



# Retributive Emotions and Retributive Intuitions

---

**Bogdan V. Faul**

National Research University Higher School of Economics (HSE University Saint Petersburg).

Saint Petersburg, Russia. Email: [faulbogdan\[at\]gmail.com](mailto:faulbogdan[at]gmail.com) ORCID <https://orcid.org/0000-0001-9680-9944>

Received: 4 August 2025 | Revised: 9 September 2025 | Accepted: 17 September 2025

## Abstract

---

The article addresses a key challenge to moral justification in normative ethics – the relationship between emotions and retributive moral judgments. Retributivism holds that there are normative reasons to reward or punish individuals on the basis of merit, independent of consequences. A common objection argues that such beliefs are grounded in retributive emotions, which are allegedly unreliable foundations for moral reasoning. The study aims to analyze and evaluate the strength of this “argument from retributive emotions” in two of its main forms: evolutionary debunking arguments and moral rationalist objections. The research demonstrates that evolutionary debunking strategies, such as those developed by Joshua Greene and Isaac Wiegman, fail to undermine retributive intuitions, since anger is neither a necessary nor a unique basis for these judgments. Similarly, rationalist objections prove unconvincing, as moral reasoning is deeply intertwined with affective processes, as shown by the broader body of empirical work discussed by Shaun Nichols and others. The findings suggest that retributive intuitions cannot be dismissed solely because of their emotional foundation.

## Keywords

---

Retributivism; Emotions; Moral Epistemology; Evolutionary Debunking; Rationalism; Anger; Moral Sentimentalism; Punishment; Normative Reasons; Moral Psychology



This work is licensed under a [Creative Commons “Attribution” 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



## Ретрибутивные эмоции и ретрибутивные интуиции

---

**Фауль Богдан Владимирович**

Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ СПб).  
Санкт-Петербург, Россия. Email: [faulbogdan\[at\]gmail.com](mailto:faulbogdan[at]gmail.com)  
ORCID <https://orcid.org/0000-0001-9680-9944>

Рукопись получена: 4 августа 2025 | Пересмотрена: 9 сентября 2025 | Принята: 17 сентября 2025

### Аннотация

---

Статья посвящена одной из ключевых проблем морального обоснования в нормативной этике – соотношению эмоций и ретрибутивных моральных суждений. Ретрибутивизм утверждает, что существуют нормативные основания для поощрения или наказания индивидов на основании заслуженности, независимо от последствий. Распространённое возражение против этой позиции заключается в том, что такие убеждения основаны на ретрибутивных эмоциях, которые считаются ненадёжным фундаментом для морального мышления. Цель исследования – проанализировать и оценить силу «аргумента от ретрибутивных эмоций» в его двух основных формах: эволюционного разоблачения и морального рационализма. В статье показано, что стратегии эволюционного разоблачения, разработанные Джоша Грином и Исааком Вигманом, не подрывают ретрибутивные интуиции, так как гнев не является ни необходимым, ни уникальным основанием этих суждений. Аналогично, рационалистские возражения оказываются недостаточно убедительными, поскольку моральное мышление глубоко связано с аффективными процессами, о чём свидетельствуют эмпирические исследования, анализируемые Шауном Никольсом и другими авторами. Результаты исследования показывают, что ретрибутивные интуиции не могут быть отклонены лишь на основании их эмоциональной природы.

### Ключевые слова

---

ретрибутивизм; эмоции; моральная эпистемология; эволюционное разоблачение; рационализм; гнев; моральный сентиментализм; наказание; нормативные основания; моральная психология



Это произведение доступно по лицензии [Creative Commons “Attribution”](https://creativecommons.org/licenses/by/4.0/) («Атрибуция») 4.0



## **Введение**

---

В этой статье я рассмотрю одну из побочных, но вместе с тем важных тем философского осмысления эмоций в медиасфере. Концепция датарсиса показывает, как цифровая среда захватывает и структурирует эмоциональное переживание, превращая его в инструмент власти и стандартизации. Текущая работа переносит акцент в сферу нормативной этики и сосредоточена на вопросе о взаимосвязи *эмоций* и *моральных суждений*. Статья фокусируется на формулировке и анализе возражения к теории широкого ретрибутивизма, которое можно обозначить как *возражение от ретрибутивных эмоций*. В первом приближении, широкий ретрибутивизм – это этическая позиция, согласно которой практики поощрения и наказания могут быть *морально обоснованными и ценными* вне зависимости от их последствий. В возражении от ретрибутивных эмоций утверждается, что наши ретрибутивные представления основаны на особом классе эмоций, которые, по тем или иным причинам, не могут служить надёжным основанием для моральных суждений. Отсюда делается вывод, что ретрибутивизм лишен обоснования и не должен приниматься. В первых двух частях статьи я ввожу теоретические положения, на основании которых формулируется аргумент от ретрибутивных эмоций. Затем выделяются две трактовки аргумента: (1) через эволюционное разоблачение ретрибутивных интуиций и (2) через теорию морального рационализма, согласно которой эмоции не должны лежать в основе моральных суждений в целом. В последующих частях эти трактовки подвергаются анализу и критике. В анализе трактовок рассмотрена релевантная литература не только из области этики, но также из областей эволюционной и поведенческой психологии. Я постараюсь продемонстрировать, что эти трактовки аргумента проблематичны и не подрывают интуитивной поддержки ретрибутивизма.

## **Основные понятия**

---

Главная деонтологическая категория статьи – это *нормативное основание*. Нормативные основания – это базовая моральная категория, которую нельзя определить в более фундаментальных терминах. Важно отметить следующее: (1) нормативные основания могут иметь разную силу (что в естественном языке выражается, к примеру, словами «обязан» или «должен»); (2) личность может быть не осведомлена о том, какие нормативные основания у нее есть (например, «Он не должен употреблять наркотики, но не знает об этом»); (3) нормативные основания могут не мотивировать, если не принимаются в качестве руководства к действию (например, «Он знает, что не должен употреблять наркотики, но ему все равно»). В рамках этой статьи предполагается, что нормативные основания являются объективными в том смысле, что истинность или ложность утверждений о нормативных основаниях



не зависит от пропозициональных установок агентов, которые выносят суждения об этих основаниях. Однако в рамках различных подходов в нормативной этике может иметься зависимость между нормативными основаниями и строением агента. Например, в рамках аретических теорий можно утверждать, что у разных людей в сходных обстоятельствах будет различный набор нормативных оснований (Hursthouse, 1999). Например, если процветание человека, который любит музыку и литературу предполагает занятия музыкой и литературой, то у такого человека будут нормативные основания заниматься музыкой и литературой. Однако у другого человека, у которого иные условия процветания, таких оснований не будет, хотя внешние обстоятельства могут быть сходными или даже тождественными. Не стоит, однако, считать, что принятие категории нормативных оснований обязательно предполагает принятие какой-то конкретной теории в рамках нормативной этики.

