшись в определенным образом структурированной реальности, такие системы в целом способны продемонстрировать сложное адаптивное поведение, несмотря на когнитивную простоту или даже примитивность составляющих их агентов. При этом создатели и ИНС, и МАС прекрасно отдают себе отчет в «логике» поведения отдельных элементов, которые они целенаправленно программируют, но, как правило, не могут предсказать или даже математически описать эмерджентные результаты деятельности распределенной системы в целом. Но очевидно, что идея ИНС имеет источник в устройстве мозга, а МАС навеяны «архитектурой» животных стай и человеческих сообществ. Поэтому мы вполне можем рассматривать предполагаемую связку таких систем, как модель когнитивно-социальной метасети, и пытаться вывести свойства и характеристики реальных обществ из формальных и алгоритмических особенностей этих моделей.

Однако для этой цели нам понадобится еще одна объяснительная модель, помогающая понять телеологию целостной натуральной вычислительной системы. И мы позаимствуем ее у концепции *предиктивного процессинга (ПП)*, которая в последние два десятилетия порождает неиссякающий вал

²¹ Михайлов И.Ф. Концепции вычислений в современных науках о человеческом познании // Философские проблемы информационных технологий и киберпространства. 2018. Т. 14. № 1. С. 4–22.

публикаций и претендует на роль фундаментальной теории, объясняющей сложное поведение живых, когнитивных и социальных систем на основе единых принципов. У истоков этого движения стоит фигура британского нейробиолога Карла Фристана. Но ее адептами стали когнитивист Якоб Хови, философы Энди Кларк и Томас Метцингер и мн. др. ПП постулирует многослойную иерархическую архитектуру когнитивного аппарата любого живого существа. Верхние слои формируют байесовские модели предполагаемой реальности и транслируют их вниз. Нижние слои сравнивают эти модели с полученными снизу или извне данными, вычисляют ошибки предсказания и транслируют их наверх. Сгенерированные верхними слоями модели уточняются, и цикл повторяется до тех пор, пока разность предсказанных и входящих данных не достигнет минимальных значений.

ПП предполагает еще один способ минимизации ошибок предсказания, а именно путем повторной выборки сенсорных данных, чтобы они соответствовали прогнозам. Это достигается изменением входных сенсорных данных, т.е. изменяя мир через действие. Так, согласно этой теории, 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 – стремления к минимизации ошибок иерархически вырабатываемых предсказаний путем выборочного тестирования сенсорных входов²².

Важным понятием в этой концепции является *ограждение Маркова* (*Markov Blanket*). В статистической сети это наименьший набор узлов, который делает огражденный ими узел сети условно независимым от всех остальных. Поведение огражденного узла можно предсказать, зная только состояния этих ближайших узлов. То же самое и с внешними узлами: огражденный узел бессмыслен для предсказания их поведения. Ограждение Маркова разделяет все важные для организма состояния на внешние и внутренние, а последние – на сенсорные и активные²³. В качестве ограждений Маркова можно рассматривать отдельные органы, функциональные системы организма (кровеносную, эндокринную и т.п.), организм в целом, а также сообщества организмов могут быть описаны как большие вложенные ограждения Маркова.

Таким образом, все организованные во вложенные статистические единства сложные неравновесные системы стремятся к минимизации определенного показателя, который Фристон называет «свободной энергией», – меры разности входных и предсказанных данных. Для этого элементы и уровни системы участвуют в порождении и коррекции моделей внешней для данного ограждения Маркова среды. Эффективность и оптимальность действий всех составных частей системы оценивается относительно оптимальных значений этого общего для системы показателя.

С учетом сказанного реальное общество может быть представлено как метасеть, являющая собой мультиагентную систему, состоящую из обучаемых нейросетей, способных строить и корректировать предиктивные статистические модели окружения. Каждая из нейросетей-агентов есть ограждение

²² Friston K.J., Frith C.D. Active inference, Communication and hermeneutics // Cortex. 2015. Vol. 68. P. 129–143.

²³ Friston K.J., Fagerholm E.D., Zarghami T.S., Parr T., Hipólito I., Magrou L., Razi A. Parcels and particles: Markov blankets in the brain // Network Neuroscience. 2021. No. 1 (5). P. 211–251.

Маркова, вложенное в одно большое ограждение Маркова, стремящееся к достижению одного параметра («свободной энергии», если Фристон прав) и подчиняющее этой общей цели деятельность всех своих узлов. Поскольку, как сказано, узлы представляют собой обучаемые нейросети, их необходимая самокоррекция относительно некоторого целевого параметра неизбежно должна индуцировать в них некие правила, следование которым с приемлемой вероятностью обеспечивает приближение к оптимальному значению этого параметра. И в этом пункте мы вплотную подходим к природе морали.

