

Основы Онтологии 0.1

30.12.18

Прапион Медведева

Замечания

Это драфт книжки по онтологии, которая создается как поддержка курса по онтологии в Школе системного менеджмента. В готовой книжке предполагаются задачи для закрепления важных когнитивных навыков и примеры использования. Текущая версия — фиксация минимального содержания курса.

Введение:

Это книжка про то, как смотреть на мир и обращаться с ним, чтобы было удобно и эффективно с точки зрения соотношения затрат и полученной точности результата.

Иногда кажется, что обычный, повседневный, бытовой способ смотреть на мир — и есть наилучший из возможных в обычной жизни: все остальное пригоняется или в узкоспециализированных областях, или не пригоняется вообще.

Наверняка, когда вы сдавали философию в вузе, вам казалось, что люди почему-то напридумывали дико вычурных способов обращаться

с миром, своим знанием, процессом познания, коммуникацией с другими людьми и так далее. Мне тоже кажется, что вычурных способов много. Но при этом я уверена, что бытового способа недостаточно: посмотрите, сколько проблем из-за недостаточных возможностей интеллекта и затрудненной коммуникации между людьми.

Это приводит к тому, что человечество все еще не умеет решать по-настоящему сложные проблемы (хотя определенных успехов все же достигло). Если не ввести в обычный человеческий обиход более сильные способы мышления, в первую очередь, коллективного, то светлого будущего нам не видать — и просто комфортной жизни тоже, потому что и так затянули с решением глобальных проблем дальше некуда. Эта тема заслуживает отдельной большой книжки (не этой), надеюсь, что вы понимаете мою мысль и не буду распространяться. Если что, можно пойти по ссылкам и почитать¹.

Так что в этой книжке я пытаюсь сжато рассказать, как по моему мнению (и не только, надо сказать, моему), нужно думать о мире и разговаривать с людьми, чтобы более сложные вещи становились возможными.

Стоит сказать, что остается за пределами книжки:

¹ <https://80000hours.org/career-guide/world-problems/>

1. Этика почти целиком. Поскольку книжка базового уровня, я предполагаю, что вы практикуете навыки, описанные в ней, в дружественной среде, где нет откровенно злонамеренных волей.
2. Конкретные подходы к решению задач, такие как системное мышление, прикладная рациональность, всякие управленческие практики и прочая прочая — все это темы для отдельных текстов, а здесь дается только кусок до-мыслительного, по большому счету, характера.
3. Целеполагание — это всегда на совести читающего и надо понимать, что именно на ней, ответственность тут не делится.
4. Проблемы управления собой и своими состояниями. Про это предостаточно книжек у психологов всех мастей, и хотя, разумеется, вы или я — такие же части мира, как и все остальное, практики там специфические и непростые, между делом не опишешь.

Полезно эту книжку читать тем, кто хочет получить способ укладывать все, что попадает в руки, в какую-то понятную канву, притом компактно и пригодно для передачи другим. Последний момент важен особенно: если мы не можем знание о мире (в широком смысле) передать другим, то максимум, что мы можем сделать — это то, что умещается в одной голове. А это очень и очень скромно.

Я опишу несколько концепций, связанных между собой, которые вместе работают как машинка для операций с картиной мира: построением, обновлением, передачей.

То, что здесь описано — материал курсов, которые я в Школе Системного Менеджмента² вела в 2018 году.

² <http://system-school.ru/ontologics>

Часть 1. Прагматизм

Что такое прагматизм?

Одна из ключевых тем, которые поднимаются в этой книжке (и на курсе, который ею поддерживается) — это тема *прагматизма*. Само слово — обозначение философского течения конца XIX - начала XX века, читайте Ч. С. Пирса³. Заметьте, это довольно старые идеи. В первоначальном виде там много всяких несущественных или ошибочных наслоений, так что читайте с опаской.

Однако прагматизм в целом, в его, так сказать, духе, лично я считаю очень хорошей идеей из-за принципов, которые лежат в его основе:

1. **Нет различия в следствиях — нет различия в идеях.** Так говорит философия прагматизма и таким образом снимает жуткие вечные споры о словах⁴. Если за разными словами стоит одинаковое содержание, то неважно, какие слова. Конечно, там есть и обратная сторона: иногда небрежное обращение со словами ведет к затруднению понимания, иногда ваша картина мира получается бессвязной, а действия бессмысленными, особенно если вы не научились замечать, что какие-то слова были неаккуратны или отсылали к несуществующим явлениям. Позже будет еще много о применении этого принципа прагматизма, а пока нужно просто над ним помедитировать. *Нет различия в следствиях — нет различия в*

³ <http://www.digitalpeirce.fee.unicamp.br/home.htm>

⁴ https://lesswrong.ru/w/Споры_об_определениях

идеях.

Если от того, верите вы в бога или нет, не происходит ровно никаких изменений в вашей жизни, то можно сказать, что нет разницы, верите вы в бога или нет.

