

## ПРЕДИСЛОВИЕ

Данная статья является переводом [Tutorial on Active Inference](#) Олега Солопчука. Так как тема Свободной Энергии является достаточно новой в научной среде, в этой статье присутствует много терминов, которые не очень популярны, с которыми русский читатель встречается впервые. В связи с чем очень тяжело перевести их на русский язык точно, потому что в русском языке такие словосочетания и слова не имеют определенного смысла. Но мы постарались перевести все максимально корректно и, более того, решили оставить некоторые термины в квадратных скобках на оригинальном (английском) языке для того, чтобы предоставить читателям возможность перевести самим, если их не устроит текущий перевод.

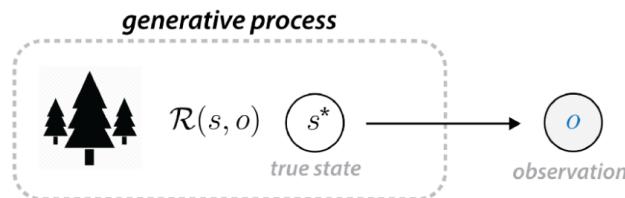
## СТАТЬЯ

Активный вывод [active inference] – это принцип Свободной Энергии мозга, применяемый к действию. Этот принцип относительно подтверждается экспериментами в области нейронаук и является популярной моделью работы мозга. Изначальная предпосылка активного вывода [initial inference]: агент (любая самоорганизующаяся система) хочет остаться в живых, поддерживая свой гомеостаз. В конце концов, агент должен следить за тем, чтобы важные для существования параметры (температура тела или насыщенность крови кислородом) не отклонялись сильно от нормы, т.е. чтобы не были неожиданными. Но поскольку эти параметры можно вывести только с помощью сенсорных измерений, агент минимизирует неожиданность [surprise] наблюдений, полученных с помощью сенсоров. Интересно то, что данная задача схожа с задачей продолжительного улучшения модели мира агента (убедимся в этом позже). Рассмотрим вышеописанную схему вкратце:

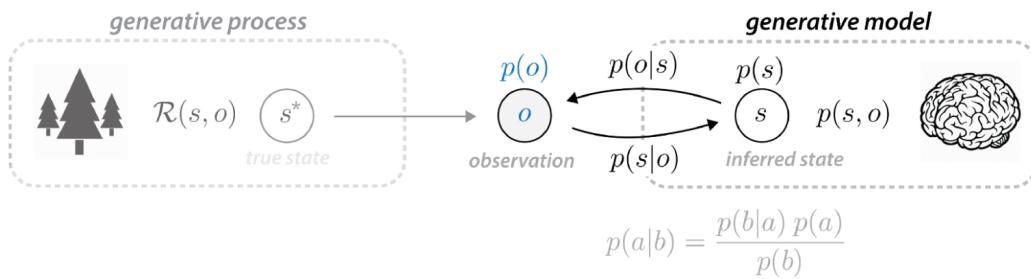
Остаться в живых  $\Rightarrow$  поддерживать гомеостазис  $\Rightarrow$  избегать неожиданных состояний  $\Rightarrow$  избегать неожиданных наблюдений  $\Rightarrow$  минимизация приближения к неожиданностям (свободной энергии)

И если [предыдущий пост](#) нацелен был на то, чтобы дать интуицию Свободной энергии [Free Energy], то здесь нам придется испачкать ручки. Никакого технического бэкграунда не нужно, за исключением [Теории вероятностей](#) и [Теоремы Байеса](#). Идея заключается в следующем.

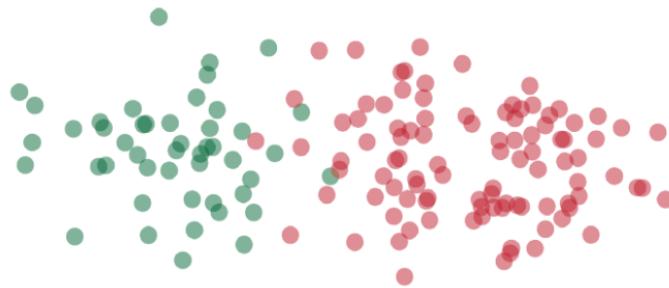
Мозг избегает неожиданностей [surprise], имея хорошую модель окружающей среды. С одной стороны, окружение имеет истинную, скрытую для агента, случайную величину  $s$  (называемую состояние), которая генерирует некоторые наблюдения [probabilistic observations]  $o$ . Важно подчеркнуть, что состояние  $s$  скрытое означает, что мы можем наблюдать только  $o$ . Например, шел дождь ночью ( $s$ ), значит трава влажная утром ( $o$ ). Это называется генеративным процессом [generative process]  $R(s,o)$ .



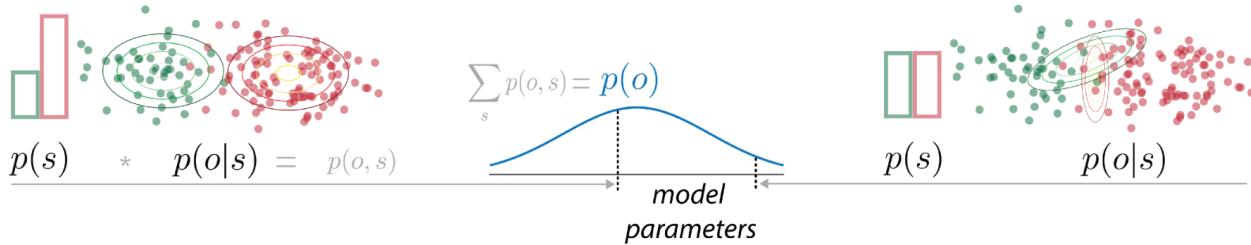
С другой стороны, мозг пытается сделать заключение [infer] о вероятности разных скрытых состояний, учитывая наблюдения,  $p(s|o)$ . И делает он это через априорные знания [prior belief]  $p(s)$  и правдоподобие [likelihood]  $p(o|s)$ . Таким образом, мозг строит генеративный процесс [generative process], определяемый как совместное распределение [as a joint]  $p(s,o)$ .



Давайте рассмотрим пример. Представим следующий генеративный процесс [generative process] – в нашем саду яблочное и апельсиновое дерево. Обозначим за скрытую переменную  $s$  – является фрукт апельсином или яблоком. Предположим, что яблочное дерево немного левее апельсинового дерева. Итак, когда фрукты падают на землю, мы наблюдаем следующее:

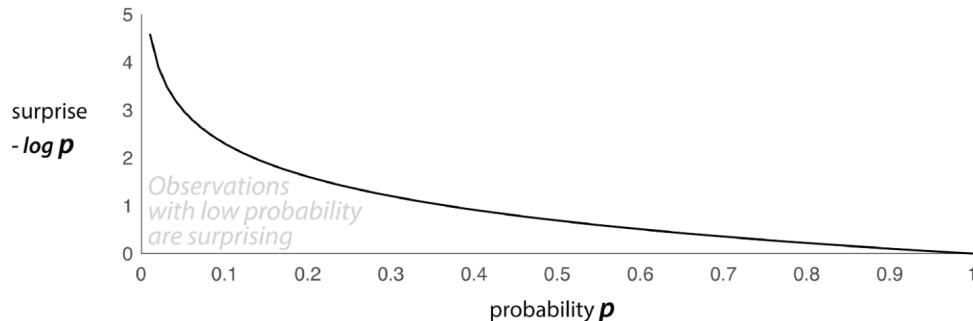


Кажется, что 70% лежащих фруктов – апельсины, а остальные 30% – яблоки. Это истинное распределение скрытого состояния  $s$ . Когда фрукты падают, они не попадают в одно и тоже местоположение, они случайно распределяются на земле. Это уже вероятность наблюдении при условии  $s$ . **Вывод** [Inference] в генеративной модели [generative model] заключается в нахождении апостериорного распределения  $p(s|o)$  – вероятность того, что фрукт является яблоком, учитывая его местоположение. **Обучение** генеративной модели [generative model] состоит из оценивания параметров распределения (таких как, например, математическое ожидание и дисперсия для нормального распределения) скрытого состояния  $p(s)$  и  $p(o|s)$ , сопоставляющего состояния и наблюдения. Вывод  $p(s|o)$  через теорему Байеса требует от нас подсчета вероятности  $p(o)$ , которая интересна сама по себе, так как чем лучше наша модель, тем выше будет вероятность наблюдаемых данных  $p(o)$ . Это также называется 1) «обоснованность модели» [«model evidence»], так как это учитывает, насколько хорошо наша модель предсказывает реальные данные, 2) «функция предельного правдоподобия» [«marginal likelihood»], потому что мы исключаем скрытое состояние  $s$  путем интегрирования. Сравним две модели внизу, которые оценивают  $p(o,s)$ : модель слева, очевидно, лучше –  $p(s)$  корректно показывает, что апельсинов больше, чем яблок, и  $p(o|s)$  хорошо центрирован по обоим кластерам. Мы можем количественно оценить качество каждой модели с помощью модели  $p(o)$ , проинтегрировав по скрытой переменной  $s$ :

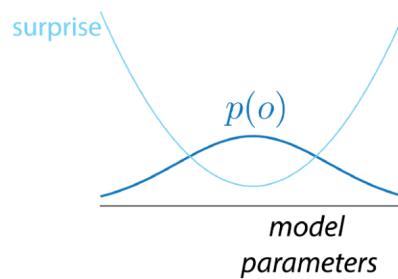


В идеале, мы хотим выбрать такие параметры модели, которые будут вести к наилучшей обоснованности модели  $p(o)$ . Как это связано с активным выводом [active inference] и с агентом, который избегает неожиданных наблюдений? По факту, задача максимизации «обоснованности модели» [«model

**evidence»]** эквивалентна **минимизации неожиданности**, которая равна всего лишь отрицательному логарифму  $p(o)$ . Если вероятность равна 1 – неожиданность [surprise] равна 0, если вероятность равна 0 – неожиданность стремится к бесконечности. Здесь неожиданность представлена, как функция от вероятности:



Здесь неожиданность [surprise],  $-\log p(o)$ , наложенная на обоснованность модели [model evidence]  $p(o)$ :



Как мы обсуждали выше для обоснованности модели [model evidence], чтобы оценить неожиданность  $-\log p(o)$ , нам нужно просуммировать по скрытой переменной  $s$  совместное распределение  $p(o, s)$ :

$$-\log p(o) = -\log \sum_s p(o, s)$$

$$p(a) = \sum_b p(a, b)$$

Так как речь идет о суммировании по всем возможным значениям  $s$ , суммирование может стать очень тяжелым занятием, если будет очень много различных значений  $s$ . Однако есть трюк, который поможет вам избежать невозможного суммирования.

**Вместо того, чтобы считать неожиданность напрямую, мы можем аппроксимировать ее** чем-то очень схожим и легко расчитываемым. Сперва, мы представляем простое распределение [dummy distribution]  $q$  с областью значений  $s$  (которая окажется безумно полезной позже). Мы можем спокойно внести ее в сумму через деление и умножение одновременно:

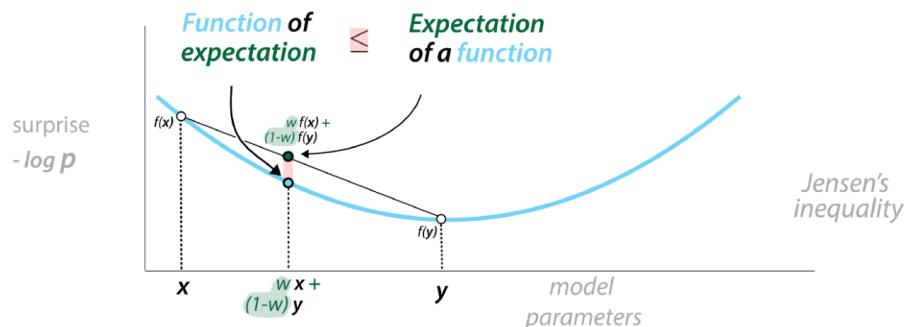
$$-\log \sum_s p(o, s) = -\log \sum_s q(s) \frac{p(o, s)}{q(s)}$$

*multiply by 1*

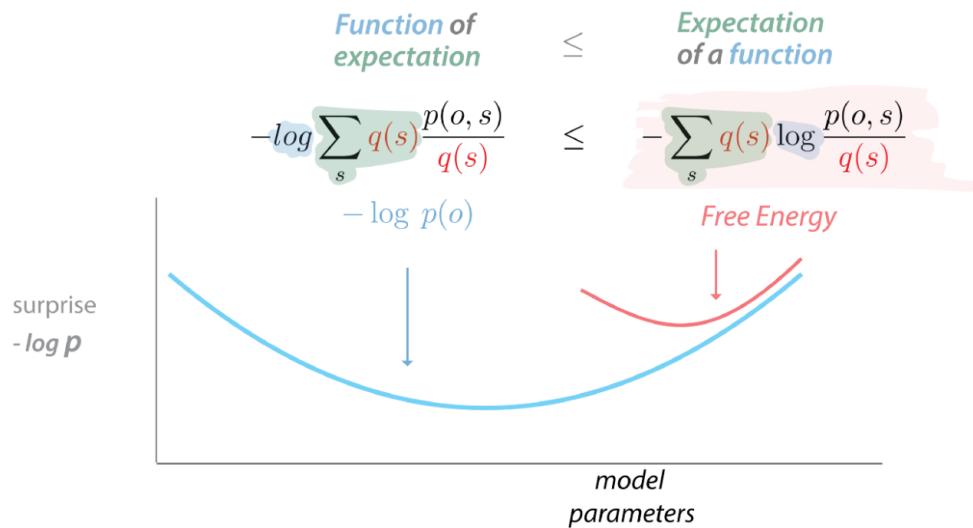
Это дает нам взвешенную сумму, в которой для каждой  $s$ , отношение  $p(o, s)/q(s)$  суммируется с весом  $q(s)$ . Это все еще та же неожиданность,  $-\log p(o)$ , но теперь нам очень удобно заменить ее аппроксимацией.

Теперь же, давайте отвлечемся от неожиданности на секунду и сфокусируемся на аппроксимации. Мы знаем, что функция неожиданности  $-\log$  выглядит как впадина или чаша (иными словами

выпуклая). Определение гласит: функция выпуклая в том случае, если когда вы уроните на нее палку, палочка окажется в функции, как в миске. Формально мы объясняем это следующим образом. Давайте возьмем две точки на оси абсцисс ( $x$  и  $y$ ), и выберем число между 0 и 1 ( $w$ ). Как показано ниже, мы можем двигаться между  $x$  и  $y$  с помощью их взвешенной суммы через изменение  $w$  от 0 до 1. То есть мы могли бы рассматривать  $w$  и  $(1-w)$  как параметры обычного распределения [simple distribution]. На самом деле, есть краткое название для «взвешенных сумм, в которых веса определяются распределением» – математическое ожидание. Теперь вернемся к тетиве в луке: если вы оцениваете функцию этого ожидания, она всегда будет ниже или равна ожиданию функции, оцененной в  $x$  и  $y$ .



Поскольку ранее мы обозначили неожиданность, как функцию ( $-\log$ ) от математического ожидания (взвешенной суммы отношения  $p(o, s)/q(s)$ ), то неравенство для выпуклых функций выглядит подходящим для нашего приближения! Формально, это называется «верхняя граница». Эта самая верхняя граница и есть свободная энергия.