Мораль как распределенный вычислительный алгоритм

Если потребности целого способны индуцировать определенные алгоритмы поведения частей, то, вероятно, некоторые из этих алгоритмов предстают пред сознательным взором последних (если те обладают сознанием) как правила, которым лучше следовать. Склонность к выполнению этих правил может оправдываться различной мотивацией: от банального конформизма – не выпасть из системы – до возвышенного «я не могу иначе». Но в целом именно в этом механизме я вижу сущность происхождения моральных норм. Тогда моральные нормы – это статистические алгоритмы распространения социальных взаимодействий, которые формируются в ответ на потребности системы как целого, но реализуются на уровне отдельных составляющих сеть процессоров (агентов, индивидов). Их функциональная роль состоит в настройке системы в целом в направлении оптимизации некоторой контрольной переменной – например, энергетического оптимума.

Естественно, что отдельным агентом эти нормы воспринимаются как внутренние регуляторы загадочного происхождения – знаменитый «нравственный закон внутри нас», столь же загадочный, как и звездное небо над нами. Поскольку такие системы развиваются эволюционно, используя случайные мутации для вероятностного угадывания подходящего способа адаптации, то и конкретное содержание запретов («заповедей») исторически случайно и в другом возможном мире могло бы быть иным. Значение имеет только их статистическое влияние на показатели системы в целом.

Моральные нормы часто формулируются как запреты («заповеди»). На уровне целого запреты действуют как механизм сокращения числа возможных состояний системы, что равнозначно нарастанию степени ее структурности²⁴ и уменьшению энтропии²⁵. Весь комплекс алгоритмов достижения системой оптимального значения контрольной переменной – предположительно, энергетического оптимума – формируется эволюционно, в процессе обучения агентов.

Общество, как было показано, может быть рассмотрено как сложный распределенный компьютер, осуществляющий вероятностные вычисления, состоящий из многих сложных распределенных компьютеров, осуществляющих вероятностные вычисления, каждый из которых состоит из таких же компьютеров меньшего масштаба – и т.д., до уровня органических молекул.

²⁴ Анохин П.К. Принципиальные вопросы общей теории функциональных систем // Анохин П.К. Философские аспекты теории функциональной системы. М., 1978. С. 68–72.

²⁵ Gao J., Liu F., Zhang J., Hu J., Cao Y. Information Entropy as a Basic Building Block of Complexity Theory // Entropy. 2013. No. 12 (15). P. 3396–3418.

Естественная социальная система, в отличие от искусственных МАС, может обеспечить их подчинение правилам только с некоторой долей вероятности. Отсюда феномен «моральной свободы» – принципиальная возможность не подчиниться правилу.

Как показывают нейрофизиологические исследования²⁶, механизм общественного одобрения/осуждения действует напрямую через гормональную сферу мозга, и поэтому реакция социального «организма» на поведение агента и в целом его социальная (не)успешность переживается им непосредственно на эмоциональном уровне. Это обстоятельство, как представляется, объясняет и субъективное ощущение «долга».

Поскольку между социальными агентами, в отличие от нейронов мозга, жестких физических связей нет, общества, особенно на ранних этапах своего существования, эмулируют такие связи всевозможными социальными рестрикциями: рабство, сословия, трудовые повинности, армейский призыв и т.п. Однако по мере развития искусственных систем коллективной памяти (священные книги, библиотеки, компьютерные хранилища и т.д.) необходимость в таких рестрикциях постепенно ослабевает, и общество становится более либеральным. Чем быстрее и эффективнее осуществляются информационные обмены, тем меньше необходимость в жестком прикреплении агентов к определенным локусам социального пространства и в конкретных моральных ограничениях. Моральные нормы становятся более абстрактными, поскольку теперь они не столько обеспечивают жесткие социальные связи, сколько, наоборот, компенсируют их отсутствие. Начинает цениться социальная мобильность. Общество как вычислительная система становится адаптивнее и, предположительно, энергетически эффективнее.

Функционирование связки «мозг – общество» можно описать следующим образом. Предположим, мозг принимает и обрабатывает большие потоки данных и, как всякая нейросеть, создает в результате некую вероятностную модель реальности. Она помогает индивиду жить и адаптироваться. Однако у философов возникает в этом пункте юмовская проблема: почему мы должны доверять этой модели? Возможно ли знание, в надежности которого мы уверены? Я бы дал ответ, отличный от кантовского: вмещивается социальная метасеть, одним из узлов которой является наш мозг. Метасеть отбирает вероятностные модели, которые соответствуют ее целям, и как бы сертифицирует их – теперь это общественно одобренное «знание».