Нормативные основания следует отличать от *мотивирующих оснований*. Мотивирующие основания не имеют деонтологического значения, в отличие от нормативных. Например, у кого-то может быть мотивирующее основание часто кричать на своего супруга или супругу в силу раздражительности, однако это не значит, что кричать на супруга позволительно. Понятие нормативного основания весьма удобно, поскольку само по себе не специфицирует силы этого основания. При первом приближении исключительно семантически основание удобнее всего мыслить так, как предлагается у Тима Скэнлона в его книге по метаэтике (Scanlon, 2014), хотя его подход, конечно, и не единственный. Он рассматривает основание как четырехместный предикат, который связывает четыре переменных: (а) факт, (б) личность, (в) обстоятельства и (г) поступок. Например, тот (а) факт, что плита горячая, является основанием для (б) меня в (в) текущих обстоятельствах (г) не прикасаться к плите. Все это не требует конкретной метаэтической теории Скэнлона. Следует также отметить, что в рамках этой статьи не говорится отдельно о *результатирующих основаниях* и о промежуточных основаниях, которые иногда еще называют основаниями *pro tanto*. Хотя это бывает необходимым в обсуждениях ретрибутивной справедливости, конкретно в рамках планируемого анализа в этом различии нет необходимости. С этого момента я не буду говорить о мотивирующих основаниях, но всегда буду говорить только о нормативных.

Перейдем к различению нескольких типов моральных теорий.

Во-первых, нужно различить *аксиологический* монизм и плюрализм. Во-вторых, *деонтологический* монизм и плюрализм. Аксиологические свойства следует отличать от деонтологических фактов. Аксиологические свойства – это свойства ценности тех или иных состояний мира. Например, у произведений искусства может быть высокая или низкая ценность; или же ценность может быть у состояния справедливости, как бы мы ее не трактовали. У состояния несправедливости, в свою очередь, может быть отрицательная ценность и т. д. Обычно мы говорим о таких состояниях через использование аксиологи-



ческой терминологии вроде хорошо или плохо, либо же через отдельные концепты, в которых соединяется описательный и аксиологический компоненты. В рамках этой статьи предполагается, что истинность или ложность суждений об аксиологических свойствах также не зависит от пропозициональных установок тех, кто выносит суждение. Тот факт, что есть аксиологические свойства, сам по себе не влечет ничего дополнительного в отношении того, *какие* это свойства и как они *распределены*. Защита реализма как в отношении должноствования, так и в отношении аксиологических свойств, выходит за пределы этой статьи.

С точки зрения *аксиологического монизма*, фундаментально ценным является только что-то одно. Например, с точки зрения некоторых форм утилитаризма фундаментально ценным является только удовольствие (в более сложных версиях, понимаемых через удовлетворение желаний, интересов или предпочтений (Nare, 1981; Singer, 1999)), а страдание является фундаментально плохим. Из этого следует, что всякая ценность в конечном итоге зависит от удовольствия или страдания. Подобные позиции могут быть мотивированы простотой построения моральной теории и ее способностью предписывать действия, а не только описывать их. Или же подобные позиции могут быть решением проблемы странности моральных фактов, с которой сталкивается моральный реализм. Например, Шэрон Роулетт аргументирует, что именно *феноменальные состояния сознания* являются наилучшим кандидатом для фундирования аксиологических свойств (Rawlette, 2016).

Согласно *аксиологическому плюрализму*, в свою очередь, существует множество фундаментальных ценностей. Например, любовь или красота могут считаться такими ценностями, и при этом не сводиться к удовольствию. В рамках этой статьи я буду говорить о подобных ценностях посредством взаимозаменяемых терминов: *недеривативная ценность*, *фундаментальная ценность*, *внутренняя ценность* и *самостоятельная ценность*. *Деривативная ценность* является ценной по причине того, что ценным является что-то другое. Например, деньги ценны не сами по себе, но по причине того, что с их помощью можно получить другую ценность вроде здоровья или свободы.

*Деонтологический монизм* устроен совсем иначе. С точки зрения этой позиции, всегда имеется *только одно нормативное основание* действовать. Моральные дилеммы лишь видятся таковыми, потому что в действительности мы всегда должны сделать *что-то одно*, хоть мы и не всегда понимаем, что именно. Нам лишь кажется, что есть конфликт оснований, тогда как в реальности его нет.

*Деонтологический плюрализм* предлагает альтернативный способ осмысления моральной дилеммы: моральная дилемма действительно создает *конкурирующие обязательства*, которые не могут быть одновременно удовлетворены. Таким образом, согласно деонтологическому плюрализму, бывает так,



что у нас есть более одного нормативного основания, и при этом мы вынуждены выбирать что-то одно. В рамках деонтологического плюрализма реальные обязательства можно мыслить в качестве деонтологических векторов, которые утягивают нас в разные стороны. Я не утверждаю, что нормативные основания обязательно мотивируют, но в качестве иллюстрации метафора утягивания вполне подходит. Например, у меня может быть основание писать статью, — и в то же время может быть основание пойти гулять, потому что это важно для моего здоровья. Однако я могу выбрать только что-то одно. Как следствие, из моральных дилемм не существует полностью непорочного выхода, что отражается в онтологии. Наиболее правдоподобным способом отстаивать деонтологический плюрализм является отрицание принципа, согласно которому из долженствования следует возможность (Tessman, 2015).

### **Ретрибутивизм**

Обычно под ретрибутивизмом имеют в виду теорию *легального наказания*, согласно которой есть основания наказывать людей за поступки *вне зависимости от последствий этого наказания*. Иными словами, согласно ретрибутивизму в теории легального наказания мы должны наказывать людей не только по причине определенных позитивных последствий, которые можно из наказания извлечь (например, исправление человека, изоляция человека, предотвращение будущих преступлений другими людьми, этическая коммуникация и пр.), но главным образом по причине, которая, как иногда это формулируется, «смотрит в прошлое» — человек совершил плохой поступок в прошлом и теперь *должен пострадать именно за это*. В этой статье ретрибутивизм в теории легального наказания я буду называть *узким ретрибутивизмом*. Чтобы очертить сферу действия узкого ретрибутивизма, полезно дать определение легального наказания. Самое распространенное определение называется определением Бенна-Флю-Харта — по причине того, что эти исследователи независимо сформулировали его (Benn, 1958; Flew, 1954; Hart, 1959). В соответствии с этим определением, легальное наказание — это (i) умышленное (ii) причинение вреда (iii) в отношении (предполагаемого) правонарушителя (iv) за правонарушение, (v) несущие экспрессивное послание осуждения, и (vi) назначаемое и исполняемое не потерпевшим, а иным субъектом, обладающим (предполагаемыми) полномочиями на его применение (как правило, государством) (Zisman & Rehren, 2024).

Узкий ретрибутивизм должен быть основан на том, что можно назвать *широким ретрибутивизмом* — теорией в нормативной этике, согласно которой есть *ретрибутивные* и *меритивные* нормативные основания в целом. Иными словами, согласно широкому ретрибутивизму, в целом бывает так, что *имеются нормативные основания поощрять или наказывать людей за их поступки*. Основания наказывать за поступок (или последствия поступка)



называют *ретрибутивными*, основания поощрять – *меритивными*. В практики поощрения и наказания входит весь набор практик, основанных на базовой *заслуженности*, и обычно в естественном языке они сопровождаются таким обоснованием: «Этот человек заслужил то-то и то-то», будь то нечто позитивное или негативное. В данный момент нет необходимости рассматривать теории базовой заслуженности, достаточно сказать, что иногда базовая заслуженность мыслится субстантивно, а иногда – нет. В первом случае заслуженность является свойством, которое создает нормативные основания для наказания или поощрения; во втором случае вся совокупность этих оснований по отношению к человеку и есть то, что называют базовой заслуженностью. Помимо деонтологического тезиса в рамках широкого ретрибутивизма есть также тезис *аксиологический*: ситуация справедливого наказания или поощрения на ретрибутивных или меритивных основаниях является ценной.