2. Содержание идеи или понятия в тех практических действиях, которые из них следуют. Я еще раз подчеркиваю здесь связь модели в ментальном пространстве (поясню подробнее в ч.2) и реальных действий в окружающей действительности. Здесь, конечно, большие пересечения с этикой, в которую я обещала не лезть (и не буду), но на более атомарном уровне можно сказать, что так мы защищаемся от пустой или благообразной но вредной болтовни.

3. Критерий истины — полезность. Если расшифровать, то получается вот что: только если возможно применить некоторое знание для воздействия на окружающую среду, причем так, чтобы результат соответствовал нашим ожиданиям, мы называем что-то истиной. Более того, отходя от классического прагматизма, будем заранее считать, что у знания — вероятностная природа. Так что получается, что мы можем присваивать высокую вероятность содержанию некоторых утверждений, только если у нас есть хорошие, сильные свидетельства, что именно так оно и работает. Здесь же мы закладываем основы байесианства (тоже расскажу позже, в ч. 3), которое пригодится нам еще не раз.

4. Иногда нет полной информации и приходится верить в то, что предположение сделано верное; потом действовать исходя из этой веры. Это тоже одно из положений классического

прагматизма. Оно, буквально, значит следующее: если наши знания о каком-то предмете несовершенны, полной информации нет, а принимать решение нужно, то можно (сюрприз) принимать его и без полной информации. И не только принимать, но и исполнять. Это вместе с вероятностной природой знания обозначает весь наш фрейм, в котором мы будем рассуждать о мире, моделировании, принятии решений.

Прагматизм и прагматика

Есть похожее слово, которое изначально обозначает совсем другую область: прагматика, или речевая прагматика. Это такой раздел семиотики (где еще есть семантика и синтактика)⁵.

Прагматика занимается тем, что выясняет, *что в каком качестве* можно использовать в речи, чтобы донести *какую* мысль и чтобы указать на *что* в мире.

С самого начала там есть Грайс⁶, который предложил общий фреймворк для рассуждений о том, как общаются люди.

Параллельно с этим всем мы наблюдаем развитие *семантики* тоже, или дисциплины о связи знака и обозначаемого объекта. Оказалось, что прагматика некоторым образом диктует семантике, что делать: как мы употребляем знак (слово) так оно и значит. Как нам с этим удобно обращаться, так мы и будем это использовать. Наконец, что

⁵ <https://ru.wikipedia.org/wiki/Семиотика>

⁶ <http://kant.narod.ru/grice.htm>

достигает наибольшей эффективности, то и верно как коммуникативная максима.

Поначалу семантика много занималась тем, что пыталась выяснить, что на самом деле значат слова и какая теория лучше объясняет то, устроен язык. Потом, конечно, переключилась на то, что соответствует языковой интуиции носителей языка, а потом и вовсе ушла или в формализмы или в тесную связь с прагматикой.

Так вот, давайте изначально договоримся: слова ничего не значат "на самом деле". Все контекстно зависимо, слова это просто знаки, которые каким-то образом отсылают или не отсылают нас к кусочкам реальности⁷.

Поэтому нас перестает интересовать "что значит X", и начинает интересовать "как нам использовать X", причем это касается не только предметов, но и речи. Или не только речи, но и предметов, это уж как понятнее.

Объяснение и предсказание

Следующее важное различие в употреблении слов, которое понадобится нам дальше — это различие между объяснением и предсказанием.

7

<https://www.hse.ru/mirror/pubs/lib/data/access/ram/ticket/14/1546163696851742b6990773b04dacbf25d67264a/Контекстуальность%20и%20композициональность.pdf>

Оно ровно то же, что позволяет отличить современную науку: можно пытаться задним числом объяснить что-то, и как правило, будет получаться. Хорошая модель — не та, которая объясняет (так, что у нас возникает чувство когнитивной легкости), а та, которая предсказывает будущие состояния мира, и отделяет возможные состояния от невозможных. Если ваша модель этого не делает, то с ней что-то не так⁸.

Любое высказывание, которое вы можете сделать о мире, нужно мочь перевести (пусть развернуто и коряво) на язык будущих наблюдений, иначе это высказывание будет никак не полезно в вашей картине мира.

Целеориентированность

Главное, что мы принимаем во внимание — это действие (конечно, речевой акт тоже акт, об этом не стоит забывать). И оно должно быть целеориентированно. Целеориентированность в действии никогда нельзя терять из виду, хотя конечно все время все пытаются отвлечься на то, как на самом деле.

Правильные вопросы: зачем я это делаю? чем полезно это высказывание? что для меня означают в плане действий эти слова?

⁸ <https://lesswrong.ru/w/Лжепричинность>

Неправильные вопросы: что это значит на самом деле? что значит то или иное абстрактное понятие? как нужно поступить чтобы быть хорошим человеком?

Спектр мышления от интуитивного до формального:

Для поддержки не совсем формальных и не совсем точных, но зато быстрых и очень нужных рассуждений о мире нам понадобится еще один инструмент: спектр формальности мышления.

