**Происхождение ощущений**

Евгений Корниенко

Воспринимаемое субъектом внутреннее состояние, не выражаемое через свойства материальных объектов, и есть "идеальное" ощущение.

**В каком смысле ощущения возникают?**

Они возникают из ничего, или существует некое первичное ощущение?

Пусть построена универсальная самообучающаяся машина с предписанной целевой функцией. Пусть она ещё ничему не научилась, абсолютно ничего не знает и не умеет. Она только что включена, и мы предполагаем, что она пока не имеет субъективных ощущений и переживаний, как и любая другая машина.

Что это за мифическая машина? Нельзя ли использовать для рассуждений менее умозрительный "ощущающий объект", например, хорошо знакомого нам "человека"?

Человек уступает машине в том, что мы не знаем точно, как он устроен. Первые ощущения появляются у человека, когда он активно развивается. Формирование его органов находится в динамике. Эти сложности можно обойти, если использовать заранее изготовленную машину, которую просто включают, чтобы, проследив за развитием её опыта, сделать выводы о её восприятии мира и себя. На примере самообучающейся машины, хотя бы и выдуманной, но обладающей воспроизводимой конструкцией, можно попробовать проследить происхождение сознания с нуля.

Машина пока не знает, как она устроена. Она может что-то узнать только на собственном опыте, которого в начале ещё нет. Мы знаем, как устроены органы и "мозг" машины. Знаем, как организован процесс самообучения, оптимизирующий заданную целевую функцию. Но эти знания не являются "знаниями машины".

В процессе накопления опыта ей придётся общаться с сознательными существами - "субъектами". Придётся показывать другим субъектам своё состояние и свои намерения. Придётся руководствоваться состоянием и намерениями других субъектов при поиске и выборе подходящего поведения.

У машины есть тело - набор органов, систем и сигналов, при помощи которых она обеспечивает свою целостность и взаимодействует с внешним миром и с другими субъектами. Состояние органов и тела в целом задаётся и контролируется специальными датчиками: эффекторами и сенсорами.

Не все состояния и действия машины являются наблюдаемыми со стороны. Только часть поведения машины наблюдаема в виде различимых состояний или динамичных действий.

Некоторые состояния, позиции или малозаметная активность машины могут быть интерпретированы, как "намерение" - предстоящее действие. Намерение как начинающееся действие на самом деле вызвано работой скрытого процесса оптимизации целевой функции. Мы пока не можем говорить, что проявляемое намерение имеет какое-то субъективное (переживаемое) значение для самой машины.

Выраженному намерению предшествует едва улавливаемое самой машиной намерение, которое ещё не развилось до наблюдаемого действия, но уже создало усилия или другие измеряемые собственными датчиками машины изменения в её системах. Как следует из собственного опыта машины, такие изменения через мгновение приведут к наблюдаемому выражению намерения, а затем и к осуществлению действия.

Такое невидимое окружающим собственное состояние можно трактовать как желание, обоснованное или необъяснимое предпочтение одного действия другому. Это "внутреннее состояние" ещё не проявилось в виде "поведения".

Если бы машине пришлось объяснять своё состояние словами, то в первом случае она бы сказала "я поднимаю руку", а во втором - ей пришлось бы использовать формулу "я хочу поднять руку". При этом понятие "хочу" ассоциируется у машины с определённым измеряемым её сенсорами (внутренним) состоянием.

Вместе с тем, внутреннее состояние является для машины таким же наблюдаемым состоянием, как явно выраженное состояние и поведение собственного тела и прочих объектов во внешнем мире. Внутренним состоянием могут обладать вполне "внешние" органы, например, та же рука. В этом контексте "внутренний" - это невидимый извне. Полное состояние органов и систем машины состоит из наблюдаемого со стороны (объективного) состояния и из воспринимаемого только самой машиной (внутреннего) состояния.

Чем вызвана идея "я хочу поднять руку", которую машина может выразить, а может и не выражать в словах? Эта идея основана не на изменениях во внешнем мире, а на наблюдаемом только этой машиной изменении её внутреннего состояния.