Мы можем пойти на шаг дальше и убрать минус перед логарифмом. Это приведет нас к определению, которое используют в статьях про Активный Вывод [Active Inference]:

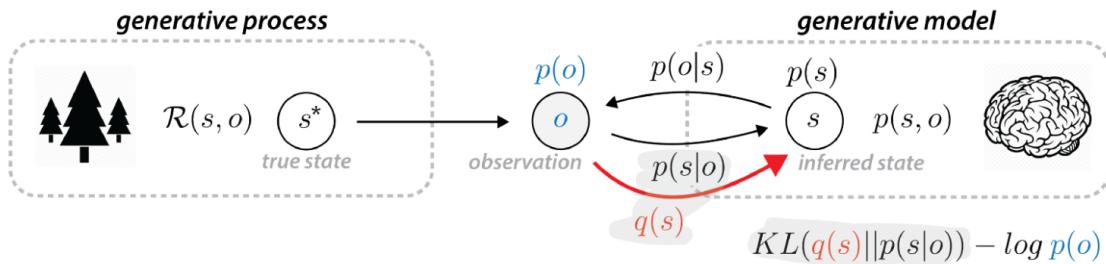
$$-\log(a) = \log\left(\frac{1}{a}\right)$$

$$-\sum_s q(s) \log \frac{p(o, s)}{q(s)} = \sum_s q(s) \log \frac{q(s)}{p(o, s)}$$

Крутой момент в Свободной Энергии заключается в том, что веса в суммировании определяются  $q(s)$ , и мы имеем полный контроль над  $q(s)$ . Получается, что мы можем изменять  $q(s)$  таким образом, чтобы минимизировать свободную энергию.

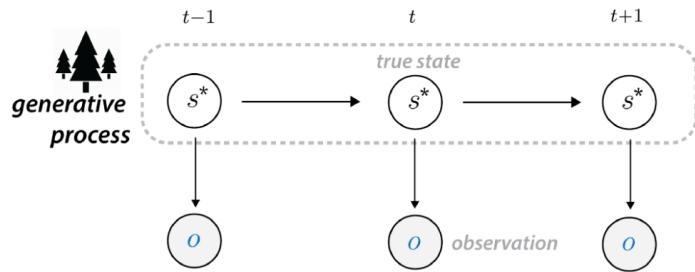
Используя пару стандартных тождеств, Свободная Энергия (обычно добавляемая в литературе как «Variational») может быть разложена двумя эквивалентными способами.

Хотя правая ветвь обычно используется на практике, давайте сосредоточимся на левой, так как она дает нам хорошее теоретическое понимание [примечание: мы можем убрать ожидание (сумму по  $s$  из  $q(s)$ ) перед  $-\log p(o)$ , поскольку  $p(o)$  не зависит от  $s$ , а  $q(s)$  как распределение вероятностей суммируется до 1]. Это говорит о том, что свободная энергия равна дивергенции  $KL$  (Кульбака - Лейблера) между  $q(s)$  и  $p(s|o)$ , и сюрпризу  $-\log p(o)$ .

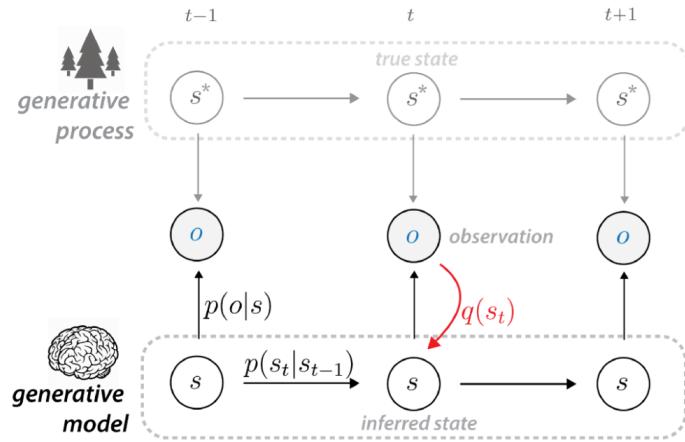


Таким образом, минимизируя свободную энергию, произвольное распределение  $q(s)$  становится ближе к апостериорному  $p(s|o)$ , и если они совпадают, дивергенция  $KL$  становится равной 0, а свободная энергия точно равна неожиданности [surprise]. Таким образом, чем больше мы минимизируем свободную энергию путем подгонки  $q(s)$ , тем ближе она становится к неожиданности (поскольку это верхняя граница), а минимизируя свободную энергию путем изменения параметров  $p(o,s)$ , мы можем минимизировать неожиданность еще больше. Это подробно показано в [посте о прогнозирующем кодировании](#) [pedictive coding], которое является следствием принципа свободной энергии, применяемого к восприятию. Вот краткое изложение на данный момент: минимизируем аппроксимацию неожиданности (свободную энергию)  $\Rightarrow$  избегаем неожиданных наблюдений  $\Rightarrow$  избегаем неожиданных состояний  $\Rightarrow$  поддерживаем гомеостазис  $\Rightarrow$  остаемся живыми.

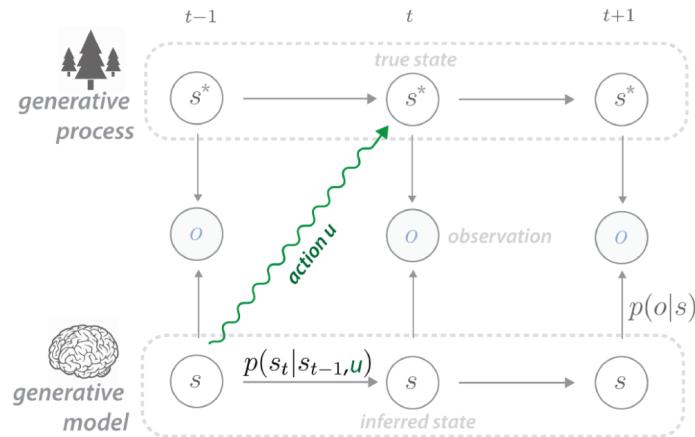
До сих пор мы имели дело со статической ситуацией, с одним скрытым состоянием и одним набором наблюдений, но реальный мир динамичен. У нас есть скрытое состояние в каждый момент времени, и, поскольку события в мире имеют тенденцию **зависеть от того, что произошло раньше**, мы будем предполагать, что  $s$  в определенный момент времени  $t$  зависит от  $s$  в предыдущий момент времени. Например, вероятность увидеть радугу напрямую зависит от того, шел ли дождь раньше.



Как и прежде, мы пытаемся смоделировать истинный генерируемый процесс [true generative process], изучая генеративную модель [generative model]  $p(o,s)$  и получая приближение к апостериорным  $q(s)$  на каждом временном шаге. Как и в простой статической ситуации, мы можем найти, каким образом нужно изменить параметры  $p(o,s)$  и  $q(s)$ , чтобы уменьшить свободную энергию, а затем сделать много маленьких шагов в этом направлении.