Типовые профили мыслительной компетенции



Это похоже на то, что у Канемана описано в "Думай медленно, решай быстро"⁹, но модель немного расширена тем, что мы говорим о спектре и компетенциях.

⁹ <https://www.litres.ru/daniel-kaneman/dumay-medlenno-reshay-bystro/>

С этой точки зрения обращение со знаниями у человека выглядит примерно так:



Часть 2. Карта, территория и прочие метафоры

В прошлой части я драматично заявила, что слова "ничего не значат на самом деле". Теперь поясню, что я имею в виду.

Карта и территория

Для более простого понимания этой концепции давайте пользоваться метафорой карты и территории. Мир — это территория. Отображение мира у нас в голове — это карта. Карта никогда не равна территории и даже никогда не дает масштаба 1:1 — это и не нужно, иначе нам было бы очень неудобно ею пользоваться.

Поскольку в содержании высказываний мы оперируем кусочками карты, мы про них не можем сказать истинны они или ложны со стопроцентной уверенностью. Все, что мы можем, это утверждать, что если мы сделаем в мире X, то получим Y. Это и есть свойства высказываний как кусочков модели: точность, предсказательная сила, наша уверенность в них. Абсолютной истины нет, и она не нужна, так как истинность (как и другие семантические понятия) могут определяться только в мета-языке по отношению к данному объектному¹⁰.

Реальность получается недоступна нам напрямую, но мы отлично научились делать простые и повседневные ее модельки: мы ходим и не врезаемся в стены, что хорошо. По этой же причине нельзя сказать

¹⁰ <http://khazarzar.skeptik.net/books/tarski01.htm>

про какое-то высказывание, что оно “объективно”. Я утверждаю, что бывают только интерсубъективные высказывания, которые разделяются бОльшим или меньшим количеством субъектов. Мы все же хотим аккуратно различать такие представления о реальности, которые разделяете только вы и такие представления, которые разделяет большая часть людей на свете. Конечно, и там может не сойтись: вы думали, что большая часть людей разделяет, а она не. Именно так и возникают заблуждения, неприятные недопонимания и так далее, но все же, по большей части это работает хорошо. Доказательством этому служит огромное множество бытовых коммуникативных актов, которые проходят удачно.

Нужно еще упомянуть, что ваша карта может стать объектом рассмотрения другого человека (или вас самих!) и тогда у вас появится карта второго, третьего, четвертого уровня и так далее. Обычно, не следует ходить выше, чем на 3-4 уровня вверх: оно ни за чем не нужно и отнимает время и внимание. Один из важных навыков, который я ожидаю, что вы получите к концу книжки (если будете делать задания!) — находить и не терять тот уровень меты (в карте-территории), который вам интересен, и не бродить по остальным впустую.

Все это время мы говорили исключительно об описательной составляющей нашей работы по познанию мира. Есть еще другие способы с ним взаимодействовать, как минимум, мы хотим менять мир

в соответствии с нашими предпочтениями, для чего неплохо было бы уметь эти самые предпочтения учитывать.

Возможные миры

Для этого обратимся к терминологии возможных миров. Возможные миры это такой конструкт, чтобы представить себе много почти одинаковых миров, которые отличаются друг от друга только одним небольшим заданным параметром. Например, в одном мире стол, за которым я сижу, белый, а в другом — черный. Теперь я могу сказать, какой из этих миров я предпочитаю и еще много всего интересного могу сказать про это множество миров. В частности, могу сказать, какой мир вероятнее, в каком мире больше информации, какой мир подойдет или не подойдет какому-то другому агенту по его предпочтениям.

Еще мы обычно умеем обращать внимание не просто на мир, или на карту, а на себя-в-мире, то есть моделировать последствия наших будущих в этом мире поступков.

4D и ментальное пространство¹¹¹²

Территорию мы будем называть 4D объектами, то есть такими, которые имеют протяженность (экстент) в пространстве и времени. Кроме уровня 4D есть еще уровень **концептов**, некоторое ментальное пространство, в котором мы это 4D мыслим, объединяем в классы, совершаем какие-то операции, моделируем, представляем себе и так

¹¹ <https://en.wikipedia.org/wiki/BORO>

¹² <https://arxiv.org/abs/1502.00121>

далее.

Итак, по большей части мы будем работать с двумя принципиально разными типами штуковин (специально беру незарезервированный термин!):

- 4д объекты — штуковины, которые занимают место в пространстве-времени;
- штуковины в ментальном пространстве, которые не занимают места в пространстве-времени (по крайней мере вне активности мозга).

Так получилось, что в языке нет автоматической разметки, какое слово чаще к чему отсылает, поэтому где необходимо, будем пользоваться уточнениями, это я сейчас о столе как 4D говорю, или о столе вообще.

Больше того, скажу еще, что все слова — уже концептуализации начального уровня. Как только мы выделили объект из фона и назвали его, мы уже сделали его концептом (имеем в ментальном пространстве его концепт).