Важно отметить, что широкий ретрибутивизм сам по себе не утверждает, в каких конкретно случаях есть основания наказывать или поощрять людей. Например, может случиться так, что у родителей никогда нет именно ретрибутивных оснований наказывать своих детей, а всегда есть только консеквенциалистские основания вроде воспитания. Другой пример: у людей может не быть ретрибутивных оснований вообще, но они могут быть у Бога в силу каких-то Его особых свойств, например, тождественности со справедливостью как таковой в классическом теизме. Более того, широкий ретрибутивизм не утверждает абсолютности этих оснований, будто они не могут быть перевешены никакими другими. Широкий ретрибутивизм не уточняет степени пропорциональности поощрения или наказания за поступок. Помимо этого, широкий ретрибутивизм сам по себе не предполагает никакой эпистемологической программы, а именно, взгляда на то, насколько часто мы можем заблуждаться в отношении наличия или отсутствия ретрибутивных оснований. Единственное, что нужно для широкого ретрибутивизма (далее просто «ретрибутивизм») – это признание того, что иногда (а) *такие основания у кого-то есть*, и что есть (б) *ценность в справедливых ситуациях*, где справедливость трактуется через ретрибутивное основание.

У ретрибутивизма есть *значительная интуитивная поддержка*. Кажется, что в мире имеется огромное количество людей, которые заслуживают наказания и, таким образом, должны быть наказаны. Например, если мужчина пытал женщину на протяжении долгого времени, то кажется, что он *заслуживает* наказания и *должен* быть наказан, вне зависимости от последствий этого наказания. Более того, при прочих равных ситуация, в которой такой человек будет наказан, кажется *лучшей* по сравнению с ситуацией, где он наказан не будет.

Помимо подобных случаев, есть и более глубокие стратегии демонстрации интуитивной правдоподобности ретрибутивизма. Предположим, что есть два мира, в которых одинаковое количество счастья и стра-



дания. Однако в одном мире страдают хорошие люди, которые не делали ничего плохого, а счастливы плохие люди, сделавшие много зла. Во втором мире все наоборот: счастливы хорошие люди, а страдают плохие. Если нет ценности в том, что благо и зло распределяются в соответствии с базовой заслуженностью, то необходимо признать аксиологическое тождество этих миров. Однако кажется очевидным, что они не равны в своей ценности именно по той причине, что один из миров превосходит другой, поскольку блага в нем распределены в соответствии с заслуженностью. Иными словами, *распределение благ соответствует ретрибутивным и меритивным основаниям*. Выходит, что единственный способ объяснить разницу между этими мирами – это принять истинность ретрибутивизма. Подобные случаи, по утверждению защитника гедонизма Фреда Фельдмана, – это крепкий орешек для гедонизма (Feldman, 2004). Введем еще одно понятие, которым будем пользоваться дальше – *ретрибутивное состояние*. Ретрибутивное состояние – это такое состояние, в котором реализованы ретрибутивные или меритивные основания. Таким образом, аксиологическое превосходство одного мира над другим из предыдущего примера может быть описано тем, что в первом мире нет ретрибутивных состояний, тогда как они есть в большом количестве во втором мире.

Как широкий, так и узкий ретрибутивизм довольно органично сочетается с аксиологическим и деонтологическим плюрализмами. Лео Зайберт, например, защищает позицию, согласно которой ретрибутивное состояние с наказанием содержит в себе элемент дилеммы в том смысле, что страдание имеет очевидно негативную аксиологическую ценность, однако она может быть компенсирована состоянием целиком, которое оказывается аксиологически положительным. Подобные утверждения невозможны в рамках аксиологического монизма. Зайберт корректно идентифицирует проблему с большинством подходов в теории наказания. Он утверждает, что аксиологический монизм транслируется на теорию наказания в силу не до конца осознанной академической инерции, хотя большинство моральных философов уже осознано то, что хорошо сформулировано Бернардом Уильямсом: «Если существует истина относительно предмета этики – истина, можно сказать, об этическом, – то почему следует ожидать, что она окажется простой?» (Williams, 2015).

В этой статье нет возможности рассматривать все аргументы в пользу и против узкого или широкого ретрибутивизма. Фактически я сконцентрируюсь лишь на двух возражениях, у которых есть сходство формы. Однако для получения самого общего представления о том, как устроен общий контур дискуссии между сторонниками и противниками ретрибутивной справедливости, нужно все же указать на одно важное соображение<sup>1</sup>.

<sup>1</sup> Есть целый класс довольно существенных возражений, которые основаны на более общих соображениях, связанных со свободой воли и базовой заслуженностью (Caruso, 2021). В цели статьи не входит рассмотрение этой темы.



Как это часто бывает в нормативной этике, при построении этической теории (если вообще признается необходимость этого) философам приходится выбирать между непосредственными моральными интуициями и моральными интуициями в пользу абстрактных принципов. Дискуссия о ретрибутивизме является ярким примером этого конфликта. С одной стороны, у нас есть совокупность моральных интуиций, согласно которым ретрибутивная справедливость должна быть, или же фактически ценна, на что указано выше; с другой – есть интуитивно правдоподобный абстрактный принцип, согласно которому *причинение страдания в ответ на страдание не может содержать никакого блага*, а является лишь приумножением зла.

К примеру, в книге, посвященной осмыслению гнева и прощения, Марта Нуссбаум критикует гнев посредством критики его ретрибутивного компонента (Nussbaum, 2016). С точки зрения Нуссбаум, гнев в том или ином виде предполагает желание отмщения. Отмщение является либо (1) *этически порочным* по причине эгоцентричности, поскольку требует восстановления утерянного социального статуса; либо (2) *иррационально*, поскольку предполагает магическое мышление, согласно которому нужно восстановить некий космический баланс мира, хотя мы понимаем, что никакого космического баланса нет<sup>1</sup>. Именно часть (2) хорошо демонстрирует дух этого возражения к теории ретрибутивной справедливости: как можно с помощью мести добиться чего-то хорошего, если это есть ни что иное, как умножение зла? Это возражение может иметь различную форму, однако сама процедура, посредством которой делается этот вывод, остается схожей. Герберт Харт формулирует эту мысль следующим образом:

Это кажется таинственным образцом моральной алхимии, в котором сочетание двух зол – нравственного порока и страдания – превращается в благо (Hart, 2008, p. 234–35).

С этой точки зрения ретрибутивное состояние рассматривается не в целом, а по отдельности: сначала внимание направляется на негативные последствия поступка, а потом на воздаяние, которое кажется очередным событием с негативным аксиологическим статусом. Когда ситуация справедливого воздаяния разбивается на части, то действительно сложно увидеть ее ценность. Суть ретрибутивистской стратегии построения возражения на это *состоит в распознании ценности целого*, которая может превосходить отдельные элементы. Зайберт, например, опирается на идеи Джорджа Мура, Франса Brentano, Эдмунда Гуссерля и Родерика Чизолма, когда стремится продемонстрировать органическую ценность целого, которая может взяться, казалось бы, из ниоткуда. Например, отдельно взятые ноты из симфонии не представляют собой никакой ценности, однако если организовать их опре-

1 Есть глубокая, вполне возможно, аналитическая связь между справедливым воздаянием и прощением: прощение возможно только в ситуации, когда происходит отказ от полагающегося воздаяния. В силу этого Марта Нуссбаум тратит значительное количество времени для демонстрации того, чем становится прощение при отказе от идеи воздаяния.