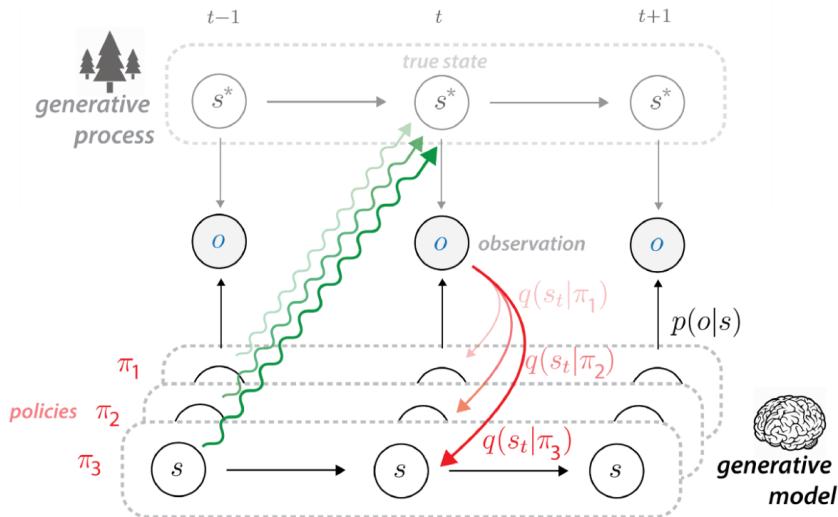


Это будет работать, но мы просто будем пассивно наблюдать за окружающей средой. Что если мы тоже будем воздействовать на нее? В этом случае  $s$  в определенный момент времени  $t$  будет зависеть от  $s$  в предыдущий момент времени и нашего действия  $u$ . Другими словами, действие может напрямую влиять на состояние мира, поэтому другое действие может привести к иному будущему (например, мы можем физически двигать вещи своими действиями).



Теперь вывод становится активным [inference becomes active]! Нам просто нужно найти способ выбрать хорошее действие на каждом временном шаге. В действительности кажется, что мы не рассматриваем последствия действий только на следующем временном шаге. Мы планируем всю политику, серию действий, ориентированных на отдаленные во времени цели. Поэтому, если есть много возможных действий и много будущих моментов времени  $\Rightarrow$  есть много потенциальных политик, которые мы можем пред-

принять. Активный вывод говорит: просто рассмотрите все варианты. Таким образом, мы делаем вывод, аппроксимируя  $p(s|o)$  с  $q(s)$  одновременно (параллельно) для каждой возможной политики  $\pi$ .



И поскольку наша цель остается прежней – минимизировать неожиданность путем минимизации свободной энергии – мы можем вычислить свободную энергию (и направление изменения  $q(s)$ , которое её минимизирует) при каждом возможном шаге политики и времени.

**Политики [Policies], которые минимизируют свободную энергию в будущем, являются предпочтительными.** Допустим, мы планируем на 10 временных шагов вперед. Если мы находимся в момент времени  $t = 1$ , для каждой политики мы суммируем свободные энергии для шагов с 1 по 10 времени и выбираем политику, которая имеет минимальную кумулятивную Свободную энергию в будущем. Давайте еще раз посмотрим на общую картину того, как можно рассчитать свободную энергию:

*Variational Free Energy F*

$$F = \sum_s q(s) \log \frac{q(s)}{p(o, s)}$$

$p(a, b) =$   
 $p(a|b) p(b) =$   
 $p(b|a) p(a)$

$= \sum_s q(s) \log \frac{q(s)}{p(s|o) p(o)}$ 
 $= \sum_s q(s) \log \frac{q(s)}{p(s) p(o|s)}$

$$KL(p(a)||q(a)) = \sum_a p(a) \log \frac{p(a)}{q(a)}$$

$$\log(a/b) = \log(a) - \log(b)$$

$$= KL(q(s)||p(s|o)) - \log p(o)$$

$$= KL(q(s)||p(s)) - \sum_s q(s) \log p(o|s)$$

Левая ветвь показывает нам важные теоретические свойства минимизации свободной энергии (например, что  $q(s)$  приближается к  $p(s|o)$ ), но это нецелесообразно, поскольку мы используем свободную энергию для аппроксимации  $-\log p(o)$  в первую очередь. Итак, давайте посмотрим на правую ветвь, которая является стандартным способом вычисления свободной энергии:

$$= KL(q(s) || p(s)) - \sum_s q(s) \log p(o|s)$$

*Complexity*                            *Accuracy*

Эти два термина обычно называют сложностью [complexity] и точностью [accuracy]. Сложность показывает, насколько аппроксимация апостериорного  $q(s)$  отклоняется от априорного  $p(s)$ , и количественно определяет, сколько дополнительных битов информации, которых нет в априорном распределении, мы хотим закодировать в приблизительном апостериорном  $q(s)$ . Точность (ожидаемое значение  $p(o|s)$ ) оценивает вероятность того, что состояниям дан конкретный результат  $\mathbf{o}$ . Хотя мы можем легко вычислить сложность, существует проблема в оценке точности для будущих временных шагов – просто потому, что мы еще не наблюдали результаты. Таким образом, нам нужен другой способ вычисления свободной энергии, поэтому мы пойдем по левой ветви её разложения. Но здесь есть одна загвоздка: поскольку у нас нет доступа к будущим наблюдениям – нам нужно угадать, как они могли бы выглядеть, и взять взвешенную сумму свободной энергии по догадкам  $p(o|s)$ . Обратите внимание, что вероятности будут при условии  $\pi$ , так как свободная энергия оценивается отдельно для каждой политики. Предполагается, что только отображение состояния на наблюдения  $p(o|s)$  одинаково для всех политик.

*Expected Free Energy G*

$$\sum_s q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(o_t, s_t|\pi)}$$

$$= \sum_s q(s_t|\pi) \sum_o p(o_t|s_t) \log \frac{q(s_t|\pi)}{p(o_t, s_t|\pi)}$$

Итак, нам нужно **предсказывать как будущие состояния  $q(s|\pi)$ , так и наблюдения**. А также оценивать свободную энергию на их основе. Давайте пройдем шаг за шагом, сначала сосредоточившись на знаменателе  $p(o,s|\pi)$ . Рассматривая левую ветвь разложения свободной энергии, показанную выше, мы можем преобразовать ее так же, как  $p(s|o) \cdot p(o)$ :

$$= \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(o_t, s_t|\pi)} \quad p(a,b) = p(b|a) p(a)$$

$$= \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(s_t|o_t, \pi) p(\mathbf{o}_t)} \leftarrow \begin{array}{l} \text{prior preferences} \\ \text{on future outcomes} \end{array}$$

$p(o)$  – это априорные предпочтения относительно будущих наблюдений [the prior preference on the future outcomes] (которые пропорциональны вознаграждению в классическом обучении с подкреплением). Разделив логарифм, мы получаем следующие две величины: отрицательное эпистемическое значение [negative epistemic value] и ожидаемое предшествующее предпочтение [expected prior preference]:

$$\begin{aligned}
 & -\text{Epistemic value} \\
 &= \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)} - \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log p(o_t)
 \end{aligned}$$

Короче говоря, эпистемическое значение [epistemic value] говорит нам, насколько будущие наблюдения могут уменьшить нашу неопределенность относительно аппроксимации апостериорного  $q(s)$  [the approximate posterior  $q(s)$ ]. Давайте сначала закончим вывод, а затем обсудим его подробно. У нас есть истинное апостериорное  $p(s|o, \pi)$  в знаменателе левого слагаемого, которое, как мы знаем, трудно вычислить, особенно в будущем (имейте в виду, что мы предсказываем будущие состояния и наблюдения). Таким образом, мы могли бы применить формулу Байеса, чтобы повторно выразить это соотношение в более вычислимых выражениях:

$$\begin{aligned}
 & -\text{Epistemic value} \\
 &= \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)} - \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log p(o_t) \\
 &\quad \boxed{\text{Bayes rule} \quad \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)} = \frac{q(s_t|\pi) q(o_t|\pi)}{p(o_t|s_t, \pi) q(s_t|\pi)} = \frac{q(o_t|\pi)}{p(o_t|s_t, \pi)}} \quad p(a|b) = \frac{p(b|a) p(a)}{p(b)} \\
 &= \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(o_t|\pi)}{p(o_t|s_t, \pi)} - \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log p(o_t) \\
 & \quad -\text{Epistemic value}
 \end{aligned}$$

Это было бы легче сделать, если бы мы пренебрегли зависимостью от  $\pi$ . Тогда  $q(o)$  это вроде функции предельного правдоподобия [marginal likelihood],  $p(o|s)$  – вероятность [likelihood] и  $q(s_t|\pi)$  – априорное распределение. Вы также можете заметить, что для некоторых распределений « $p$ » заменяется на « $q$ ». Это результат аппроксимации, так как, например,  $p(s|o, \pi)$  является истинным апостериорным распределением, которое мы можем аппроксимировать с помощью  $q(s|o, \pi)$ , что при разложении приведет к тому, что все компоненты формулы Байеса будут формой  $q$ . Также обратите внимание, что  $q(o|s, \pi)$  это то же самое, что и  $p(o|s, \pi)$ , поскольку это все еще относится к той же вероятности [likelihood].