Как сама машина, так и внешний наблюдатель, могут не видеть объективных внешних изменений, приводящих к определённому внутреннему состоянию машины. Тем более что это состояние зависит не только от внешних, но и от внутренних причин. Обычно бывает наоборот: намерение выступает как причина последующих событий во внешнем мире. Из-за отсутствия логической цепочки от внешних объектов к внутреннему состоянию не всегда возможно описать своё состояние в "объективных" терминах.

Если некоторые особенности внутреннего состояния машины наблюдаются специализированными "внутренними" сенсорами или имеют необычную для внешних сенсоров природу, то они тоже не могут быть выражены через "внешние" понятия подобно тому, как ощущения голода или страха не представимы через тактильные или слуховые ощущения.

Со временем и при необходимости машина выберет подходящие сигналы для выражения различных оттенков своего внутреннего состояния. За тысячи лет общения друг с другом мы тоже изобрели слова, жесты и другие "внешние" сигналы для выражения своего внутреннего состояния.

**Воспринимаемое субъектом внутреннее состояние, не выражаемое через свойства материальных объектов, и есть "идеальное" ощущение.**

Проводимый мысленный эксперимент направлен на демонстрацию того, что внутреннее состояние специально сконструированной машины может стать для неё субъективным ощущением. Но это не значит, что зрение или другие внешние органы поставляют машине только "объективные" сведения о мире. Внешние материальные объекты тоже воспринимаются как совокупность ощущений. Наблюдение можно трактовать, как ощущение качеств внешнего мира, а ощущение можно считать наблюдением своего внутреннего состояния.

В соответствии с конструкцией нашей машины "внутреннее" и "внешнее" наблюдение выполняется обычными датчиками. Ничего "идеального" в этом нет. Идеальность состоит в том, что машина воспринимает и использует для выбора поведения лишь "качества", ассоциативно связывающие состояния различных датчиков. Универсальность машины запрещает её мозгу использовать конкретные данные для взаимодействия с сенсорами и эффекторами. Мозг оперирует лишь "узнаванием" ситуаций. Видимо, элементарные "узнавания" - это и есть основа "качеств", доступных восприятию машины как субъективные ощущения.

Мы знаем, что на выходе каждого сенсора имеется "результат текущего измерения". Но машина не может выполнить "измерение" по нашей просьбе, так как не использует никаких предписанных логических моделей внешнего мира или своего состояния, например, не используются сведения о конструкции датчиков и о способах их подключения к системе. Машина не знает о существовании датчиков.

Чтобы не спутать наши знания о машине и собственные знания машины, достаточно спросить её, "что она знает о своей конструкции". И если бы она могла говорить, то не сказала бы ничего вразумительного. Для чистоты проводимого мысленного эксперимента, мы лишили её возможности изначально обладать такими знаниями. Но это не мешает ей обучаться. Она даже может научиться говорить.

Таким образом, наблюдаемое универсальной самообучающейся машиной собственное состояние и внешний мир воспринимаются ею в форме ощущений, а не в виде результатов измерений, числовых или символьных данных.

Машина может сделать или не сделать то, что она намеревается сделать, так как обстоятельства меняются, изменяется оценка важности и осуществимости различных действий - алгоритм оптимизации целевой функции постоянно работает. В результате появляется впечатление о "свободе воли" и о "силе воли", например, о способности отказаться от желаемого или терпеть неприятное. Это впечатление - тоже одно из ненаблюдаемых извне внутренних состояний.

Ощущения постоянно детализируются и наполняются всё новым смыслом в результате взаимодействия субъектов, которым приходится учитывать "душевное" состояние друг друга.

Ощущения, как своего состояния, так и внешнего мира появляются у машины не сразу, а после накопления некоторого опыта, когда она в результате самообучения обнаруживает ассоциативную связь между разными событиями, в частности, между своим состоянием и своими последующими действиями. Первое ощущение - это первое узнавание и первое знание о себе и мире.