деленным образом в крупное симфоническое произведение, оно становится ценным. Есть много тонкостей в обсуждении темы органических целых и их ценности, однако в данный момент погружаться в это нет необходимости. Важно другое: если мы считаем, что такая ситуация возможна и распространена, то возражение, которое разделяет злой поступок и воздаяние, теряет свою кажущуюся силу (Zaibert, 2018). Наши ретрибутивные интуиции объясняются тем, что бывают органически целые события, ценность которых может превосходить ценность частей, и ретрибутивное состояние является просто очередным примером этого.

### **Аргумент от ретрибутивных эмоций**

В данном разделе я предлагаю общую форму аргумента от ретрибутивных эмоций, в котором (широкий) ретрибутивизм критикуется по причине того, что основывается на эмоциях. В этой общей форме нет спецификации того, как именно осуществляется эта критика. После введения аргумента я предложу две его наилучшие трактовки.

1. *Тезис эмоционального фундирования*: утверждения о позитивной ценности ретрибутивных состояний, или о том, что есть нормативные основания достигать ретрибутивных состояний, основаны на ретрибутивных эмоциях.
2. *Тезис отрицания*: если верен тезис эмоционального фундирования, то широкий ретрибутивизм не обоснован.

Из этих двух тезисов следует вывод.

3. *Вывод*: широкий ретрибутивизм не обоснован.

Аргумент от ретрибутивных интуиций может быть проинтерпретирован двумя способами, которые кратко можно описать следующим образом.

1. *Эволюционное разоблачение*. Ретрибутивные эмоции получают эволюционное объяснение с точки зрения их практической полезности. Подобные объяснения должны продемонстрировать, что эти эмоции не отслеживают реальных моральных свойств. Как следствие, (широкий) ретрибутивизм не обоснован.
2. *Моральный рационализм*. Моральные суждения не должны основываться на эмоциях вообще, а должны основываться на корректном моральном мышлении. Ретрибутивные интуиции, однако, основываются на эмоциях. Как следствие, (широкий) ретрибутивизм не обоснован.

В следующих разделах рассматриваются эти две трактовки аргумента от ретрибутивных эмоций и демонстрируются проблемы, с которыми они сталкиваются.



## Эволюционное разоблачение

В этом разделе я рассмотрю трактовку аргумента, которая является разновидностью аргументов разоблачения. Общее устройство аргументов разоблачения можно описать следующим образом. Представьте, что вечером вы решили встретиться со своим знакомым и во время прогулки увидели НЛО. Вам кажется, что тарелка парит в небе, и у вас нет оснований сомневаться в том, что это так и есть. Однако ваш знакомый, который стоит рядом с вами, сообщает, что утром вам дали специальную таблетку, после которой люди видят НЛО. И действительно, вы вспоминаете об этом. Более того, вам говорят, что таблетка устроена так, что вы начинаете видеть НЛО ровно в десять вечера. Вы смотрите на часы — и сейчас как раз десять вечера. Кажется, что правильная реакция после этого — заключить, что на самом деле вы не видите НЛО, а вам лишь это кажется. Это не означает, что НЛО перед вами нет с какой-то абсолютной необходимостью. Это означает лишь то, что, если оно перед вами сейчас действительно есть, то это окажется невероятным совпадением<sup>1</sup>.

Эволюционные аргументы разоблачения могут быть направлены на подрыв наших эпистемических способностей как в области рационального познания мира, так и в этической сфере (Vavova, 2015). Если истинность или ложность моральных суждений не зависит от установок агентов, и естественный отбор направлен на адаптацию, а не на отслеживание истинных нормативных фактов, то есть основания считать, что наши убеждения о нормативных фактах необоснованны. Спецификация подобной аргументации, конечно, зависит от того, как конкретно демонстрируется, что знание моральных истин не даёт эволюционного преимущества. Например, Шэрон Стрит, среди прочего, предполагает, что наши моральные суждения нацелены на репродуктивный успех, и поэтому наши способности к оценочным суждениям сформированы *неправильным типом причин*. Это даёт основание заключить, что с большой долей вероятности они ошибочны (Street, 2006, p. 122).

Аргументация в этом разделе ориентирована на разоблачение не всех моральных интуиций, но только их подкласса. Я рассмотрю две версии такой аргументации. Одна принадлежит Джошуа Грину, и она нацелена на разоблачение *деонтологических моральных интуиций*, в противовес консеквенциалистским (Greene, 2008). Вторая версия принадлежит Исааку Вигману (Wiegman, 2017, 2023). Он стремится продемонстрировать эпистемологическую дефективность *именно ретрибутивных интуиций*.

Джошуа Грин апеллирует к концепции *двойного процесса* формирования моральных суждений: один из них базируется на эмоциях — *эмоциональный процесс*, другой основан на мышлении — *когнитивный процесс*. Соответственно, эмоции являются основой деонтологических моральных интуиций,

1 Для классификации и анализа устройства аргументов разоблачения, см. (Korman, 2019).



а когнитивные процессы – основой консеквенциалистских интуиций. После этого формулируется локальный аргумент разоблачения и утверждается, что эмоции, в отличие от мышления, ненадёжны в силу эволюционных причин. Как результат, деонтологические интуиции разоблачаются, а консеквенциалистские – нет.

Для иллюстрации различия Грин рассматривает два классических мысленных эксперимента с вагонеткой. Обе ситуации устроены так, что можно спасти пятерых людей, пожертвовав кем-то одним. Однако в одном случае нужно толкнуть человека с моста ради спасения пятерых, а во втором нажать на рычаг. Насилие в первом сценарии является *персональным*, а во втором – нет. Персональное насилие эволюционно древнее: оно существовало задолго до того, как сформировалась способность к абстрактному мышлению. Поэтому персональное насилие вызывает сильные эмоции. Неперсональное насилие в виде переключения рычага таких эмоций не вызывает, и поэтому открывается возможность *более взвешенного морального размышления о ситуации* через анализ издержек и выгод (Greene, 2008, р. 43). Размышление о случае с переключением рычага, согласно Грину, должно актуализировать регионы мозга, ассоциированные с более высокоуровневым мышлением. Помимо этого, Грин утверждает, что данные указывают на различие во времени реакции на эти случаи. Реакции в ситуациях, подобных случаю со сталкиванием человека с моста, гораздо быстрее, чем в ситуациях, которые подобны мысленным экспериментам с переключением рычага. Это также хорошо описывается его моделью: эмоциональные реакции быстрее, чем взвешенное размышление о моральных ситуациях (Greene, 2008, р. 46).

Каким конкретно образом концепция Грина позволяет разоблачить только деонтологические моральные интуиции? Для выполнения этой задачи нужна истинность следующего контрфактуала: в соответствии с тем, что нам известно, если бы не было деонтологических фактов, то эволюция всё равно сформировала бы в нас эмоции, которые были бы основой деонтологических интуиций (Wiegman, 2017, р. 198). На первый взгляд проблема состоит в том, что совершенно не очевидно, почему нельзя сформулировать подобного контрфактуала по отношению к аксиологическим суждениям, которые участвуют в консеквенциалистских соображениях. Например, почему нельзя сказать, что если бы боль не была *на самом деле плохой*, то есть не имела бы негативного аксиологического свойства, то мы все равно бы имели те же представления о боли?

Чтобы объяснить это, Грин отделяет эмоциональный процесс от когнитивного в том числе через введение понятия *эмоциональной валентности*, согласно которой у «быстрых эмоций»<sup>1</sup>, с его точки зрения, имеются следующие свойства:

---

1 В оригинале “alarm-like emotions”.



