КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2024 Т. 16 № 5 С. 1323–1335
DOI: 10.20537/2076-7633-2024-16-5-1323-1335

Ки&М

МОДЕЛИ ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

УДК: 51-77

Моральный выбор: математическая модель

С. Ю. Малков[1,a], О. А. Шпырко[2,b], О. И. Давыдова[3,c]

[1]Центр долгосрочного прогнозирования и стратегического планирования МГУ имени М. В. Ломоносова,
Россия, 119991, г. Москва, ул. Ленинские горы, д. 1
[2]Филиал МГУ имени М. В. Ломоносова в городе Севастополе,
Россия, 299001, г. Севастополь, ул. Героев Севастополя, д. 7
[3]ООО «АйДесайд Консалтинг»,
Россия, 141070, г. Королев, ул. Калинина, д. 6б, оф. 32

E-mail: [a] s@malkov.org, [b] shpyrko@mail.ru, [c] davydova.olga.msk@gmail.com

Получено 13.04.2024, после доработки — 04.07.2024.
Принято к публикации 18.07.2024.

В работе приведены результаты исследований по созданию математической модели морального выбора, основанной на развитии подхода, предложенного В. А. Лефевром. В отличие от В. А. Лефевра, который рассматривал весьма умозрительную ситуацию морального выбора субъекта между абстрактными добром и злом под давлением на него внешнего мира с учетом субъективного восприятия субъектом этого давления, в нашем исследовании рассмотрена более приземленная и практически значимая ситуация. Рассматривается случай, когда субъект при принятии решений ориентируется на свое индивидуальное восприятие внешнего мира (которое может быть искаженным, например, вследствие внешнего целенаправленного информационного воздействия на субъекта и манипулирования его сознанием), а добро и зло не абстрактны, а обусловлены системой ценностей, принятой в конкретном рассматриваемом обществе и привязанной к конкретной идеологии/религии, которые могут быть различными для разных обществ.

В результате проведенных исследований разработана базовая математическая модель, рассмотрены частные случаи ее применения. Выявлены некоторые закономерности, связанные с моральным выбором, приведено их формальное описание. В частности, на языке модели рассмотрена ситуация манипулирования сознанием, сформулирован закон снижения моральности общества, состоящего из так называемых *свободных* субъектов (то есть таких, которые стремятся действовать в соответствии со своими интенциями и соответствовать в своих действиях образу своего «я»).

Ключевые слова: моральный выбор, математическая модель, интенция, функция готовности, система ценностей, свободный субъект

Работа выполнена при поддержке Программы развития МГУ, проект № 24-Ш05-12.

COMPUTER RESEARCH AND MODELING
2024 VOL. 16 NO. 5 P. 1323–1335
DOI: 10.20537/2076-7633-2024-16-5-1323-1335

Ки&М

MODELS OF ECONOMIC AND SOCIAL SYSTEMS

UDC: 51-77

Features of social interactions: the basic model

S. Yu. Malkov[1,a], O. A. Shpyrko[2,b], O. I. Davydova[3,c]

[1]Center for Long-Term Forecasting and strategic planning of Moscow State University,
1 Leninskie gory st., Moscow, 119991, Russia
[2]Branch of Moscow State University named after M. V. Lomonosov in the city of Sevastopol,
7 Geroev Sevastopolja st., Sevastopol, 299001, Russia
[3]iDecide Consulting LLC,
6b/32 Kalinina st., Korolev, 141070, Russia

E-mail: [a] s@malkov.org, [b] shpyrko@mail.ru, [c] davydova.olga.msk@gmail.com

Received 13.04.2024, after completion — 04.07.2024.
Accepted for publication 18.07.2024.

The paper presents the results of research on the creation of a mathematical model of moral choice based on the development of the approach proposed by V. A. Lefebvre. Unlike V. A. Lefebvre, who considered a very speculative situation of a subject's moral choice between abstract "good" and "evil" under pressure from the outside world, taking into account the subjective perception of this pressure by the subject, our study considers a more mundane and practically significant situation. The case is considered when the subject, when making decisions, is guided by his individual perception of the outside world (which may be distorted, for example, due to external purposeful informational influence on the subject and manipulation of his consciousness), and "good" and "evil" are not abstract, but are conditioned by a value system adopted in a particular society under consideration and tied to a specific ideology/religion, which may be different for different societies.