Поскольку машина не сразу обнаруживает какие-то закономерности, то её первые ощущения возникают, оформляются и детализируются постепенно. В отсутствие необходимого опыта некоторые ощущения могут быть забыты, как забываются непрочные знания. Некоторые чувства могут появиться достаточно поздно, не раньше, чем будут обнаружены соответствующие закономерные связи между разными внутренними состояниями или между состоянием машины и внешними обстоятельствами. Некоторые чувства можно развить путём специальных упражнений.

Не так просто связать духовный мир человека с определёнными объективными явлениями и действиями или с субъективными намерениями. Эта связь часто не прослеживается из-за высокой детальности и из-за социальной обусловленности наших чувств.

Хотя внутреннее состояние не наблюдаемо в виде "чужого ощущения", но оно может быть определено инструментально - измерено. Вполне можно установить, каким именно объективным состояниям организма соответствуют чувства голода, страха, боли. Такие простые и сильные первичные чувства можно надёжно связать с объективным внутренним состоянием. Некоторые "сложные" переживания развиваются только при наличии глубокого знания, недоступного животным и нашей наивной самообучающейся машине.

Что касается человека и животных, то они не были созданы, как экспериментальные машины, и не были однажды "включены". Первые чувства развиваются у ещё не родившегося ребёнка одновременно с развитием самих органов чувств.

**Что чувствует искусственное существо?**

Органы измеряют качественные свойства мира, мозг встраивает эти измерения в ассоциативную смысловую сеть, и при этом они становятся ощущениями.

Обоснованием того, что у нас получается существо, а не автомат, является его способность к универсальному обучению и наличие у него заинтересованности в новых достижениях. Универсальность обучения обосновывается конструкцией "универсального мозга". Стимулом самообучения является внутренняя объективная целевая функция Wish.

Мы умышленно не вносим в машину механизмов или алгоритмов, которые могли бы диктовать, чему и как надо учиться. Машина сама научится чему-то в зависимости от того, какие она получит органы, и каков будет окружающий мир.

Видимое любопытство и заинтересованность в результатах своих действий, избегание невыгодных ситуаций должны убедить нас в том, что машина испытывает эмоции. Какие? Как мы сможем увидеть её любопытство и эмоции? Ведь мы не программировали никакого поведения, в том числе и способа выражения эмоций.

То, что какое-то поведение машины является выражением её эмоций, станет ясно только потом, после общения с ней, и после того, как она достигнет определённого уровня интеллекта. Ведь нигде не сказано, что означает виляние хвостом у собаки. Мы сами разгадали смысл этого виляния, сопоставляя различные поведения собак с тем, что они должны были бы чувствовать по нашему мнению.

И вот, наконец, мы убедились в том, что машина что-то чувствует. Что она чувствует? Каким образом возникли её ощущения и её субъективный внутренний мир?

Допустим, машина освоила точные манипуляции с предметами с помощью рук и надёжное различение предметов при помощи глаз. Теперь она осязает и видит. Объективно можно сравнить остроту нашего и её зрения, диапазоны видимых световых волн.

Наша машина имеет универсальный мозг, которому всё равно, что за датчики он обслуживает. Несмотря на абсолютно равное отношение мозга ко всем датчикам, и всем сигналам, поступающим от них, машина научилась видеть разные цвета. Жёлтый и зелёный цвет - это разные качества внешнего мира. Строго говоря - это разные состояния органа зрения. Приписывание этих состояний свойствам мира - это просто способ описания внешнего мира в терминах нашего восприятия. Итак, у машины есть различные ощущения внешнего мира.

После того, как машина хорошо освоила свой орган зрения, проведём такой эксперимент. Отключим орган зрения - глаза, и на соответствующие входы мозга подадим сигналы, имитирующие какие-то видимые образы. В этом случае машина будет продолжать видеть. Она будет ощущать зрительные качества в отсутствие органа зрения.