Разница только в том, что про некоторые слова все более-менее согласны про значение, то есть пространство смысла¹³ там не сильно большое и размазанное, а в другом — такое огромное, что прямо легче другие слова подобрать. Тут наверно надо пояснить про эти самые пространства смыслов. Для нас очевидно, что в языке все со

¹³ https://en.wikipedia.org/wiki/Conceptual_space

всем связано “как-нибудь”. И если связь крепка (например “стол — столик”) то мы говорим, что эти слова похожи. Бывают разные способы связей, не только по подобию. Бывают еще всякие предметы, которые не в материальном мире, например “любовь” или “дружба” и про них мы тоже кое-как можем друг с другом общаться, несмотря на то, что, вообще говоря, имеем в виду скорее всего совершенно разные штуки.

Совершенно, да не совершенно. Есть явления, которые большинство людей отнесет к “любви”, например, картинку, где двое держатся за руки, или что-то такое. И есть явления, которые менее вероятно отнесут к любви, типа забирания последнего куска хлеба.

Все подобные понятия, собирательные, зонтичные, которые трудно потрогать или указать на них, являются абстрактными. Я обычно еще говорю, что понятия бывают размытыми: то есть, если вы с собеседником или какой-то группой договорились о том, **что** значит какое-то абстрактное слово, то оно продолжает быть абстрактным, но становится менее размытым. Если не договорились — то пиши пропало, будет непонятно о чем речь, будут разные ожидания, будут ошибки в коммуникации.

Любые наши высказывания должны быть сводимы к высказываниям о 4d, а еще лучше к высказываниям-предсказаниям.

Построение онтик

Первым делом определяем, что у нас домен: какую осязаемую часть реальности мы сейчас будем рассматривать. Чтобы ее отметить или обозначить совершенно необязательно использовать слова: можно представить себе границы предметов, можно указать на них пальцем, можно обозначить границу, можно вообще указать на требуемую часть любым доступным способом.

После того, как у нас появился **домен**, мы можем выделить там объекты и отношения. Здесь начинается сложное, потому что само то, как мы выделяем объекты из фона (на одном и том же куске реальности) в разные моменты времени или с разными людьми, очень сильно задано тем, какой у нас в ситуации интерес и какой мы **стейкхолдер**. Раз выделенные объекты будут разные, то и отношения будут разные, разных типов и разных качеств, и все операции, которые мы сможем совершать тоже будут отличаться.

Так что получается, что онтик для одного и того же кусочка мира может быть много — столько, сколько взглядов на этот кусочек мира существует.

После того как у нас есть первичная концептуализация, очень близкая к реальности, мы можем абстрагироваться дальше. Обычно нас интересует вовсе не уровень отдельной вещи или отдельного процесса, а уровень хотя бы некоторых обобщений над ними, уровень **классов**.

Чтобы объединять какие-то объекты в классы очень важно верно нащупать то *существенное* в объекте, что мы оставим, и отбросить именно несущественные детали.

Онтологическое описание

То, что у нас получилось по итогу такой работы — *онтологическое описание*. Конечно, оно еще кое-какое, но вполне себе описывает некоторую часть реальности.

Итак, домен у нас в физическом мире, это часть реальности, онтика это отображение части реальности в ментальном пространстве, онтология это такая онтика, которая описывает весь мир, а онтологическое описание — это зафиксированное на носителе отображение части мира. Да, конечно, сам носитель тоже часть мира. Но бумага, на которой описан ваш план проекта вряд ли интересует вас именно как бумага.

Онтологии разных уровней

Обычно мы имеем дело с онтологией уровня M1 — мы описываем конкретную задачу, предметы и отношения в ней. Есть еще уровень обобщения, где мы хотим описывать все подобные задачи и то, что может для них потребоваться. Далее идет онтология предметной области — там не только задачи, похожие на нашу, но и прочие задачи, которые могут нам встретиться в этой области.

При этом есть области, в которых до уровня 4D очень близко — например, производство чего угодно. А есть такие, в которых ваша непосредственная деятельность будет так далеко от 4D, что вы и в глаза никогда не увидите ваш вклад в конечный продукт. Что ж, это ваш уровень абстракции, и вы занимаетесь им. Важно понимать, что его нужно уметь проследить аж до 4D, иначе очень может быть, что вы будете делать непонятно что, непонятно зачем и вообще непонятно.

Основные типы отношений

Хотя вы всегда можете сами придумать и описать отношения, которые есть в ваших онтиках, я перечислю тут основные, по которым можно вкладывать концепты друг в друга, как матрешки. Нам важно уметь вкладывать что-то во что-то другое, потому что это базовый способ справиться с многообразием мира: если не собирать штуки в кучи, категории, матрешки, классы, типы (что угодно), то придется их все воспринимать по одной и тогда у нас очень быстро кончится место в голове.

Первое отношение, которое я хочу описать, это отношение композиции, или отношение часть-целое. Оно означает буквально “один объект является частью другого объекта”, “один объект занимает в пространстве часть места другого объекта”. Например, ножка стола является частью стола. Или мои органы являются частью меня.