As a result of the conducted research, a basic mathematical model has been developed, and special cases of its application have been considered. Some patterns related to moral choice are revealed, and their formal description is given. In particular, the situation of manipulation of consciousness is considered in the language of the model, the law of reducing the "morality" of a society consisting of so-called free subjects (that is, those who strive to act in accordance with their intentions and correspond in their actions to the image of their "I") is formulated.

Keywords: moral choice, mathematical model, intention, readiness function, value system, free subject

Citation: *Computer Research and Modeling*, 2024, vol. 16, no. 5, pp. 1323–1335 (Russian).

This work was done with the support of MSU Program of Development, Project No. 24-SCH05-12.

Введение. В. А. Лефевр и его аналитическая модель субъекта

Несмотря на то что научная литература по теории принятия решений включает в себя огромное количество работ (см., например, [Keeney, Raiffa, 1976; Saaty, 1980; Айзерман, Алескеров, 1990; Малков, 2008; Розен, 2002; Подиновский, Потапов, 2003; Подиновский, Ногин, 2007; Соболь, Статников, 2006; Соколов, Токарев, 2011; Aleskerov, Bouyssou, Monjardet, 2007; Brams, Taylor, 1996; Clemen, 1996; Raiffa, 1997; Roth, Sotomayor, 1990; Smith, 1988]), серьезной проблемой остается учет в соответствующих математических моделях деонтологических аспектов, иными словами, учет влияния ценностных/моральных/религиозных факторов на принятие решений субъектами в различных ситуациях. Как правило, в исследованиях по теории принятия решений рассматриваются ситуации рационального/утилитарного выбора, когда возможны *количественные* оценки последствий принимаемых решений (с учетом того, что эти оценки могут иметь нечеткий или интервальный характер), с последующим решением оптимизационных задач. При этом деонтологические аспекты учитываются путем введения *функции полезности*, формируемой экспертным образом и/или на основе проведенных социологических исследований. Особенно остро задача учета деонтологических аспектов при принятии решений встала в последние годы в связи с широким распространением технологий искусственного интеллекта и машинного обучения, поскольку автоматизированное (с использованием искусственного интеллекта) принятие решений, затрагивающее социальную сферу, должно учитывать моральные аспекты последствий принимаемых решений (см., например, работы [Strimling et al., 2019; Schramowski et al., 2020; Baird, Schuller, 2020]). Проблема заключается в том, что *функции полезности* формируются экспертами на основе их личного понимания того, что в рассматриваемом обществе считается правильным и неправильным, моральным и аморальным и т. п. Процессы формирования в сознании ЛПР (то есть лиц, принимающих решения) того, что морально, а что аморально для них в конкретной ситуации и как ЛПР будут поступать, если утилитарная выгода в конкретной ситуации входит в противоречие с общепринятыми понятиями нравственности, исследуются весьма редко.

Наиболее детально и глубоко вопросы математического моделирования морального выбора были рассмотрены в работах В. А. Лефевра (см., например, [Лефевр, 1991; Лефевр, 2000; Лефевр, 2003а; Лефевр, 2005]). Его подход заключался в следующем[1]: он рассматривал ситуацию, когда некий субъект находится перед выбором одной из двух альтернатив, первая из которых олицетворяет для субъекта добро (В. А. Лефевр называет ее позитивным полюсом), а вторая — зло (называемая негативным полюсом). В. А. Лефевр вводит функцию *готовности* субъекта к действию:

$$X = f(x_1, x_2, x_3), \tag{1}$$

где значения X, x_1, x_2, x_3 изменяются в пределах отрезка [0; 1], причем 1 — это «добро» (позитивный полюс), а 0 — это зло (негативный полюс). Функция (1) показывает, как влияет давление внешнего мира на принятие решения субъектом, изменяя его первоначальную *интенцию* (намерение); x_3 — интенция (вероятность, с которой субъект изначально намеревался выбрать позитивный полюс), x_1 — давление внешнего мира в сторону позитивного полюса, x_2 — представление субъекта о давлении внешнего мира в сторону позитивного полюса; величина X отражает итоговую *готовность* субъекта (вероятность, с которой субъект *готов* выбрать позитивный полюс).