Аналогично происходит обработка нормативных моделей поведения, которые также вырабатываются мозгом как вероятностные индуктивные обобщения «лучших практик». Но здесь возникает уже кантовская проблема: каким образом эти случайные нормы могут стать всеобщими и необходимыми? Как и в предыдущем случае, сетевой компьютер, использующий вычислительные мощности множества мозгов, отбирает нормы, которые, будучи приняты как элементы распределенного алгоритма, будут работать на достижение целевых показателей предполагаемых контрольных переменных, в которых заинтересована целостная социальная система. Примитивная иллюстрация: почему шулерство в карточной игре полегит суровому моральному осуждению,

²⁶ Mehta P.H., Josephs R.A. Social endocrinology: Hormones and social motivation // Social motivation. N.Y., 2011. P. 171–189.

а виновных, согласно преданию, всегда наказывали канделябрами? Потому что, если шулерами станут все, игра будет невозможна.

В искусственно созданной МАС правила поведения агентов чаще всего жестко программируются в них самих и, следовательно, не имеют характера моральных норм – искусственный агент «не может поступить иначе» по своей природе, а не по своей «воле». Однако если даже в искусственной МАС программируемых агентов заменить на обучаемые нейросети, то, согласно моей гипотезе²⁷, такие агенты, способные к обучению, способные отличать себе подобных от других элементов окружения, способные к выработке языка²⁸ и правил, в конце концов выработают правила, подчинение которым, собственное и других, будут сами контролировать, порождая видимость следования долгу по собственной воле²⁹.

Выводы

В рамках вычислительного подхода в социальном познании общество и человек понимаются как вложенная вычислительная система мульти-агентной архитектуры. «Вычислительность» как понятие расширяется за пределы машины Тьюринга и рассматривается как способность рассчитывать и предсказывать статистические распределения, максимально соответствующие условиям внешней среды и контрольной переменной – предположительно, некоторого оптимального показателя энергетического обмена со средой.

Общество как мультиагентная система действует на основе правил, имплементированных в агентах. Агенты как обучаемые НС «находят» наилучшие правила в ходе многочисленных итераций обучения. То есть, метафорически говоря, общество *индуцирует* в агентах правила их поведения подобно тому, как магнитное поле индуцирует ток в движущемся проводнике. Индукция осуществляется путем отбора эмпирически найденных индивидами паттернов поведения и сертификации тех из них, которые наиболее полезны для достижения целевых показателей системы. Сертифицированные правила воспринимаются индивидами как, с одной стороны, имманентные им, а с другой – как в принципе допускающие их нарушение, поскольку они формируются в результате статистических, а не детерминированных вычислений. Иными словами, отдельные агенты, как и МАС в целом, – статистические машины, поэтому сохраняется вероятность неисполнения требования правила («моральная свобода»).

Регуляция обществом индивидов осуществляется в т.ч. средствами гормональной модуляции нейронной активности, которая имеет непосредственный эмоциональный эффект (как в случаях влюбленности или наркозависимости). Отсюда – ощущение «морального долга».

²⁷ Которую, наверное, возможно проверить эмпирически – но это не задача философа.

²⁸ Clark B. Facebook's AI accidentally created its own language // TNW. 2017. URL: https://thenextweb.com/news/facebook-ai-accidentally-created-its-own-language#.tnw_YGLIG5bd (дата обращения: 21.04.2021); Schulz R., Glover A., Milford M.J., Wyeth G., Wiles J. Lingo-droids: Studies in spatial cognition and language // IEEE International Conference on Robotics and Automation. 2011. P. 178–183.

²⁹ А адекватна ли такая «слишком человеческая» интерпретация или нет – зависит от реалистической или антиреалистической позиции внешнего интерпретатора.

Вычислительный подход к этике обладает некоторыми теоретическими преимуществами. В частности, он преодолевает индуктивизм натуралистических и утилитаристских теорий и в то же время научно объясняет «трансцендентальность» нравственного закона, не прибегая к идеалистической метафизике деонтизма кантовского толка.