1. Они содержат *склонности* вести себя определенным образом или судить об этом поведении как об уместном.
2. Они актуализируются в ответ на некоторый *ограниченный* набор факторов.
3. Когда быстрые эмоции запускаются, они *пересиливают* когнитивный процесс.

В противоположность этому когнитивный процесс является *медленным, гибким и контролируемым*. Существенное различие между этими процессами состоит в том, что мышление может учитывать потенциально бесчисленное количество факторов, в отличие от эмоциональных реакций, которые актуализируются в ответ на ограниченный набор. Очевидное возражение состоит в том, что даже с этими уточнениями когнитивный процесс все еще может включать в себя неконсеквенциалистские основания. Например, можно размышлять о наказании человека по ретрибутивному основанию и взвешивать эти основания с соображениями исправления человека. Получается, что когнитивный процесс, который, как предполагается, является консеквенциалистским, не исключает деонтологических интуиций, тогда как в модели Грина такого быть не должно. Помимо этого, возможна и обратная ситуация: быстрые эмоции могут быть частью консеквенциалистских соображений (Wiegman, 2017, 202). Из этого следует, что позиция Грина не может сохранить разоблачение только части моральных убеждений, но коллапсирует в глобальное разоблачение моральных суждений.

Другая проблема, с которой сталкивается подход Грина – это недостаточная эмпирическая обоснованность модели двойного процесса. Валерий Зисман и Пауль Ререн указывают на проблемы с использованием эмпирических данных для обоснования или опровержения ретрибутивизма (Zisman & Rehren, 2024). Основная идея в том, что в рамках философских размышлений, которые поддерживают или опровергают ретрибутивизм, эмпирические послышки на самом деле далеко не так хорошо обоснованы, как утверждают авторы аргументов. В частности, модель двойного процесса крайне проблематична как в целом (Keren & Schul, 2009; Melnikoff & Bargh, 2018), так и в той версии, которую предлагает Джошуа Грин (Berker, 2009; Kahane, 2012). Более того, Зисман и Ререн провели отдельное исследование, посвященное эмпирической проверке модели двойного процесса. В нем участников тестировали в двух разных условиях: (1) условия *думайте внимательно*, где им давали инструкцию, что они должны внимательно обдумывать свой выбор; и (2) *контрольные условия*, где участникам не давали никаких инструкций. Во-первых, согласно модели двойного процесса, участники первой группы должны были предпочитать консеквенциалистские элементы. Во-вторых, участники в контрольных условиях должны были в большей степени предпочитать ретрибутивные элементы. Ничего из этого не подтвердилось (Rehren &



Zisman 2022). Таким образом, стратегия разоблачения, предложенная Грином, кажется несостоятельной.

Исаак Вигман модифицирует позицию Грина и стремится подорвать не все деонтологические интуиции, а *только ретрибутивные*. Более того, ретрибутивные интуиции подрываются только в отношении их применения к построению теории наказания. Вигман утверждает, что у ретрибутивных интуиций может быть полезная роль в силу того, что, как правило, эти интуиции могут приводить к решениям, которые верны по консеквенциалистским, а не ретрибутивным основаниям. Например, я могу опираться на ретрибутивную интуицию, что преступник должен быть наказан, и преступник действительно должен быть наказан, но не по ретрибутивной причине, а по причине, смотрящей в будущее: скажем, потому что это предотвратит будущие преступления. Ретрибутивная интуиция может быть полезной в том смысле, что на её основании я могу приходиться к правильным выводам о наказании. С точки зрения Вигмана, нельзя, однако, использовать ретрибутивные интуиции для поддержания *ретрибутивных теорий* наказания или справедливости (Wiegman, 2017, 2023).

Для демонстрации мысли он приводит такой пример. Представьте, что Джепетто создаёт киборга Пиноккио. Он делает его таким, что с большой долей вероятности Пиноккио будут травить в школе. Поэтому он вкладывает в Пиноккио такую установку, согласно которой *сопротивление травле является ценным само по себе*. Джепетто не может гарантировать, что Пиноккио всегда сможет осознавать долгосрочную пользу от сопротивления травле, поэтому вкладывает в Пиноккио убеждение В.

В: ценность акта сопротивления хулиганам не выводится (или выводится не только) из его последствий (или из последствий политики сопротивления).

Если бы Пиноккио узнал о том, что Джепетто задумал его таким, то он должен был бы заключить, что его склонность считать В верным не является свидетельством в пользу В. Таким образом, знание о замысле Джепетто подрывает тот способ, которым Пиноккио приходит к убеждению В (Wiegman, 2017, p. 205–6). Точно таким же образом устроена ситуация с самостоятельной ценностью ретрибутивизма, что выражается в нашем убеждении R.

R: ценность (или оправданность) акта наказания не выводится (или выводится не только из) из последствий акта (или практики) наказания.

Чтобы продемонстрировать устройство этой аналогии, я реконструирую мысль Исаака Вигмана чуть детальнее. Сначала он утверждает, что (а) ретрибутивные интуиции в том или ином виде основаны на эмоции *гнева*. Далее он указывает на (б) эволюционный функционал гнева, который объясняет причину, по которой мы склонны ко гневу и, как следствие, имеем ретрибутивные интуиции. Вигман утверждает, что убеждение R и соответствующие установки есть у нас по следующим причинам: во-первых, импульсы,



связанные с гневом и возмездием, эволюционировали как мотивация для конфронтации, у которой есть издержки на короткой дистанции, но польза на длинной; во-вторых, наказание является ключевым для объяснения человеческой кооперации в имеющихся масштабах (Wiegman, 2023, p. 141). Таким образом, поскольку мы имеем объяснение наших убеждений о самостоятельной ценности возмездия без какой-либо его реальной самостоятельной ценности, мы (в) не должны считать возмездие самостоятельно ценным.

Когда мы рассматривали теорию Грина, было легко продемонстрировать проблемы его главного различения и привести контрпримеры к нему. В случае с Вигманом все сложнее. Однако сама суть возражения остается прежней: действительно ли Вигману удастся избежать разоблачения всех моральных интуиций? Это возражение он рассматривает и отвечает следующим образом: если некоторая склонность к поведению развилась в силу ее полезности, то мы не можем считать, что само это действие является *самостоятельно ценным* (Wiegman, 2017, p. 212). Может показаться, что это не является содержательным ответом на проблему, потому что все наши склонности к моральным суждениям развились, как минимум частично, в силу их полезности. Например, можно заявить, что склонность ценить друзей была полезной для выживания. Это значит, что Вигману недоступно обоснование самостоятельной ценности дружбы на основании интуиции. Эту процедуру можно продолжить, и, таким образом, мы, кажется, получим проблему *глобального скептицизма в отношении моральных интуиций*. Подход Вигмана, однако, может сохранить часть интуиций, но только те, что свидетельствуют в пользу *деривативной ценности*. Эта позиция близка к глобальному скептицизму в отношении моральной интуиции, однако не тождественна ей. Тем не менее, подход Вигмана влечет довольно проблематичную аксиологическую позицию *радикального инструментализма*, согласно которой *все ценности являются деривативными*.