Обычно мы говорим об отношении часть-целое не только в пространстве, но и во времени. Тогда это называется "темпоральная часть". Например, я в момент, когда пишу это — темпоральная часть "меня вообще". Такой финт позволяет нам легко и удобно представлять процессы: как совокупность всех входящих в него объектов (их экстенгов, то есть протяженности в пространстве) и время, которое этот процесс разворачивается.

Отношение часть-целое важно еще и потому, что по такому принципу строятся холархии — основные единицы, которыми вы будете оперировать при изучении системного мышления.

Второе важное отношение, которое я хочу здесь описать, это отношение специализации. Еще его называют отношением класс-надкласс-подкласс. Например "стулья -- мебель -- коричневые стулья".

И третье отношение — классификация, то есть отношения между классом и экземпляром класса, вроде "стулья вообще -- конкретный стул -- стулья по типу материалов". Последнее еще называют классификаторами.

Как вы уже заметили, по этим отношениям можно делать иерархии произвольной длины. В таких иерархиях важно не пропускать уровни, когда их описываете. Если пропускать, то можно случайно проскочить тот уровень, который вам окажется важен. В частности, по отношению "часть-целое" уровень — это системный уровень и его обычно можно

выделить по тому, какой язык для его описания используется: в каком домене и с каким интересом люди об этом говорят и думают.

Но о людях (и стейкхолдерах с их интересами) мы еще немного поговорим позже, а сейчас сначала разберемся как происходит, собственно, моделирование.

Моделирование

Выше мы говорили об онтиках и о том, как мы их строим: выбираем метод описания, выбираем инструменты описания и делаем у себя в голове ментальное отображение некоторого кусочка реальности некоторым способом. Еще иногда нам нужно описывать что-то для предсказания поведения этого чего-то. Такая репрезентация называется **моделью**. Модель может быть в ментальном пространстве, может быть в физическом. Например, всем известны деревянные модели кораблей (они моделируют конструкцию). Мы можем на этой модели понять, как будет вести себя реальный корабль, если из него вынуть какой-то ключевой элемент. Конечно, по большей части эти штуки используются исключительно в эстетических целях, и по сути тогда они не являются моделью, а являются больше скульптурой или макетом.

Я скажу больше: в каком-то смысле что угодно может быть моделью чего угодно, вопрос только в том, насколько точной и какой именно аспект поведения оно предсказывает. Допустим, ручка может быть моделью смартфона в том, как оно ведет себя брошенное с высоты.

Мы, кстати, часто так делаем: когда проверка какой-то гипотезы на целевом объекте дорогая и/или сложная, мы проверяем гипотезу на модели.

К модели обычно нужен интерпретатор. Только интерпретатор позволяет соотнести то, что вы наблюдаете на модели с тем, что вы будете наблюдать на реальном объекте. Без этого, можно сказать, что основа модели — онтологическое описание, чаще всего, схема или схемоид.

Часть моделей существует только в нашем ментальном пространстве: большинство моделей мира живут там. Вы успешно ходите и не врезаетесь в стены, вы умеете ловить летящие в вас объекты, вы понимаете выражения лиц и можете предсказывать довольно сложное поведение людей (знакомых и не очень). Конечно, эти предсказания иногда точны, иногда не очень, иногда ваша уверенность в них неоправданно высока, иногда нет, но об этом мы тоже поговорим позже, когда будем говорить о байесианском обновлении моделей.

Для того, чтобы построить модель, неважно, в ментальном пространстве или нет, вам придется определить, какие свойства целевого объекта вы собираетесь предсказывать, после этого станет понятно с выбором инструментов, способом моделирования и тем, что нужно отбросить как неважное.

Большую часть моделей мы либо строим изначально для передачи кому-то, кто будет ими пользоваться, либо строим вместе с кем-то,

либо потом их достраивают другие. Короче, процесс это очень коллективный. Поэтому дальше придется разбираться, почему мы с разными людьми очень по-разному можем видеть мир, и как из этого собирать общие модели.

Вьюпоинт, стейкхолдеры и коллективное моделирование

Как я уже говорила в начале, то, как мы выделяем объекты из фона сильно зависит от того, что мы собираемся с ними делать, или, иначе говоря, какой наш интерес в обсуждаемом куске реальности. То, что позволяет нам по-разному выделять объекты из фона и распределять в группы, называется методом описания или вьюпоинтом. То, какой у вас вьюпоинт связано с тем, какой у вас в ситуации *интерес*, и какой вы, следовательно, в данный момент *стейкхолдер*.

Стейкхолдер — это, если говорить по-простому, как-то заинтересованный в ситуации актер. Мы представляем личность, как состоящую из трех уровней — уровень воли, уровень актера и уровень стейкхолдера. Стейкхолдер имеет интерес в конкретной ситуации.

Актер играет роль то одного стейкхолдера, то другого, а воля позволяет выбирать, быть ли сейчас вообще актером, или, например, пойти поесть.