Принципиальным является то, что В. А. Лефевр различает *реальное* давление внешнего мира (x_1) в сторону позитивного полюса (например, наличие общепринятых норм морали и нравственности) и *представление* субъекта о давлении внешнего мира (x_2), которое может быть искаженным, ошибочным, не соответствовать x_1. Причина этого несоответствия может быть как

[1] Мы излагаем подход В. А. Лефевра так, как он изложен в его работе [Лефевр, 2005, с. 40 и далее].

внутренней (обусловленной психологическими и ментальными особенностями субъекта, ограниченностью его информации о внешнем мире), так и внешней (обусловленной внешним целенаправленным воздействием на сознание субъекта, манипуляцией).

Свои рассуждения В. А. Лефевр строит на следующих постулатах.

Постулат 1. *При фиксировании любых двух переменных из тройки x_1, x_2, x_3 функция X линейна по третьей переменной*:

$$X = p_0 + p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 + p_4 \cdot x_1 \cdot x_2 + p_5 \cdot x_1 \cdot x_3 + p_6 \cdot x_2 \cdot x_3 + p_7 \cdot x_1 \cdot x_2 \cdot x_3, \quad (2)$$

где p_i — постоянные вещественные коэффициенты.

Постулат 2. *В* четком *состоянии*[1] *субъект совершает* четкий *выбор* (1 *или* 0).

Далее В. А. Лефевр вводит две аксиомы, связывающие интенцию и поведение субъекта, находящегося в *четком* состоянии.

Аксиома 1. *Если интенция позитивна, то субъект всегда выбирает позитивный полюс, за исключением случая, когда он не ведает, что творит*[2].

Аксиома 2. *Если интенция негативна, то субъект всегда выбирает негативный полюс, за исключением случая, когда мир склоняет его к выбору позитивного полюса.*

Из аксиомы 1 и постулата 2 следует

$$f(0, 0, 1) = 1, \quad (3)$$
$$f(1, 0, 1) = 1, \quad (4)$$
$$f(1, 1, 1) = 1, \quad (5)$$
$$f(0, 1, 1) = 0. \quad (6)$$

Из аксиомы 2 и постулата 2 следует

$$f(0, 0, 0) = 0, \quad (7)$$
$$f(0, 1, 0) = 0, \quad (8)$$
$$f(1, 1, 0) = 1, \quad (9)$$
$$f(1, 0, 0) = 1. \quad (10)$$

Из (2) и (3)–(10) следует

$$p_0 + p_3 = 1, \quad (11)$$
$$p_0 + p_1 + p_3 + p_5 = 1, \quad (12)$$
$$p_0 + p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 = 1, \quad (13)$$
$$p_0 + p_2 + p_3 + p_6 = 0, \quad (14)$$
$$p_0 = 0, \quad (15)$$
$$p_0 + p_2 = 0, \quad (16)$$
$$p_0 + p_1 + p_2 + p_4 = 1, \quad (17)$$
$$p_0 + p_1 = 1. \quad (18)$$

[1] Под *четким* состоянием понимается ситуация, когда x_1, x_2, x_3 принимают значения либо 1, либо 0.

[2] Ситуация «не ведает, что творит» означает следующее: помыслы субъекта позитивны ($x_3 = 1$), он думает, что мир склоняет его к совершению добра ($x_2 = 1$), но не знает, что на самом деле мир склоняет его к совершению зла ($x_1 = 0$), и, делая то, к чему его склоняет мир, совершает зло ($X = 0$).