Вторая проблема, с которой сталкивается аргументация Вигмана – это обоснование того, что ретрибутивные интуиции действительно *уникальным образом* основаны на гневе, в противном случае разоблачению будет подвергнуты и другие типы размышления о наказании. Вигман предлагает эмпирическую поддержку этого тезиса. Однако Ререн и Зисман указывают на ее недостаточность. Прочитую их заключение раздела, посвященному этой проблеме:

Вызывает ли гнев ретрибутивное наказание? В этом разделе мы утверждали, что лишь немногие исследования напрямую касаются этого вопроса – либо потому, что они используют корреляционные дизайны, либо потому, что не контролируют или не измеряют мотивы участников к наказанию (или и то и другое одновременно). Мы также утверждали, что немногие исследования, преодолевающие эти ограничения, не поддерживают аргументы таких авторов, как Грин и Вигман. Мы заключаем, что на данный момент аргументы, суммированные нами в разделе 2 и опирающиеся на утверждение, что гнев вызывает карательное наказание, не представляются убедительными (Zisman & Rehren, 2024).



Помимо этого, гнев также не является обязательным для фундирования ретрибутивных интуиций, в пользу чего можно выдвинуть феноменологический аргумент. В разделе «Ретрибутивизм» я пытался продемонстрировать, что широкий ретрибутивизм поддерживается аксиологическими интуициями. В качестве примера приводилось два мира, которые равны с точки зрения удовольствия, но отличаются в том, что один мир несправедливый: в нем страдают те, кто этого не заслужил, а плохие люди счастливы. Во втором мире все наоборот: страдают только те, кто заслужил это. Разница между этими мирами в том, что первый не содержит ретрибутивных состояний, а во втором их очень много. Уместное описание разницы между этими ситуациями такое: *справедливый мир лучше, чем несправедливый*. У моего ментального состояния, которое я могу охарактеризовать через аксиологические концепты, есть как феноменальный характер, так и содержание. Однако несомненно то, что это ментальное состояние очень далеко отстоит от эмоции гнева, или даже чего-то близкого ко гневу. Даже если его можно охарактеризовать как эмоциональное состояние (на этом я остановлюсь в следующем разделе), мне не кажется, что эта аксиологическая интуиция каким-то образом зависит именно от гнева. Я не исключаю того, что гнев может обосновывать моральные суждения о наказании (хотя кажется правдоподобным, что гнев может фундировать моральное суждение о том, что нужно, например, изолировать человека), однако общая интуитивная поддержка (широкого) ретрибутивизма необязательно зависит от гнева. Таким образом, гнев не только не представляет собой специфической поддержки интуиций в пользу ретрибутивизма, но, кажется, он и не является обязательным для поддержки всех этих интуиций. Если это так, то аргументация в стиле Вигмана просто не разоблачает всего комплекса ретрибутивных интуиций.

Таким образом, аргумент от ретрибутивных эмоций, проинтерпретированный через эволюционное разоблачение, проблематичен. В формулировке аргумента через концепцию двойного процесса в стиле Джошуа Грина есть две проблемы. Во-первых, такая форма аргумента приводит в конечном итоге к глобальному скептицизму в отношении моральных интуиций. Во-вторых, теория двойного процесса на сегодняшний день уже устарела и с трудом выдерживает критику. Если же взять аргументацию в стиле Исаака Вигмана, то она тоже сталкивается с трудностями. Во-первых, позиция Вигмана в конечном итоге сближается с глобальным скептицизмом в отношении моральных интуиций, хотя и не отождествляется с ним. Вигман вынужден принимать вывод, что всякая ценность является деривативной, что само по себе кажется довольно проблематичным выводом (хотя, конечно, у этой позиции есть сторонники). Во-вторых, нет достаточных научных оснований считать, что гнев – это эмоция, которая является основой исключительно ретрибутивных интуиций в отношении наказания. Если гнев лежит в основе морального мышления о наказании в целом, то тогда аргумент Вигмана подрыв-



вает и неретрибутивные обоснования наказания, которые, на самом деле, он желает оставить. Более того, кажется, что есть ряд аксиологических интуиций в пользу широкого ретрибутивизма, которые вообще не основаны на гневе. Из этого следует, что гнев не является ни *уникальным* для поддержки ретрибутивизма, ни *обязательным*. Можно вполне правдоподобно заключить, что рассмотренные стратегии разоблачения конкретно ретрибутивных интуиций не являются успешными.

## **Моральный рационализм**

В этом разделе рассматривается вторая трактовка аргумента от ретрибутивных эмоций. В рамках обсуждения этой темы со специалистами я неоднократно сталкивался с рассуждением, что эмоции, которые фундируют ретрибутивные интуиции, являются либо слишком сильными, и потому неправильным образом воздействуют на процесс морального размышления, либо же корректный процесс морального размышления вообще не должен основываться на эмоциях ни в каком виде. Я не склонен отрицать очевидный факт, что сильные эмоции неправильным образом воздействуют на моральное мышление. Однако кажется, что ретрибутивные интуиции не должны основываться на сильных эмоциях, что я постарался продемонстрировать в предыдущем разделе. Я буду исходить из того, что, даже если в моральном размышлении ретрибутивные основания базируются на эмоциях, есть множество ситуаций, в которых эти эмоции не так сильны, чтобы каким-то существенным образом *искажать этот процесс*. В связи с этим в данном разделе отстаивается тезис, что моральное мышление вообще не должно основываться *ни на каких эмоциях, даже на слабых*. Эту позицию я, вслед за Шоном Никольсом, буду называть *моральным рационализмом* (Nichols, 2004).

Моральный рационализм является частью этической традиции, современные сторонники которой включают в себя Алана Гевирца (Gewirth, 1981), Стивена Дорволла (Darwall, 1983), Томаса Нагеля (Nagel, 1978, 1986), Кристин Корсгаард (Korsgaard, 1986), Майкла Смита (Smith, 1996, 1997, 2009) и других. Никольс корректно утверждает, что этический рационализм является одним из привлекательных способов сохранения морального реализма. Действительно, одна из проблем морального реализма состоит в объяснении возможности глобальной супервентности моральных фактов на естественных фактах, потому что в противном случае распределение моральных фактов случайно, и неясно, почему они связаны с теми или иными *естественными* состояниями мира. В силу каких фактов, например, объективно плохо оскорблять людей и хорошо быть вежливым, и почему не наоборот? Моральный рационализм позволяет объяснить это. С этой точки зрения, *аморальное*, в конечном итоге, – это не что-то таинственное, а наоборот, вполне знакомое для нас: *аморальное* – это *иррациональное*. Более того, для познания этих моральных фактов не нужно постулировать специальной способности, с помощью которой



мы получаем доступ к моральным фактам, достаточно лишь разума (Korsgaard, 1986).

Для демонстрации того, как подобные позиции могут быть устроены, сконцентрируюсь на конструктивизме в метаэтике (Street, 2010). Шэрон Стрит предлагает следующую характеристику метаэтического конструктивизма: истинность нормативного утверждения состоит в том, что это утверждение логически дедуцируется из *практической точки зрения*, тогда как *практической точке зрения* дается формальная характеристика (Street, 2010, p. 367). Стрит объясняет практическую точку зрения следующим образом: быть может, мы и не уверены в том, что такое ценность, однако мы точно знаем, что мы склонны что-то *оценивать*. Практическая позиция занимается субъектом тогда, когда субъект выносит суждение о ценности, или *принимает* нечто как ценное (Street, 2006, p. 366). После этого Стрит выделяет кантианскую и юмовскую версии конструктивизма. Согласно Кантианской версии, из практической точки зрения следуют моральные суждения, которые, как предполагается, истинны в полностью реалистском смысле. Примерно так теорию Иммануила Канта трактует Кристин Корсгаард: вне зависимости от того, понимает ли это некий отдельный субъект или нет, его фактическая практическая точка зрения влечет определенные этические выводы (Korsgaard, 1996). Согласно юмовской версии, то, что следует из практической позиции, говорит лишь о том, что рационально принимать тому или иному субъекту, не более того. Этой версии придерживается и сама Шэрон Стрит (Street, 2010, 370). Кантианская версия метаэтического конструктивизма может выступать в качестве хорошего примера морального рационализма, если, конечно, дополнительно указать на то, что универсальная практическая точка зрения совершенно независима от эмоций.