Очень хорошо бывает понимать, разговаривая с человеком, какой он в данный момент стейкхолдер, какой у него интерес. Еще лучше — понимать это про себя тоже. Тогда у вас появляется взгляд на ситуацию со стороны анализа интересов и вам становится *компактно*

понятно, почему человек говорит те или иные слова и чего он вообще хочет. Это помогает не затевать конфликты и вообще вести переговоры. Еще это помогает строить совместные модели, в которых учтены все основные вьюпоинты и интересы стейкхолдеров. По этим совместным моделям потом можно заниматься совместной деятельностью. Это то, что называется "найти общий язык".

Мы можем определять то, какой человек сейчас стейкхолдер по тому, какой язык он использует: какие слова и выражения (концепты) занимают его внимание? Как он группирует понятия? Как только вы поняли, что для него важно, вы можете с помощью уточняющих вопросов понять, какой у него в ситуации интерес, проще говоря, что он хочет чтобы было, и чего он хочет, чтобы не было.

Об этом же шла речь, когда мы говорили о возможных мирах: мы умеем описывать разные миры, которых еще нет, и указывать, какой из них для нас предпочтительнее.

Все эти операции можно проделывать, пользуясь разными инструментами из спектра мышления, о котором мы говорили в начале. Вы можете делать это формально, обозначив заранее язык, а можем делать это очень неформально и пользуясь догадками. Об этом хорошо пишет Грайс по имплицатуре ([ссылка выше](#)).

Попробуйте это поделать и последить, в какой части спектра мышления вы находитесь.

Язык для описания

Очень важный кусок для понимания, это что один и тот же кусочек реальности можно описывать не только с разных точек зрения, а еще и на разных системных уровнях. Это включает в себя смену языка: в частности, язык для описания ментальных состояний отличается от языка для описания физических объектов. Иногда мы пытаемся описать сложные абстрактные понятия, вроде любви или дружбы, описанием процессов в мозге — это обычно оказывается бесполезно. Типичный пример выбора не того системного уровня.

Иногда чтобы хорошо предсказывать поведение какой-то штуки вам достаточно тыкать в нее палочкой и смотреть, что будет. Иногда этого недостаточно и приходится смотреть, как она там внутри устроена. Лучше всего сочетать эти методы, потому что в чистом виде ни один из них нормально не работает. Для этого и приходится выбирать нужный системный уровень плюс 2 уровня туда и сюда¹⁴.

Нужно уметь сконцентрироваться именно на том системном уровне, который вас интересует. В данном случае простой способ сделать это — попытаться построить гипотезу про мир или обозначить желаемый мир. Технически, обозначение предпочтительного мира — это тоже гипотеза о вашем будущем внутреннем состоянии. Поэтому мы можем привести вообще все высказывания о важных для нас кусках мира в

¹⁴ Подробнее про системные уровни и вообще системный подход читать <https://www.litres.ru/anatoliy-levenchuk/sistemnoe-myshlenie/>

форму гипотез о состоянии мира или о нашем внутреннем состоянии в связи с состоянием мира.

Некоторые высказывания легко декомпозируются как высказывания о фактах, которые просто проверить: если сейчас я выйду на улицу, я увижу голубое небо. Некоторые проверить не так просто, потому что они включают в себя сложные запутанные предпочтения и оценки, а также свернутые, абстрактные, размытые концепты. Например: “коммунизм — лучший общественный строй”. Есть способ работать с такими размытыми концептами — нужно проговорить про себя эту фразу и посмотреть, какая сенсорная информация вам сразу представляется: что такое в вашей голове коммунизм, и что конкретно вы имеете в виду, говоря, что это — лучший общественный строй. Если вам будет удаваться в случаях таких размытых и неясных концептов приходить к уровню описания конкретных 4D- объектов: людей, процессов, явлений, занимающих место в физическом пространстве и времени, то прекрасно, вы знаете, что вы имеете в виду. Если такое не получается, и у вас есть только смутное ощущение — придется что-то с таким убеждением (то есть, любым осмысленным высказыванием о мире) сделать, как-то его переформулировать или вообще выкинуть.

Еще один прием для проверки ваших убеждений на прочность — это спросить себя, что могло бы это ваше убеждение поколебать. Если у вас есть такие убеждения, что вы не можете представить себе информацию из мира, которая бы его изменила,

(нефальсифицируемое¹⁵) то очень возможно, что это убеждение не очень четко сформулировано и вообще я бы рекомендовала от таких убеждений отказываться, они не имеют никакой связи с реальностью (см. раздел про прагматизм).

Моделирование в классах

Моделирование происходит в классах. Мы хотим описывать законы, по которым взаимодействуют не конкретные 4D-объекты, а “подобные” объекты. Иначе и нельзя — чтобы точно понимать, как поведет себя тот или иной объект, нам нужно все проверять именно на нем. В этом смысле моделью объекта X мог бы являться только сам объект X, что бесполезно. Кроме того, моделирование в классах позволяет переносить выясненные закономерности из одного кусочка мира в другой, похожий кусочек, что существенно упрощает почти всем людям жизнь, несмотря на то, что они проделывают это по большей части неосознанно.