Список литературы

- Анохин П.К.* Принципиальные вопросы общей теории функциональных систем // *Анохин П.К. Философские аспекты теории функциональной системы.* М.: Наука, 1978. С. 49–106.
- Апресян Р.Г.* Нормативная регуляция поведения // *Человек.* 2018. № 1. С. 5–20.
- Бажанов В.А.* Вычисляющая природа – реальность или метафора? // *Философия науки и техники.* 2021. № 2 (26). С. 38–43.
- Михайлов И.Ф.* Концепции вычислений в современных науках о человеческом познании // *Философские проблемы информационных технологий и киберпространства.* 2018. № 1 (14). С. 4–22.
- Михайлов И.Ф.* Вычислительный подход в социальном познании // *Философия науки и техники.* 2021. № 1 (26). С. 23–37.
- Фритт К.* Мозг и душа: как нервная деятельность формирует наш внутренний мир / Пер. с англ. П. Петрова. М.: Астрель; CORPUS, 2010. 340 с.
- Шалак В.И.* Алгоритмические явления в природе: модель объяснения // *Вопросы философии.* 2020. № 11. С. 120–124.
- Шалак В.И.* Алгоритмическая модель социальных процессов // *Философские проблемы информационных технологий и киберпространства.* 2021. № 1. С. 46–62.
- Clark B.* Facebook's AI accidentally created its own language // *TNW.* URL: https://thenextweb.com/news/facebooks-ai-accidentally-created-its-own-language#.tnw_YGLIG5bd (дата обращения: 21.04.2021).
- Friston K.J., Fagerholm E.D., Zarghami T.S., Parr T., Hipólito I., Magrou L., Razi A.* Parcels and particles: Markov blankets in the brain // *Network Neuroscience.* 2021. No. 1 (5). P. 211–251.
- Friston K.J., Frith C.D.* Active inference, Communication and hermeneutics // *Cortex.* 2015. Vol. 68. P. 129–143.
- Gao J., Liu F., Zhang J., Hu J., Cao Y.* Information Entropy as a Basic Building Block of Complexity Theory // *Entropy.* 2013. No. 12 (15). P. 3396–3418.
- Goodman N.* Languages of Art. Indianapolis: Hackett Publishing Co, 1976. 291 p.
- MacLennan B.J.* Natural computation and non-Turing models of computation // *Theoretical Computer Science.* 2004. No. 1 (317). P. 115–145.
- Marr D.* Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Cambridge (Mass.): The MIT Press, 2010. 432 p.
- Mehta P.H., Josephs R.A.* Social endocrinology: Hormones and social motivation // *Social motivation* / Ed. by D. Dunning. N.Y.: Psychology Press, 2011. P. 171–189.
- Milkowski M.* From Computer Metaphor to Computational Modeling: The Evolution of Computationalism // *Minds and Machines.* 2018. Vol. 28. P. 515–541.
- Poggio T.* The Levels of Understanding Framework, Revised // *Perception.* 2012. No. 9 (41). P. 1017–1023.
- Pylyshyn Z.W.* Computation and cognition: Toward a foundation for cognitive science. Cambridge (MA): The MIT Press, 1986. 292 p.
- Schulz R., Glover A., Milford M.J., W̄yeth G., Wiles J.* Lingodroids: Studies in spatial cognition and language // *IEEE International Conference on Robotics and Automation.* 2011. P. 178–183.
- Turing A.M.* On computable numbers, with an application to the entscheidungsproblem. A Correction // *Proceedings of the London Mathematical Society.* 1938. No. 1 (s2–43). P. 544–546.

Social computations and the origin of moral norms

Igor F. Mikhailov

Institute of Philosophy, Russian Academy of Sciences. 12/1 Goncharnaya Str., Moscow, 109240, Russian Federation; e-mail: ifmikhailov@gmail.com

The author proposes a conception in which society and a person are understood as a nested computing system of a multi-agent architecture. The 'Computational' as a concept extends beyond the Turing machine and is considered as the ability to compute and predict statistical distributions that best match the characteristics of the external environment to achieve the optimal value of an assumed control variable. Society induces the rules of behavior in agents similarly to the way by which a magnetic field induces a current in a moving conductor. Induction is carried out by selecting the patterns of behavior empirically derived by individuals and certifying the patterns that are most useful for achieving the target indicators of the system. Certified rules are perceived by individuals as, on the one hand, immanent to them, and on the other, as generally allowing for their violation, since both individual agents and the whole society are statistical machines. The objective probability of non-compliance with a rule is perceived as moral freedom, and hormonal reinforcement of adherence to the general rule is perceived as a sense of moral duty. The computational approach to morality overcomes the inductivism of naturalistic and utilitarian theories, and at the same time scientifically explains the 'transcendentality' of moral law, without resorting to idealistic metaphysics, as does Kantian deontology.