Усложним эксперимент и подадим на входы зрения сигналы, соответствующие полной темноте. Всё равно различные зрительные ощущения сохранятся. Мы видим зрительные образы во сне, когда орган зрения фактически не видит ничего подобного. Эти образы ассоциативно вспоминаются под влиянием сигналов от других органов чувств. Ощущения сохраняются, пока опыт таких ощущений имеется в памяти.

Из медицинской практики известны фантомные ощущения: движение отсутствующими пальцами, боль в отсутствующей ноге. Интересно, как долго сохраняется это явление - какова ёмкость оперативной памяти органов?

Следовательно, сами датчики (органы) не являются поставщиками ощущаемых качеств. Они поставляют в мозг что-то другое, некую информацию; а ощущение цвета, боли, касания и т.п. могут быть извлечены изолированным мозгом из прошлого опыта.

Воспользуемся тем, что и глаза и руки имеют и моторные и сенсорные функции.

Проведём более радикальный опыт. Отсоединим руки машины от нервных каналов связи с мозгом и подключим их к тем каналам, которые ранее обслуживали орган зрения. А глаза подключим к каналам, которые раньше были соединены с руками.

Как и в первом опыте, машина поначалу будет продолжать осязать и видеть, но это будут непривычные или бессмысленные образы. В силу универсальности, её мозг постепенно освоит правильное взаимодействие с переставленными органами, такое, которое способствует оптимизации целевой функции. У машины вновь сформируются адекватные ощущения цвета и видимых образов, осязание и точные движения руками.

Похожая, но менее радикальная перестройка, происходит у человека, который начинает носить переворачивающие очки. Можно провести подобный эксперимент по имитации слуха зрительными образами или наоборот. Впечатления слепой женщины, которая использовала звуковой локатор для компенсации отсутствующего зрения есть на страничке "Видеть при помощи слуха".

Из этого опыта следует, что конкретные ощущения качеств, относящиеся к зрению, не зависят от того, к каким входам универсального мозга подключены глаза. Зрительные ощущения формируются постепенно и встраиваются в систему прочих ощущений свойств мира по мере накопления опыта использования зрения.

По способу формирования эта система неизбежно становится моделью внешнего мира, устройство которого не зависит от способа подключения органов к мозгу.

Итак, первоисточником зрительных ощущений является внешний мир, но после того, как система зрительных ощущений сформирована, она остаётся в сознании, а объективно - в памяти, ещё некоторое время даже и без участия органа зрения.

При отсутствии положительного (приводящего к оптимизации целевой функции "Wish") опыта взаимодействия с внешним миром ощущения качеств этого мира вообще отсутствуют. Они не присущи самообучающейся системе самой по себе.

"Датчик Wish" является существенной и обязательной частью универсального мозга. Основная функция универсального мозга - минимизация сигнала Wish.

По мере подключения к универсальному мозгу всё новых и новых физически различных органов, у нашей ощущающей машины появляется всё больше качественно различных ощущений свойств внешнего мира, но только в том случае, если использование новых органов влияет на функцию Wish.

Важно не только улучшение значения Wish, а практическое влияние на величину Wish. Какое-то хронически правильное поведение перестаёт влиять на Wish и становится неосознаваемым, в то время как небольшое изменение этого поведения могло бы ухудшить значение Wish.

Можно сказать, что органы измеряют качественные свойства мира, мозг встраивает эти измерения в ассоциативную смысловую сеть, и при этом они становятся субъективно воспринимаемыми или невоспринимаемыми (автоматическими) ощущениями. У детерминированной, алгоритмической машины сигналы датчиков всегда являются автоматическими.

У универсальной обучающейся машины эти сигналы, если они способствуют достижению лучшего значения целевой функции, становятся ощущениями, а после того, как их использование перестаёт изменять целевую функцию, например, при достижении оптимума, эти ощущения перестают осознаваться. Таким образом, осознавание существует только во время обучения, или в любой ситуации, когда текущее использование соответствующих органов влияет на величину целевой функции Wish.