Коллективное моделирование и внешние модели

Как я уже говорила, модели, которые вы создаете — по большей части внешние и создаются для передачи другим людям. Существенная часть моделей находится в экзокортексе, “внешней коре”, то есть на всяких девайсах, бумаге, отчужденных и задокументированных схемах. Люди используют разные инструменты для моделирования: от бумаги и до специального софта, для разных задач разного. Еще

¹⁵ <http://filosof.historic.ru/books/item/f00/s00/z0000197/st098.shtml>

используются, конечно, разные языки моделирования. Самый простой и стандартный вид модели, которым обмениваются люди — это текст. Я здесь имею в виду не только написанный текст, но и произнесенный, конечно.

Умение восстанавливать из текста модель (с гипотезами и стейкхолдерскими интересами, породившими эту модель) — ключевой навык, если вы хотите заниматься моделированием осознанно. Хотя, конечно, большинство людей моделируют все подряд неосознанно, не очень точно и получают не очень хорошие результаты в предсказательной силе модели.

Самый распространенный способ получить модель — *спросить стейкхолдера*. Вообще, мы крайне редко в сколько-то сложных и интересных для рассмотрения вещах получаем информацию напрямую из мира — большая часть интересующих объектов реальности нам недоступны, или их очень много или что-то еще такое, что делает “снятие модели с реальности” невозможными. Поэтому основная практика моделирования большинства интересных вещей — расспрашивание стейкхолдеров, выявление их интересов и их собственных моделей происходящего, и построение большой модели моделей.

Еще люди обмениваются частями одной модели, чтобы все получили больше релевантных предсказаний, то есть, по сути, научились порождать из описания новые высказывания, которым можно было бы

присвоить некоторую вероятность, обоснованную понятным способом старыми высказываниями. Для этого нужны правила вывода. Они могут быть более или менее формальными для каких-то областей, и они помогают как раз получить новые высказывания из старых.

Люди получают большинство бытовых моделей от других людей в процессе взросления, поэтому модели у людей немного разные, но в целом совместимы между собой.

Иногда бывает так, что людям надо совместить две модели одного кусочка реальности для того чтобы получить более полную картинку для совместной деятельности. Единственный полностью рабочий способ сделать это и быть уверенными, что договорились — это “стучать по столу”, то есть договориться на уровне 4D, что вы оба ожидаете или хотите от мира.

Байесианство и причинный вывод¹⁶

Как я уже писала выше, мы считаем, что все знание — вероятно, то есть вы никогда не можете быть уверены в чем-то на 100%, но и не можете про что-то думать, что этого не может быть, потому что не может быть никогда.

Здесь есть подвох: эта штука с вероятностями относится к *наблюдаемой реальности*, а не к согласованности внутри модели. Поэтому, скажем, утверждение о том, что в евклидовой геометрии параллельные прямые никогда не пересекаются — верно, потому что

¹⁶ <https://www.amazon.com/Book-Why-Science-Cause-Effect/dp/046509760X>

оно целиком про модель и внутреннюю согласованность модели, а не про мир.

Технически, даже про свершившиеся события мы можем говорить только в ключе будущих наблюдений за последствиями этого события, и язык для описания этих наблюдений — вероятностный.

Для того, чтобы вообще менять какие-то свои представления о реальности или хоть как-то обучаться и усваивать новую информацию, мы постоянно и по-бытовому делаем модели с причинно-следственными связями. То есть мы говорим “шел дождь, и поэтому дорожка мокрая”, для нас такие события становятся связанными, у нас в голове есть общее представление о том, что происходит, когда идет дождь и почему (все становится мокрым, потому что дождь — это вода). Мы к этому привыкли и даже не замечаем, а между тем, с этой способностью (строить причинно-следственные модели) связана другая человеческая способность, которая вообще позволяет нам что-то планировать: способность строить контрфактические суждения. Это значит, что мы способны вообразить и промоделировать то, чего нет. Нас не ставят в тупик вопросы типа: “а что было бы, если бы дождь не пошел? какие еще причины могут быть у того, что дорожка мокрая?”.

Так что прежде чем мы поговорим об обновлении моделей, нужно усвоить вот что: в модели нужно иметь связи между явлениями, в том

числе причинно-следственные, чтобы хоть сколько-нибудь правильно присваивать веса и вероятности.

Модели могут быть более или менее сложными, а еще они могут быть более или менее точными, и иногда бывает так, что это не особенно коррелирует с их предсказательной силой. Поэтому мы всегда ищем баланс между сложностью, точностью, и приростом предсказательной силы.

Часть 3. Обновление моделей

С помощью инструментов моделирования и понимания того, чего ради вы моделируете, вы можете построить довольно полезный кусочек карты, который будет хорошо вас водить по знакомому кусочку реальности.

Но мир берет и меняется. Что делать? Нужно обновлять убеждения и перестраивать модели. Сейчас подробно разберемся, как это делается.