Откуда

$$p_0 = 0, \quad p_1 = 1, \quad p_2 = 0, \quad p_3 = 1, \quad p_4 = 0, \quad p_5 = -1, \quad p_6 = -1, \quad p_7 = 1. \tag{19}$$

Подставляя в (2), получаем выражение для функции готовности X:

$$X = x_1 + x_3 - x_1 \cdot x_3 - x_2 \cdot x_3 + x_1 \cdot x_2 \cdot x_3 = x_1 + x_3 \cdot (1 - x_1 - x_2 + x_1 \cdot x_2), \tag{20}$$

которая может быть представлена в виде

$$X = x_1 + x_3 \cdot (1 - x_1) \cdot (1 - x_2). \tag{21}$$

Таким образом, *готовность* субъекта к действию достаточно сложным образом зависит от его первоначальной *интенции*, от давления внешнего мира и от представления субъекта об этом давлении.

Особое внимание В. А. Лефевр уделяет ситуации *свободного выбора*, то есть ситуации, когда *готовность* субъекта (X) соответствует его *интенции* (x_3), то есть субъект действует в соответствии со своим первоначальным намерением. В рамках модели ситуация *свободного выбора* записывается в виде равенства

$$X = x_3. \tag{22}$$

Тогда, подставляя (22) в (21), получаем

$$X = \frac{x_1}{x_1 + x_2 - x_1 \cdot x_2} \quad \text{при } x_1 + x_2 > 0. \tag{23}$$

То есть субъект *свободен* и делает *реалистический* (по терминологии В. А. Лефевра) выбор, то есть действует в соответствии со своей интенцией, только в случае, если эта интенция согласуется с давлением среды следующим образом:

$$x_3 = \frac{x_1}{x_1 + x_2 - x_1 \cdot x_2} \quad \text{при } x_1 + x_2 > 0. \tag{24}$$

Формулу (24) В. А. Лефевр называет формулой человека и использует ее для анализа различных психологических феноменов во многих своих работах (см., например, [Лефевр, 1991; Лефевр, 2003а; Лефевр, 2005]). Однако следует отметить, что понимание им *свободного выбора* и *реалистичности* выглядит достаточно абстрактно. Из (24) следует, что *реалистичный* субъект должен сложным образом формировать свою интенцию (x_3), четко осознавая реальное давление на себя внешней среды (x_1) и при этом рефлексируя свое субъективное восприятие этого давления (x_2), которое может быть иным, чем x_1.[1] Возникает вопрос: если формирование субъектом своей интенции такое сложное, то в чем тогда заключается *свободность* его выбора?

Таким образом, предложенный В. А. Лефевром инструментарий математической формализации когнитивных и рефлексивных процессов, безусловно, представляет большой интерес, однако поведенческая ситуация, рассмотренная с помощью данного инструментария, представляется достаточно абстрактной, прежде всего в части сформулированной В. А. Лефевром **Аксиомы 2**. Формулировка данной аксиомы предполагает, что субъект всегда знает, к чему объективно склоняет его внешний мир (x_1), однако это очень часто не так, поскольку он может заблуждаться или его сознанием могут целенаправленно манипулировать, оказывая информационное воздействие.

[1] Здесь, впрочем, возникает вопрос: почему, если при формировании своей интенции x_3 в соответствии с формулой (24) субъект *осознает* реальное давление на него внешнего мира x_1, его субъективное восприятие этого давления x_2 не совпадает с x_1?

Модель деонтологического выбора

В рамках предложенного В. А. Лефевром методического подхода рассмотрим более реалистичные, на наш взгляд, формулировки аксиом. Будем считать, что в обществе задана общепринятая система моральных правил, определяющих понимание того, что «хорошо» (позитивный полюс) и что «плохо» (негативный полюс). Изначально интенция субъекта (x_3) может быть произвольной, но перед тем, как обратить свою интенцию в действие (X), он учитывает давление внешней среды, исходя из своего понимания этого давления (x_2).

При этом **Аксиома 1** сохраняется, а **Аксиома 2** изменяется на **Аксиому 2а**.