Мы также могли бы объединить логарифмы и снова разделить их по-другому, что приводит нас к окончательной форме ожидаемой свободной энергии:

$$\begin{aligned}
 &= \sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(o_t|\pi)}{p(\mathbf{o}_t) p(o_t|s_t, \pi)} \quad \leftarrow \text{predicted outcomes} \\
 &= \sum_{o,s} q(s_t|\pi) p(o_t|s_t) \log \frac{q(o_t|\pi)}{p(\mathbf{o}_t)} - \sum_s q(s_t|\pi) \sum_o p(o_t|s_t) \log p(o_t|s_t) \\
 KL(p(a)||q(a)) &= \sum_a p(a) \log \frac{p(a)}{q(a)} \quad H[p(a)] = - \sum_a p(a) \log p(a) \\
 &= KL(q(o_t|\pi)||p(\mathbf{o}_t)) + \sum_s q(s_t|\pi) H[p(o_t|s_t)] \\
 &\quad \text{Expected cost} \quad \quad \quad \text{Expected Ambiguity}
 \end{aligned}$$

Примечание: на левой стороне мы можем повторно выразить  $q(s|\pi)p(o|s)$  как совместное распределение  $q(s,o)$  и просуммировать по всем  $s$ , чтобы получить  $q(o)$ . Это даст нам ожидаемую дивергенцию Кульбака-Лейблера. А на правой стороне мы можем удалить  $\pi$  из  $p(o|s,\pi)$ , потому что вероятность [likelihood] одинакова для каждой политики.

Левое слагаемое (называемое "Издержки" или "Риск") – это дивергенция Кульбака-Лейблера между двумя распределениями: ожидаемыми в рамках политики  $\pi$  наблюдениями  $q(o|\pi)$  и априорными предпочтениями [prior preferences]. Таким образом, минимизация ожидаемой свободной энергии будет способствовать политике, которая приведет к наблюдениям, которые нам нравятся. И правое слагаемое, неоднозначность [Ambiguity], количественно определяет, насколько неопределенным является отображение между состоянием и наблюдениями  $p(o|s)$ . И это окончательная формула свободной энергии в будущем. Давайте теперь посмотрим на эпистемическое значение (в формуле выше, окрашено в желтый цвет).

**Эпистемическое значение говорит нам, как много мы могли бы извлечь из окружающей среды, если бы следовали этой политике.** Так происходит потому, что эпистемическое значение представляет собой взаимную информацию [mutual information] между скрытыми состояниями  $s$  и ожидаемыми наблюдениями  $o$ . Взаимная информация количественно определяет, насколько неопределенность ( $H$ ) по одной переменной уменьшается, если мы знаем другую.

$$\begin{aligned}
 MI(a,b) &= H[p(a)] - H[p(a|b)] \quad H[p(a)] = - \sum_a p(a) \log p(a) \\
 &= H[p(b)] - H[p(b|a)]
 \end{aligned}$$

Аналогично, взаимная информация может быть повторно выражена как дивергенция Кульбака-Лейблера между совместным распределением двух переменных (если взаимная информация велика, то знание одной переменной говорит нам много о распределении другой) и произведением их маргинальных плотностей [marginals] (как если бы они были полностью независимы). Нам просто нужно перевернуть дробь (потому что эпистемическое значение отрицательно в уравнениях). Обратите внимание, что мы можем удалить  $\pi$  из  $p(o|s,\pi)$ , потому что вероятность одинакова для каждой политики.

-Epistemic value

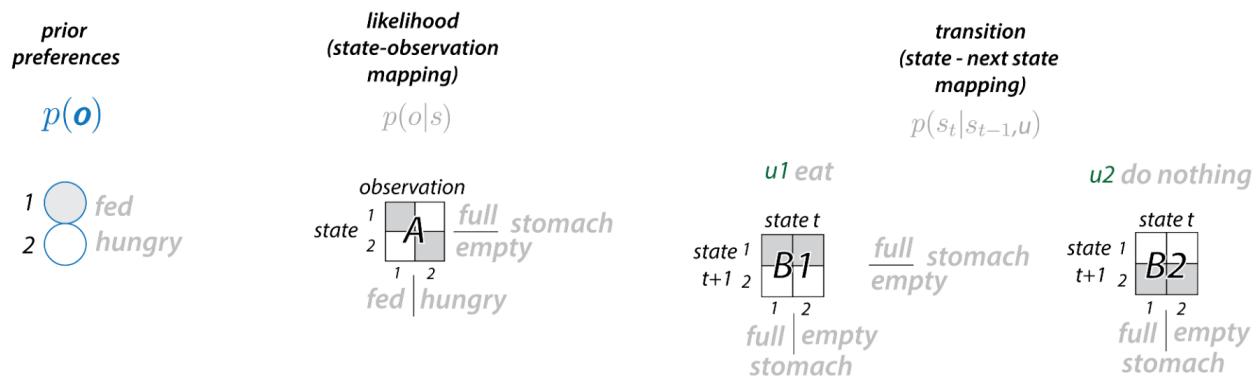
$$\sum_{o,s} p(o_t|s_t) q(s_t|\pi) \log \frac{q(o_t|\pi)}{p(o_t|s_t, \pi)}$$

$$\begin{aligned} MI(a, b) &= \sum_{a,b} p(a, b) \log \frac{p(a, b)}{p(a) p(b)} \\ &\quad \text{Mutual information} \\ &= \sum_{a,b} p(a|b) p(b) \log \frac{p(a|b) p(b)}{p(a) p(b)} = \sum_{a,b} p(a|b) p(b) \log \frac{p(a|b)}{p(a)} \end{aligned}$$

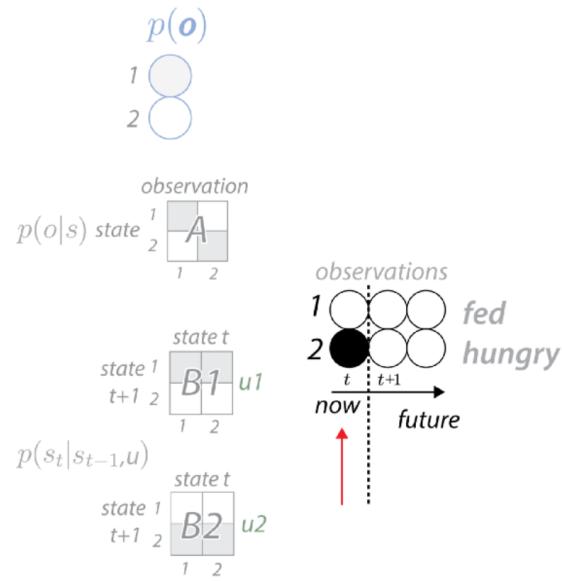
$$-\log(a) = \log\left(\frac{1}{a}\right)$$

На практике эпистемическое значение зависит от неопределенности относительно будущих состояний  $q(s|\pi)$ . Если вы абсолютно уверены, тогда  $H[q(s|\pi)]$  невелико, вам больше нечего учиться, поэтому эпистемическая ценность будет низкой. Но если вы не уверены,  $H[q(s|\pi)]$  высоко, и существует сильная зависимость между состояниями и наблюдениями (поэтому  $H[q(s|o)]$  низок), то взаимная информация будет высокой (см. формулы взаимной информации в терминах энтропий выше). Надеюсь, что прочитав приведенным описаниям несколько раз и пройдя их самостоятельно, оно станет простым.