Во-первых, как заметить, что пора обновить свои убеждения, реальность изменилась? Очень просто, вы **удивитесь** какому-то событию. Это и есть маркер того, что что-то в вашей картине мира не так, раз она не предсказывала этого события, а оно случилось. Вы можете удивиться настолько, что не захотеть признавать то, что вы увидели реальностью. Это плохо и про это можно почитать как про “предвзятость подтверждения”¹⁷.

Итак, маркер того, что вам пора обновить убеждения — возникшее чувство удивления, замешательства. Что-то вроде “ой, как так-то!”.

Во-вторых, нужно хорошо атрибутировать удивление. Это значит, что нужно сформулировать, что там было за удивление: к чему в мире оно относится. Это нужно формулировать в терминах описания наблюдений за реальность. Например “я удивлена, что идет дождь”.

¹⁷ https://ru.wikipedia.org/wiki/Склонность_к_подтверждению_своей_точки_зрения

Потом можно попытаться сформулировать, почему возникло удивление, то есть где была беда именно с моделью. “вчера было солнечно, поэтому я удивлена, что сегодня идет дождь”. Приходится немного сдвинуться в своих моделях: не всегда сегодня будет солнечно, если накануне было солнечно.

Технически, на этом моменте можно было бы представлять себе веер возможных миров, которые все похожи на наш актуальный, за исключением интересующего события и модели вокруг него.

Формула Байеса и зачем она нужна

Когда вы испытали удивление или замешательство, нужно посмотреть, какие события его вызвали, какие бы его **не** вызвали и как выглядит кусочек картины мира, который нужно перестроить.

Дальше процедура довольно простая: нужно взять событие, которое вас удивило и учесть его как свидетельство по Байесу (понять, какие у вас есть гипотезы о мире, которые следует обновить, и обновить их).

Формула такая:

Теорема Байеса

шансовая форма

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\bar{H})}$$

$$\text{Шансы: } O(X) = \frac{P(X)}{P(\bar{X})}$$

Сила свидетельства, коэффициент правдоподобия: $\frac{P(E|H)}{P(E|\bar{H})}$

- ❖ Шансы гипотезы H с учётом свидетельства E равны априорным шансам H , умноженной коэффициент правдоподобия свидетельства.
- ❖ Коэффициент правдоподобия — вероятность получения свидетельства E в мире, где гипотеза H истинна разделить на вероятность получения свидетельства E в мире, где гипотеза H ложна.

Смысл у нее довольно очевидный из здравого смысла: берете ваши предыдущие представления о реальности (априорные шансы события) и умножаете на *силу свидетельства*, то есть то, насколько новое событие резко отличает между собой мир, где гипотеза истинна от мира, где гипотеза ложна.

Здесь вот нужно еще одну важную штуку держать в голове.

Технически, все события между собой как-то да связаны, но иногда эта связь настолько далекая, что нет никакого смысла при наших естественных вычислительных ресурсах что-то считать. Поэтому надо

изначально определиться в модели, какие события связаны с какими, и как именно.

Это важный кусок, который никакая формула за вас не сделает: сказать, как *устроен* ваш мир. Между какими событиями есть причинно-следственные связи, между какими — другие отношения, и насколько они вообще друг на друга, эти события, влияют.

Ошибки в Байесинстве

Несмотря на то, что интуитивно люди и так работают как стихийные байесианцы, в эти вычисления постоянно вкрадываются фатальные ошибки:

1. Игнорирование априорной вероятности: когда получаете свидетельство, резко забываете, что вообще думали о том, как бы устроен этот кусочек мира раньше. Или, например, именно об этом кусочке мира вообще раньше не задумывались (но имели, разумеется, какие-то имплицитные представления; мало найдется кусочков, о которых вы *вообще* ничего бы не знали)¹⁸.
2. Игнорирование контрфактических суждений, которые про то, какова вероятность получить свидетельство в мире, где гипотеза ложна, или, проще говоря, "а если бы не".

Так что бытовой смысл формулы простой: проверьте все три компонента, прежде чем делать вывод.

¹⁸ https://en.wikipedia.org/wiki/Base_rate_fallacy

Очевидно, что вы не сможете постоянно все аккуратно считать по байесу. Но это и не требуется. Достаточно зашить себе эвристику вида “удивилась — пошла обновлять картину мира байесиански” и “удивилась — проверяю все три части формулы Байеса: я вообще их знаю?

Что дальше?

Дальше можно пройти по ссылкам и прочитать всю литературу, удовольствие интеллектуального характера наверняка получите.

Можно написать мне с комментариями по книжке на t.me/pionmedvedeva

После прочтения этой книжки можно идти изучать системное мышление¹⁹, прикладную рациональность²⁰ или принятие решений, это интересно и полезно в работе.

Можно ничего дальше не изучать, а просто применять полученные знания в работе и в быту.

¹⁹ <http://system-school.ru/>

²⁰ <https://kocherga-club.ru/workshop>