Аксиома 2а. *Если изначально интенция негативна* ($x_3 = 0$) *и субъект осознает это, сопоставляя свою интенцию с системой принятых в обществе моральных правил, то он следует тому, к чему, по его мнению, склоняет его внешний мир. Но при этом при* $x_2 = 1$ *и* $x_1 = 0$ *складывается ситуация, что субъект, думая, что внешний мир склоняет его к добру, реально делает зло, поскольку его представление о давлении внешнего мира искажено.*

В соответствии с **Аксиомой 2а** выражения (3)–(9) сохраняются, а выражение (10) принимает вид

$$f(1, 0, 0) = 0. \tag{10а}$$

Соответственно, сохраняются выражения (11)–(17), а выражение (18) принимает вид

$$p_0 + p_1 = 0. \tag{18а}$$

Откуда следует

$$p_0 = 0, \quad p_1 = 0, \quad p_2 = 0, \quad p_3 = 1, \quad p_4 = 1, \quad p_5 = 0, \quad p_6 = -1, \quad p_7 = 0. \tag{25}$$

Соответственно, функция X приобретает вид

$$X = x_3 + x_1 \cdot x_2 - x_2 \cdot x_3 = x_1 \cdot x_2 + x_3 \cdot (1 - x_2). \tag{26}$$

Ситуация *свободного выбора* (по В. А. Лефевру), когда *готовность* субъекта (X) соответствует его *интенции* (x_3), то есть субъект действует в соответствии со своим первоначальным намерением, в рамках модели записывается в виде равенства (22). Подставляя (22) в (26), получаем

$$X = x_3 = x_1. \tag{27}$$

То есть интенция (x_3) и готовность (X) совпадают, если интенции субъекта формируются в соответствии с реальным давлением среды (x_1), то есть в соответствии с реальными жизненными обстоятельствами (такого субъекта можно назвать *реалистичным*).

Еще одна ситуация, при которой выполняется равенство $X = x_3$, это когда $x_2 = 0$, то есть когда субъект убежден, что внешний мир склоняет его к злу. Тогда субъект отказывается подчиняться этому давлению и делает то, что он считает нужным без оглядки на мир (то есть к чему имеет изначальную интенцию x_3).

Рассмотренный выше вариант поведения субъекта под давлением внешнего мира можно характеризовать как *моральный конформизм*. Это подразумевает следующее: во-первых, в обществе имеются общепринятая система моральных принципов, понимание о правильном (моральном) и неправильном (аморальном) поведении; во-вторых, субъект корректирует свои первоначальные интенции с поправкой на моральное давление, которое, как ему представляется, оказывает на него внешняя среда (в этом, собственно, и выражается конформизм субъекта). При

этом первоначальные интенции x_3 могут формироваться на основе разнообразных мотиваций (в том числе утилитарных).

В более общем случае необходимо учесть то, что субъект часто следует давлению среды лишь частично, стремясь реализовать свою интенцию x_3 вопреки тому, к чему, как ему кажется, склоняет его внешний мир. С этой целью можно ввести коэффициент *оппортунизма* k, изменяющийся в интервале от 0 до 1. При $k = 0$ (оппортунизм по отношению к давлению внешнего мира отсутствует) готовность X определяется в соответствии с **Аксиомами 1** и **2а**. При $k = 1$ субъект поступает вопреки давлению внешнего мира x_2, реализуя свою интенцию.

В соответствии с этим выражения (3)–(7) сохраняются, а выражения (8)–(10) принимают вид

$$f(0, 1, 0) = k, \tag{8b}$$
$$f(1, 1, 0) = 1 - k, \tag{9b}$$
$$f(1, 0, 0) = 0. \tag{10b}$$

Соответственно, выражения (11)–(15) остаются неизменными, а выражения (16)–(18) принимают вид

$$p_0 + p_2 = k, \tag{28}$$
$$p_0 + p_1 + p_2 + p_4 = 1 - k, \tag{29}$$
$$p_0 + p_1 = 0. \tag{30}$$