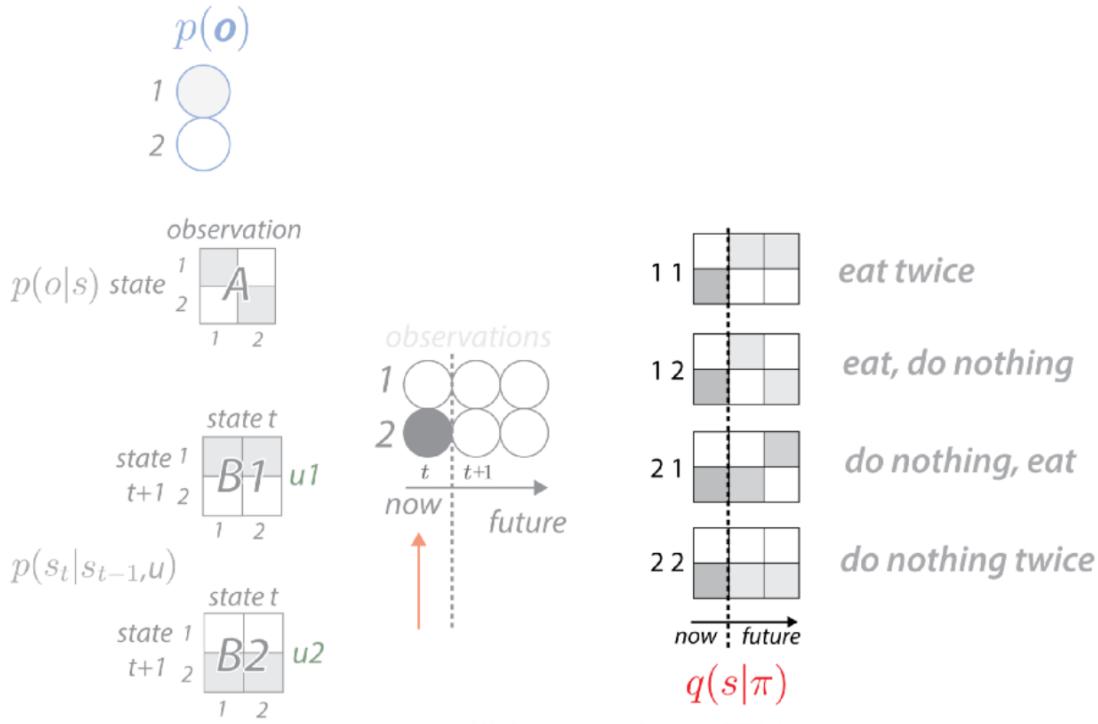
Вот небольшой пример того, что на самом деле произошло бы в мозгу агента, если бы он использовал активное умозаключение [active inference]. Для простоты предположим, что окружающая среда имеет только два состояния  $s$ : «1» и «2», например, есть пища в вашем желудке (1) или нет (2). Аналогично, есть только два возможных наблюдения : «1» и «2», вы чувствуете себя сытым (1) или голодным (2). Предположим, что мы уже знаем параметры генеративной модели [generative model]  $p(o,s)$ . Вероятность (называемая матрицей  $A$ )  $p(o|s)$  сопоставляет состояния с наблюдениями – если у вас есть еда – вас кормят и наоборот. Вероятность перехода  $p(s_t|s_{t-1}, u)$  отображает предыдущее состояние в следующее. Но поскольку переход также зависит от действия  $u$ , мы можем выразить его в виде отдельной матрицы переходов ( $B$ ) для каждого действия. Предположим, что мы можем либо пойти за едой ( $u_1$ ), либо ничего не делать ( $u_2$ ). Так что если мы выберем  $u_1$  – у нас будет еда в следующем состоянии, независимо от того, есть ли она у нас сейчас, и наоборот. Наконец, у нас также есть предварительные предпочтения  $p(o)$  – мы любим, чтобы нас кормили и не голодали, поэтому мы приписываем более высокую вероятность наблюдению «1» – кормили. Другими словами, Мы выражаем предпочтения по отношению к наблюдениям как вероятность  $p(o)$ .



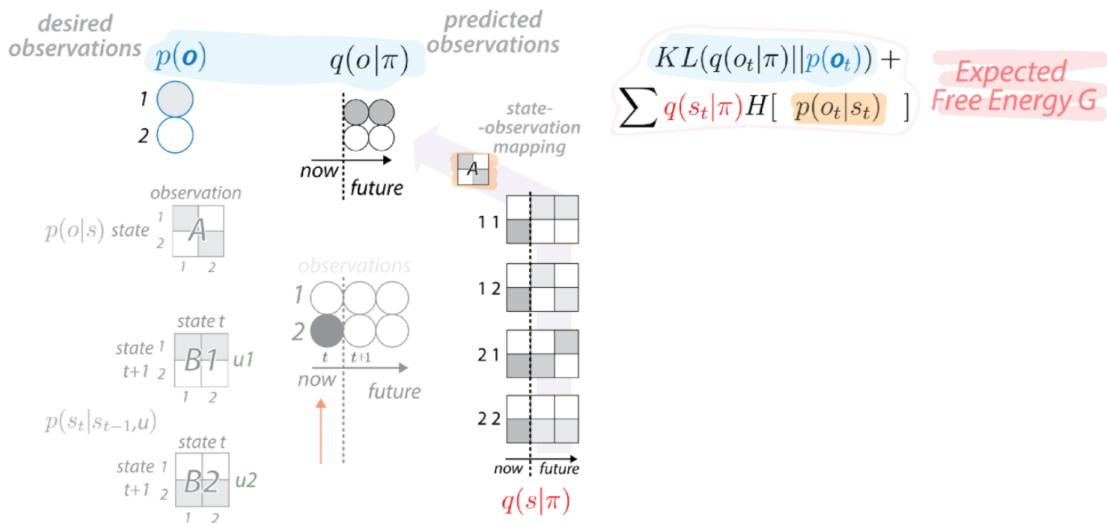
Теперь представьте, что мы наблюдаем, что мы голодны, и должны планировать свои действия на два шага вперед.



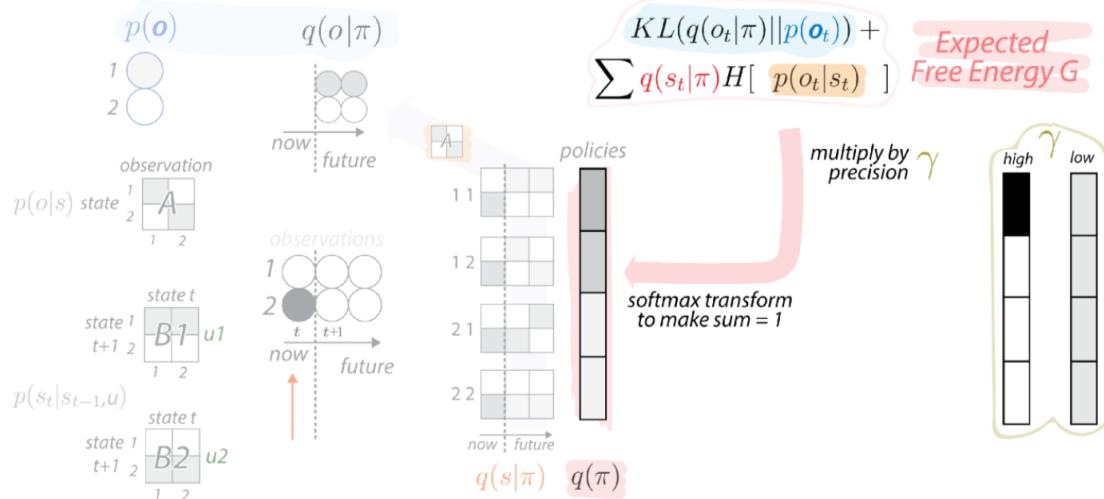
Поскольку есть только два возможных действия и два шага, мы можем оценить все возможные политики: 1 – 1 (идти за едой оба раза), 1 – 2, 2 – 1, 2 – 2. На самом деле мы будем оценивать апостериорное распределение по будущим состояниям (будет ли пища находиться в нашем желудке), при каждой из этих политик:



Поскольку мы знаем, как состояния связаны с наблюдениями ( $p(o|s)$ , матрица  $A$ ), мы можем оценить прогнозируемое наблюдение для каждой политики  $q(o|\pi)$  и вычислить дивергенцию Кульбака-Лейблера (заштрихована синим цветом) – левое слагаемое ожидаемой свободной энергии. Аналогично, мы также можем оценить неоднозначность [Ambiguity] (заштрихована оранжевым цветом), которая зависит от  $p(o|s)$  – правое слагаемое:

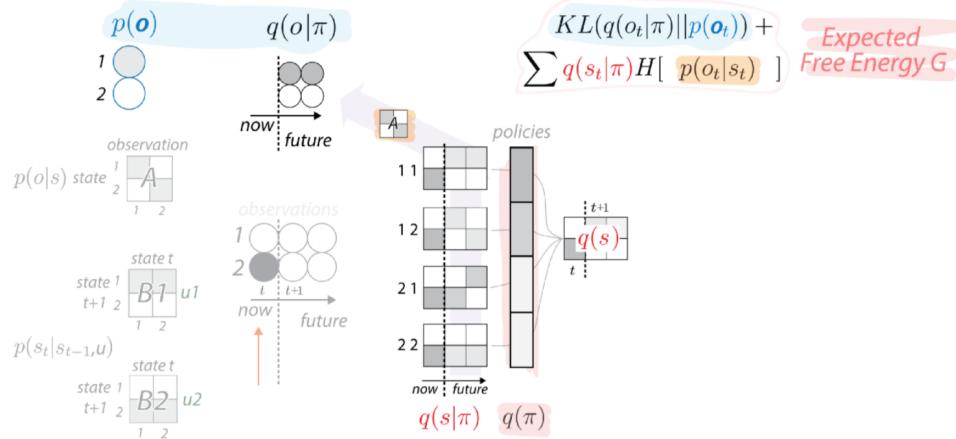


Затем мы суммируем ожидаемую свободную энергию по будущим действиям и преобразуем ее в распределение вероятностей для политики [probability distribution over policy]  $q(\pi)$ . Так что чем меньше свободная энергия, тем выше вероятность проведения политики. Интересно, что при преобразовании свободная энергия взвешивается (умножается) на точность [precision], которая определяет, насколько мы уверены в своих убеждениях по отношению к политике (то есть, изменяя точность до ее крайних значений, наши убеждения могут концентрироваться на одной политике или распределяться равномерно на несколько). Это важно для определения направления исследования/использования, так как чем больше вы уверены в том, что у вас есть хорошая политика (т. е. высокая точность), тем меньше вы исследуете и наоборот.

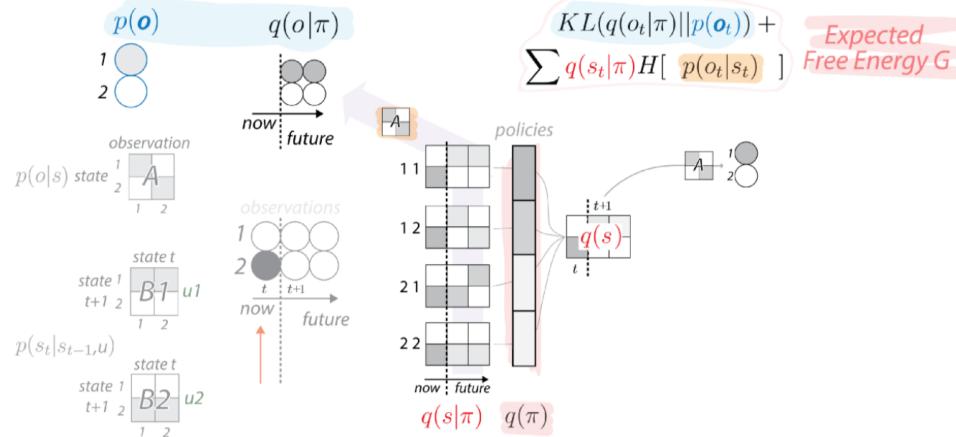


Теперь вы можете просто выбрать политику, которая максимально минимизирует свободную энергию, но есть более аккуратный способ: вместо выбора одной политики мы будем делать усреднение по ним. Подумайте об этом так, как если бы в определенный момент времени существовала одна политика, которая давала бы самую низкую свободную энергию, а другие политики были бы просто немного хуже. По мере того как вы будете получать новые наблюдения с течением времени, может оказаться, что лучшая в мире политика была среди тех, которые были «немного хуже» раньше. Но что делать, если вы уже выбрали действие, которое не позволяет вам следовать этой недавно обнаруженной политике? Таким образом, мы позволяем каждой политике голосовать во время выбора действий, причем политики, которые максимально минимизируют свободную энергию, имеют наибольший вес. Мы берем математическое ожидание  $q(s|\pi)$  при известном  $q(\pi)$  [expectation of  $q(s|\pi)$  under  $q(\pi)$ ] – взвешенная сумма, где

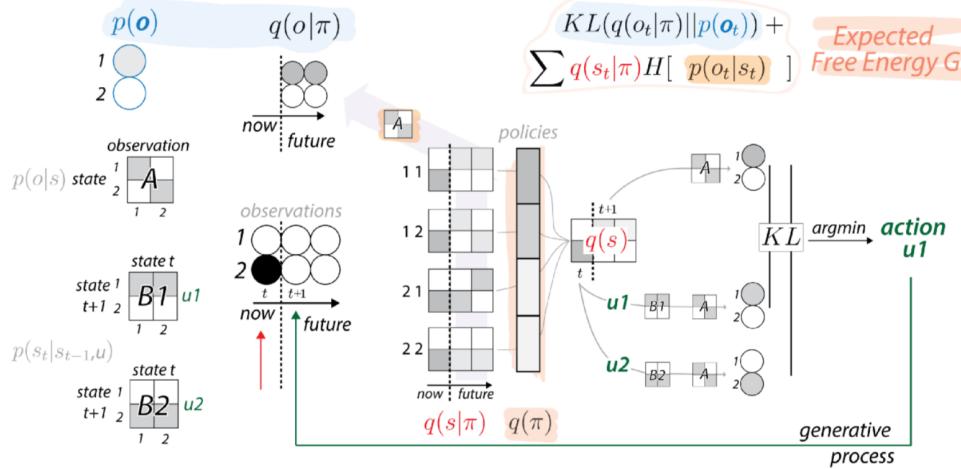
веса определяются вероятностью каждой политики. Это приводит к маргинальному распределению  $q(s)$ , которое неявно включает политику (и, следовательно, ожидаемую свободную энергию).