В данном разделе я постараюсь продемонстрировать, что моральный рационализм, который является основой для подрыва суждений, основанных на эмоциях, является крайне неправдоподобной позицией, в том числе в свете современных научных данных. Данный раздел главным образом опирается на размышления Шона Никольса, о котором уже упоминалось.

С точки зрения Никольса, необходимо различать *концептуальную* и *эмпирическую* трактовки рационализма. Согласно *концептуальной* трактовке, концепт морального обязательства – это концепт *рационального основания действовать*. Согласно *эмпирической* трактовке, фактические моральные суждения людей следуют из рациональных способностей человека.

Одним из ярких сторонников концептуальной трактовки является Майкл Смит (Smith, 2003). С его точки зрения, концептуальный рационализм влечет *требование практичности*, согласно которому агенты, которые выносят моральное суждение, *мотивированы* этим суждением. Это должно быть так по той причине, что в теориях в стиле Смита *моральные суждения отождествляются с суждениями о рациональных основаниях действовать*.



Однако это место теории легко атаковать посредством введения фигуры полностью разумного, но аморального психопата, который способен выносить моральные суждения, однако не мотивирован ими. Ответ Смита на критические замечания такого рода устроен так: на самом деле психопаты не выносят моральных суждений вообще, они выносят только суждения, которые *выглядят так*, словно это моральные суждения. Из этого следует, что если психопаты не мотивированы морально, то они и не понимают моральных суждений.

В ответ на интерпретацию Смита Никольс выстраивает аргументацию посредством концептуального анализа моральных понятий, на основании метода рассмотрения банальных суждений (Lewis, 1970). Он рассуждает так: как правило, люди считают, что психопаты понимают моральные утверждения, но *не действуют на их основании*. Получается, что устройство моральных концептов таково, что их можно *понимать* — и не быть при этом мотивированным к *моральному поведению*. Из этого следует, что сами по себе моральные концепты не могут являться концептами о разумных причинах поведения, потому что в противном случае фигура психопата не могла бы быть осмыслена так, как она осмысливается обычными людьми в рамках банальных суждений (Nichols, 2004, p. 71–76). Этот аргумент, конечно, основан на конкретной методике концептуального анализа, однако этой методикой пользуется и сам Смит для доказательства своего тезиса. Поэтому данная стратегия атаки концептуальной трактовки морального рационализма кажется вполне успешной.

Эмпирическая трактовка морального рационализма, в свою очередь, опровергается исследованиями конкретных психопатов. Дело в том, что реальные психопаты часто с трудом различают *моральное* и *конвенциональное*, однако у них, кажется, все в порядке с разумными способностями (Blair, 1995). Если верен эмпирический рационализм, то неспособность психопатов полностью понимать моральные концепты должна объясняться исключительно дефектами в рациональных способностях. Проблема, однако, в том, что такого объяснения нет. Все объяснения отклонений в области рационального недостаточны, поскольку люди с гораздо большими отклонениями в своих рациональных способностях (например, люди с аутизмом или синдромом дауна (James & Blair, 1996)) способны различать моральное и конвенциональное. Наиболее правдоподобным объяснением неспособности различать моральное и конвенциональное является гипотеза *аффективного расстройства*, согласно которой у психопатов нет достаточной развитости именно аффективных механизмов (Nichols, 2004, p. 82). Это делает дополнительный вклад в пользу *морального сентиментализма*, согласно которому моральные суждения не основаны исключительно на разуме, но зависят от аффективных состояний и эмоций. Моральный сентиментализм отстаивается в книге Никольса и поддерживается рядом других эмпирических иссле-



дований. Несмотря на то, что я считаю, что какая-то версия сентиментализма должна быть верна в силу современных психологических исследований (Kauppinen, 2022), у меня нет возможности подробно аргументировать в его пользу.

Вернемся обратно к аргументу от ретрибутивных эмоций и посмотрим на то, какие следствия для этого у нас имеются, исходя из обсуждения в разделе. Еще раз подчеркнем: как предполагается в аргументе, *моральные суждения о ретрибутивной справедливости не должны основываться на эмоциях*. Я постарался предложить наиболее щедрую трактовку этого тезиса через концепцию морального рационализма. Даже если отложить концептуальную трактовку тезиса, проблема в том, что наше фактическое моральное мышление основано на эмоциях и аффектах в том или ином виде. Из этого следует, что: либо (а) мы вообще не должны выносить большинства моральных суждений по той причине, что многие из них основаны на аффектах, как минимум частично; либо (б) мы не должны выносить именно ретрибутивных моральных суждений. Конечно, вариант (а) кажется крайне неправдоподобным, если мы хотим атаковать именно широкий ретрибутивизм, но сохранить моральный реализм и какую-то форму эпистемического доступа к нормативным фактам на основании интуиции. Проблема варианта (б) состоит в том, что очень трудно объяснить, что именно не так с эмоциями и аффектами, которые обосновывают широкий ретрибутивизм. Более того, объяснение (б) должно быть *уникально* направлено на эмоции и аффекты, которые частично фундируют широкий ретрибутивизм. Все это кажется крайне непростой задачей. Из этого следует, что аргумент от ретрибутивных эмоций, проинтерпретированный через концепцию морального рационализма, также не представляется успешным.

## **Заключение**

В статье рассмотрен аргумент против широкого ретрибутивизма, который указывает на эмоциональную природу широкого ретрибутивизма и на основании этого заключает, что это тем или иным образом подрывает теорию. Я рассмотрел этот аргумент в двух видах по той причине, что в практике обсуждения этой темы с коллегами аргумент озвучивается предельно широким образом, после чего уточняется в процессе обсуждения. Предположив, что в рамках статьи стоит повторить такой естественный способ обсуждения этой проблемы, сначала я рассмотрел трактовку *в стиле аргумента разоблачения*. Согласно наилучшей версии этой трактовки, широкий ретрибутивизм основывается на *гневе*, однако по той причине, что у гнева есть эволюционный функционал, мы не должны основывать наши суждения на нем. Я продемонстрировал, что эта аргументация несовершенна главным образом по той причине, что гнев не является ни *уникальной* основой ретрибутивных интуиций, ни *обязательной*. Гнев может участвовать в других типах морального



мышления, не делая эти размышления проблематичными, а также есть ретрибутивные интуиции, которые не основаны на гневе. В рамках второй трактовки аргумента я рассмотрел тезис о том, что мы вообще не должны формировать моральные убеждения на внерациональных основаниях. В ответ на эту форму аргумента я указал на то, что мы в целом являемся *аффективными существами* и основываем наши моральные суждения (как минимум их большую часть) на аффектах и эмоциях в большей или меньшей степени, включая ретрибутивные суждения. Моральная теория, которая не учитывает этого, кажется оторванной от реальности нашего устройства и задает невозможный стандарт морального суждения. Надеюсь, мне удалось продемонстрировать, что есть достаточно правдоподобный общий контур ответа на аргумент от ретрибутивных эмоций, отдельные аспекты которого, при необходимости, могут быть доработаны в будущем.