Keywords: naturalism, utilitarianism, deontology, morality, computation, brain, neural network

For citation: Mikhailov, I.F. "Sotsial'nye vychisleniya i moral'ny normogenez" [Social computations and the origin of moral norms], *Filosofskii zhurnal / Philosophy Journal*, 2022, Vol. 15, No. 1, pp. 51–68. (In Russian)

References

- Anokhin, P.K. "Printsipial'nye voprosy obshchei teorii funktsional'nykh sistem" [Principal questions of the general theory of functional systems], in: P.K. Anokhin, *Filosofskie aspekty teorii funktsional'noi sistemy* [Philosophical aspects of the functional system theory]. Moscow: Nauka Publ., 1978, pp. 49–106. (In Russian)
- Apresyan, R.G. "Normativnaya regulyatsiya povedeniya" [Normative regulation of behavior], *Chelovek*, 2018, No. 1, pp. 5–20. (In Russian)
- Bazhanov, V.A. "Vychislyayushchaya priroda – real'nost' ili metafora?" [Computing nature – reality or metaphor?], *Filosofiya nauki i tekhniki*, 2021, No. 2 (26), pp. 38–43. (In Russian)
- Clark, B. "Facebook's AI accidentally created its own language", *TNW* [https://thenextweb.com/news/facebooks-ai-accidentally-created-its-own-language#.tnw_YGLIG5bd, accessed on 21.04.2021].
- Friston, K.J. & Frith, C.D. "Active inference, Communication and hermeneutics", *Cortex*, 2015, Vol. 68, pp. 129–143.
- Friston, K.J., Fagerholm, E.D., Zarghami, T.S., Parr, T., Hipólito, I., Magrou, L. & Razi, A. "Parcels and particles: Markov blankets in the brain", *Network Neuroscience*, 2021, No. 1 (5), pp. 211–251.
- Frit, K. *Mozg i dusha: kak nervnaya deyatel'nost' formiruet nash vnutrennii mir* [Brain and the soul: how neural activity forms our internal world], trans. by P. Petrova. Moscow: Astrel Publ.; CORPUS Publ., 2010. 340 pp. (In Russian)
- Gao, J., Liu, F., Zhang, J., Hu, J. & Cao, Y. "Information Entropy as a Basic Building Block of Complexity Theory", *Entropy*, 2013, No. 12 (15), pp. 3396–3418.
- Goodman, N. *Languages of Art*. Indianapolis: Hackett Publishing Co, 1976. 291 pp.

- MacLennan, B.J. "Natural computation and non-Turing models of computation", *Theoretical Computer Science*, 2004, No. 1 (317), pp. 115–145.
- Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, Mass.: The MIT Press, 2010. 432 pp.
- Mehta, P.H. & Josephs, R.A. "Social endocrinology: Hormones and social motivation", *Social motivation*, ed. by D. Dunning. New York: Psychology Press, 2011, pp. 171–189.
- Mikhailov, I.F. "Kontseptsii vychislenii v sovremennykh naukakh o chelovecheskom poznanii" [Concepts of computations in modern theories of human cognition], *Filosofskie problemy informatsionnykh tekhnologii i kiberprostranstva*, 2018, No. 1 (14), pp. 4–22. (In Russian)
- Mikhailov, I.F. "Vychislitel'nyi podkhod v sotsial'nom poznanii" [Computational approach to social knowledge], *Filosofiya nauki i tekhniki*, 2021, No. 1 (26), pp. 23–37. (In Russian)
- Milkowski, M. "From Computer Metaphor to Computational Modeling: The Evolution of Computationalism", *Minds and Machines*, 2018, Vol. 28, pp. 515–541.
- Poggio, T. "The Levels of Understanding Framework, Revised", *Perception*, 2012, No. 9 (41), pp. 1017–1023.
- Pylyshyn, Z.W. *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: The MIT Press, 1986. 292 pp.
- Schulz, R., Glover, A., Milford, M.J., Wyeth, G. & Wiles, J. "Lingodroids: Studies in spatial cognition and language", *IEEE International Conference on Robotics and Automation*, 2011, pp. 178–183.
- Shalak, V.I. "Algoritmicheskaya model' sotsial'nykh protsessov" [Algorithmic model of social processes], *Filosofskie problemy informatsionnykh tekhnologii i kiberprostranstva*, 2021, No. 1, pp. 46–62. (In Russian)
- Shalak, V.I. "Algoritmicheskie yavleniya v prirode: model' ob'yasneniya" [Algorithmic Phenomena in Nature: An Explanation Model], *Voprosy filosofii*, 2020, No. 11, pp. 120–124. (In Russian)
- Turing, A.M. "On computable numbers, with an application to the entscheidungsproblem. A Correction", *Proceedings of the London Mathematical Society*, 1938, No. 1 (s2–43), pp. 544–546.