И вот теперь мы готовы действовать! Мы выбираем действие, которое минимизирует разницу (дивергенцию Кульбака-Лейблера) между тем, что мы ожидаем увидеть на следующем временном шаге, и тем, что мы должны были бы увидеть, если бы выбрали конкретное действие. Сперва чтобы получить (вероятность) ожидаемых наблюдений, мы умножаем наше убеждение о скрытом состоянии на следующем временном шаге  $q(s_{t+1})$  на  $p(o|s)$ , сопоставляющее состояния и наблюдения (матрица  $A$ ) [by the state-observation mapping  $p(o|s)$  («A» matrix)] :

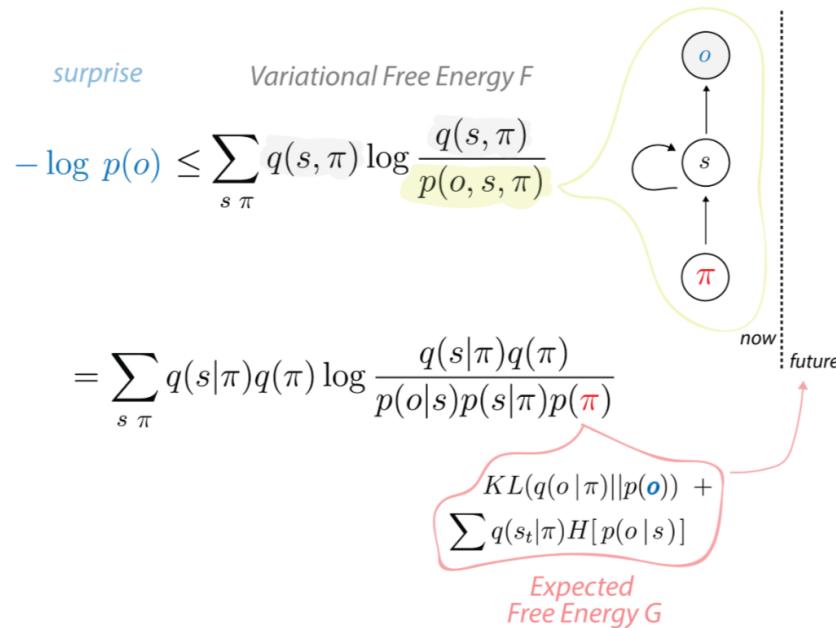


И чтобы получить ожидаемые наблюдения, которые мы увидели бы, если бы предприняли определенные действия, мы умножаем нашу веру в скрытое состояние в текущий момент времени  $q(s_t)$  на переходную матрицу  $B$  для данного действия, чтобы получить гипотетическое следующее состояние, а затем на матрицу  $A$ , сопоставляющую состояния и наблюдения, чтобы получить гипотетическое наблюдение. Выполняется действие, минимизирующее дивергенцию Кульбака-Лейблера, генеративный процесс возвращает нам следующее наблюдение в момент времени  $t + 1$ , и процесс начинается снова. И это все.



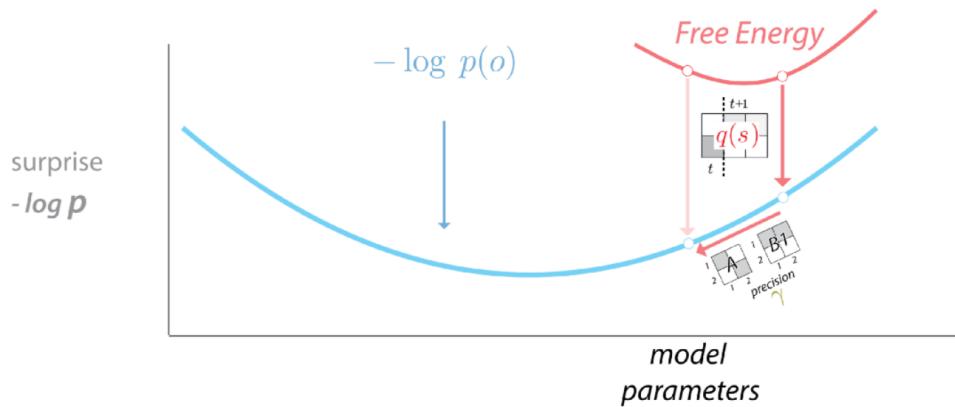
Итак, мы никогда не моделируем действие непосредственно, а вместо этого: оцениваем скрытые состояния при каждой политике → оцениваем ожидаемую свободную энергию для каждой политики → преобразуем ее в вероятность политики → усредняем состояния для всех политик и, наконец, → действуем так, как будто мы наблюдаем то, что мы ожидаем.

Формально связывание вероятности каждой политики с ожидаемой свободной энергией  $G$  осуществляется следующим образом. Модель  $p(o,s)$  фактически также включает в себя политику, больше похожую на  $p(o,s,\pi)$ . Аналогично, аппроксимация апостериорного  $q$  также включает политику и выглядит как  $q(s,\pi)$ . Решение для оптимальной политики показано аналитически, и в дополнение к ожидаемой свободной энергии  $G$ , включает в себя предыдущие предпочтения [prior preferences] по политике  $p(\pi)$  и свободную энергию с прошлых временных шагов (где мы фактически можем оценить точность, так как наблюдения уже известны). Однако, если мы предположим, что все политики в прошлом одинаковы – состоят из уже выполненных действий – аппроксимация апостериорного распределения политики  $q(\pi)$  действительно зависит только от ожидаемой свободной энергии  $G$  в будущем. Минимизация свободной энергии в будущем – это «сильно закодированный» априор [«hard-coded» prior], потому что вы связываете вероятность выбора определенной политики с ее ожидаемой свободной энергией, и это дополнительное предположение, которое вы должны сделать. Это приводит к минимизации одной свободной энергии ( $G$ ) внутри другой ( $F$ ). Вот вариационная свободная энергия с совместным распределением, развернутым во второй строке [Here is the variational FE with the joint expanded in the second line]:



Наконец, в дополнение к выводу, мы также изучаем параметры модели, такие, как матрица  $A$ , со-

поставляющая состояния и наблюдения, переходная матрица  $B$  и точность ( $\gamma$ ). Таким образом, минимизируя свободную энергию относительно параметров модели, мы не только делаем свободную энергию лучшим приближением неожиданности [surprise], но и минимизируем саму неожиданность.



Мы могли бы либо найти точечную оценку параметров, либо искать все распределение. В последнем случае мы включаем неопределенность параметров в нашу модель  $p$ , а также в аппроксимацию апостериорного  $q$ . Чтобы получить обоснованность модели/предельное правдоподобие/неожиданность [model evidence/marginal likelihood/surprise], мы затем суммируем не только латентные переменные, но и параметры. Таким образом, технически, поскольку мы не убрали параметры в этом посте, мы работали с вероятностью параметров  $p(o|parameters, model)$  [likelihood of the parameters], но не с обоснованностью модели  $p(o|model)$  [model evidence].

Конечно, есть и некоторые ограничения. Например, мы предполагаем, что аппроксимации апостериорного распределения ( $q_t|\pi$ ) независимы в каждый момент времени, и что они раскладываются на латентные переменные и параметры. Но что еще более важно, масштабируемость этой схемы ограничена, и до сих пор моделирование включает простые ситуации с небольшими пространствами состояний и наблюдений, так что вычисления (например, определение априорных предпочтений [aprior preferences]  $p(o)$ , что трудно, если существует много возможных наблюдений) остаются отслеживаемыми. Однако эта схема скорее предназначена для доказательства принципа, и ее вполне достаточно, чтобы быть полезной в вычислительной психиатрии. Например, считается, что точность [precision] отражает функцию дофамина, а это означает, что неспособность определить оптимальную точность будет иметь неблагоприятные психопатологические последствия. Аналогично, в недавнем исследовании использовался активный вывод [active inference] с акцентом на мотивацию, предполагая, что люди имеют либо целенаправленные, либо однородные априорные предпочтения [prior preferences]  $p(o)$ . Хотя эта схема была применена к игре DOOM в OpenAI Gym, она по существу сводится к тому же самому принципу, обсуждаемому здесь, поскольку пространство состояний было дискретизировано до максимума из 10 возможных состояний.