## **Заявления автора**

---

### **Согласие с политиками и положениями сетевого издания**

Автор данной рукописи был уведомлен и согласен с действующими политиками и положениями сетевого издания, изложенными на веб-сайте.

### **Оригинальность и эксклюзивность рукописи**

Автор подтверждает, что рукопись является оригинальной, ранее не публиковалась ни полностью, ни частично и не находится на рассмотрении для публикации в других изданиях.

### **Авторство и вклад (CRediT)**

Проведение исследования, концептуализация и написание статьи — Фауль Б. В.

### **Политика в отношении препринтов и предыдущих версий**

Автор подтверждает, что нет препринта/рабочей версии.

### **Конфликт интересов**

Конфликтов нет.

### **Финансирование**

Исследование выполнено при поддержке Российского научного фонда в рамках проекта «Экзистенциальный опыт в цифровой среде: “бытие к цифре”, онтология виртуального и человеческого Я», №25-18-00208, в НИУ ВШЭ в Санкт-Петербурге.

### **Использование генеративного ИИ и вспомогательных инструментов**

Инструменты искусственного интеллекта не указаны в качестве авторов и не использовались при проведении исследования.



### **Разрешения на использование (тексты, изображения, карты, иллюстрации)**

Все материалы, защищённые авторским правом, воспроизводятся с разрешения правообладателя; соблюдаются условия лицензии (включая Creative Commons).

### **Благодарности и вклады без авторства**

Отсутствуют.

## Список литературы | References

---

- Benn, S. I. (1958). An Approach to the Problems of Punishment. *Philosophy*, 33(127), 325–341. <https://doi.org/10.1017/S0031819100055017>
- Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs*, 37(4), 293–329. <https://doi.org/10.1111/j.1088-4963.2009.01164.x>
- Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29. [https://doi.org/10.1016/0010-0277\(95\)00676-P](https://doi.org/10.1016/0010-0277(95)00676-P)
- Caruso, G. D. (2021). *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*. Cambridge University Press. <https://doi.org/10.1017/9781108689304>
- Darwall, S. L. (1983). *Impartial Reason*. Cornell University Press.
- Feldman, F. (2004). *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Clarendon Press. <https://doi.org/10.1093/019926516X.001.0001>
- Flew, A. (1954). The Justification of Punishment. *Philosophy*, 29(111), 291–307. <https://doi.org/10.1017/S0031819100067152>
- Gewirth, A. (1981). *Reason and Morality*. University of Chicago Press.
- Greene, J. D. (2008). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: Vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (pp. 35–80). MIT Press.
- Hare, R. M. (1981). *Moral Thinking: Its Levels, Method, and Point*. Oxford University Press. <https://doi.org/10.1093/0198246609.001.0001>
- Hart, H. L. A. (1959). The Presidential Address: Prolegomenon to the Principles of Punishment. *Proceedings of the Aristotelian Society*, 60, 1–26. <https://doi.org/10.1093/aristotelian/60.1.1>
- Hart, H. L. A. (2008). *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford University Press.
- Hursthouse, R. (1999). *On Virtue Ethics*. OUP Oxford.
- James, R., & Blair, R. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders*, 26(5), 571–579. <https://doi.org/10.1007/BF02172277>
- Kahane, G. (2012). On the Wrong Track: Process and Content in Moral Psychology. *Mind & Language*, 27(5), 519–545. <https://doi.org/10.1111/mila.12001>
- Kauppinen, A. (2022). Moral Sentimentalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2022/entries/moral-sentimentalism/>



- Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science*, 4(6), 533–550. <https://doi.org/10.1111/j.1745-6924.2009.01164.x>
- Korman, D. Z. (2019). Debunking arguments. *Philosophy Compass*, 14(12), e12638. <https://doi.org/10.1111/phc3.12638>
- Korsgaard, C. M. (1986). Skepticism about Practical Reason. *The Journal of Philosophy*, 83(1), 5. <https://doi.org/10.2307/2026464>
- Korsgaard, C. M., & O'Neill, O. (1996). *The Sources of Normativity* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511554476>
- Lewis, D. (1970). How to Define Theoretical Terms. *The Journal of Philosophy*, 67(13), 427. <https://doi.org/10.2307/2023861>
- Melnikoff, D. E., & Bargh, J. A. (2018). The Mythical Number Two. *Trends in Cognitive Sciences*, 22(4), 280–293. <https://doi.org/10.1016/j.tics.2018.02.001>
- Nagel, T. (1978). *The Possibility of Altruism*. Princeton University Press. <https://doi.org/10.1515/9781400884490>
- Nagel, T. (1986). *The View From Nowhere*. Oxford University Press.
- Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford University Press. <https://doi.org/10.1093/0195169344.001.0001>
- Nussbaum, M. C. (2016). *Anger and Forgiveness: Resentment, Generosity, Justice*. Oxford University Press.
- Rawlette, S. H. (2016). *The Feeling of Value: Moral Realism Grounded in Phenomenal Consciousness*. CreateSpace Independent Publishing Platform.
- Rehren, P., & Zisman, V. (2022). Testing the Intuitive Retributivism Dual Process Model. *Zeitschrift Für Psychologie*, 230(2), 152–163. <https://doi.org/10.1027/2151-2604/a000461>
- Scanlon, T. (2014). *Being Realistic about Reasons*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199678488.001.0001>
- Singer, P. (1999). *Practical ethics*. Cambridge University Press.
- Smith, M. (1996). The argument for internalism: Reply to Miller. *Analysis*, 56(3), 175–184. <https://doi.org/10.1093/analys/56.3.175>
- Smith, M. (1997). In Defense of “The Moral Problem”: A Reply to Brink, Copp, and Sayre-McCord. *Ethics*, 108(1), 84–119. <https://doi.org/10.1086/233789>
- Smith, M. (2003). *The Moral Problem*. Blackwell.
- Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 127(1), 109–166. <https://doi.org/10.1007/s11098-005-1726-6>
- Street, S. (2010). What is Constructivism in Ethics and Metaethics? *Philosophy Compass*, 5(5), 363–384. <https://doi.org/10.1111/j.1747-9991.2009.00280.x>
- Tessman, L. (2015). *Moral Failure: On the Impossible Demands of Morality*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199396146.001.0001>
- Vavova, K. (2015). Evolutionary Debunking of Moral Realism. *Philosophy Compass*, 10(2), 104–116. <https://doi.org/10.1111/phc3.12194>



- Wiegman, I. (2017). The Evolution of Retribution: Intuitions Undermined. *Pacific Philosophical Quarterly*, 98(2), 193–218. <https://doi.org/10.1111/papq.12083>
- Wiegman, I. (2023). Doubts about Retribution: Is Punishment Non-Instrumentally Good or Right? In M. C. Altman (Ed.), *The Palgrave Handbook on the Philosophy of Punishment* (pp. 125–147). Springer International Publishing. [https://doi.org/10.1007/978-3-031-11874-6\\_6](https://doi.org/10.1007/978-3-031-11874-6_6)
- Williams, B. (2015). *Ethics and the Limits of Philosophy*. Routledge.
- Zaibert, L. (2018). *Rethinking Punishment* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108151740>
- Zisman, V., & Rehren, P. (2024). The Problems of Empirically-Informed Arguments for and against Retributivism. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-024-00753-w>