Откуда следует

$$p_0 = 0, \quad p_1 = 0, \quad p_2 = 0, \quad p_3 = 1, \quad p_4 = 1 - k, \quad p_5 = 0, \quad p_6 = -1, \quad p_7 = k. \tag{31}$$

Выражение для функции X приобретает вид

$$X = x_3 + x_1 \cdot x_2 \cdot (1 - k) - x_2 \cdot x_3 + x_1 \cdot x_2 \cdot x_3 \cdot k = x_1 \cdot x_2 \cdot (1 - k) + x_3 \cdot (1 - x_2 + x_1 \cdot x_2 \cdot k). \tag{32}$$

При $k = 0$ выражение (32) совпадает с (26). При $k = 1$ выражение (32) приобретает вид

$$X = x_3 \cdot (1 - x_2 + x_1 \cdot x_2). \tag{33}$$

Ситуация *свободного выбора*, когда *готовность* субъекта (X) соответствует его *интенции* (x_3), то есть субъект действует в соответствии со своим первоначальным намерением, в рамках данной модели записывается в виде равенства (22). Подставляя (22) в (32), при $k = 0$ получаем выражение (27). В этом случае равенство $X = x_3$ выполняется в том случае, если субъект согласует свои интенции с тем, к чему его объективно подталкивает внешний мир. При $k = 1$ равенство $X = x_3$, как следует из (33), выполняется при условии

$$1 - x_2 + x_1 \cdot x_2 = 1, \quad \text{то есть} \quad x_2 \cdot (1 - x_1) = 0. \tag{34}$$

Это означает, что либо $x_1 = 1$, либо $x_2 = 0$. В первом случае внешний мир объективно склоняет субъекта к добру. Во втором случае субъект считает, что внешний мир склоняет его к злу, но не реагирует на это и действует по-своему (то есть в соответствии со своей первоначальной интенцией).

Результаты моделирования

1. На рис. 1–3 в соответствии с формулой (32) представлены профили *готовности* X как функции переменных x_1 и x_2 при разных значениях параметра k и переменной x_3. Профили представлены в виде тепловых карт, где значение X как функции переменных x_1 и x_2 отображается цветом, изменяемым в диапазоне от темно-красного (соответствующего значению 1,0) до темно-синего (соответствующего значению 0,0).

Видно, что функция X существенно нелинейна, что отражает сложный характер влияния x_1, x_2, x_3, k на поведение субъекта. Обращает на себя внимание то, что рис. 1, *в*; 2, *в*; 3, *в*

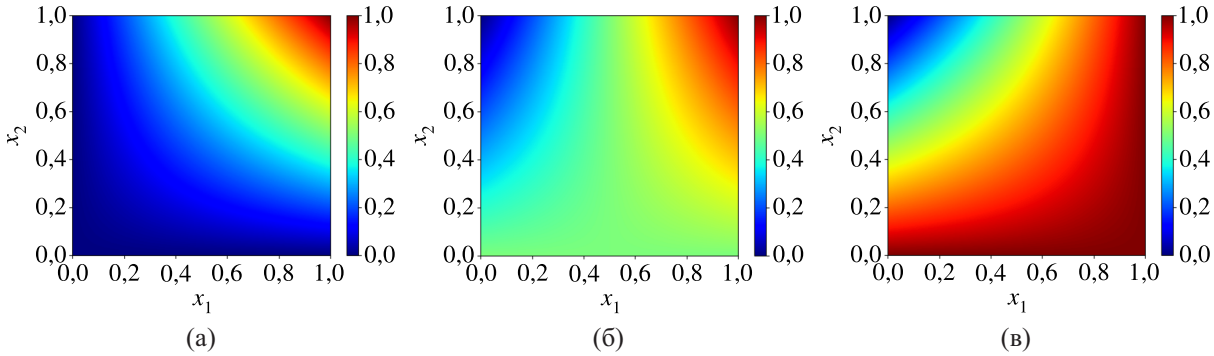


Рис. 1. Значения величины X как функции переменных x_1 и x_2 в соответствии с формулой (32) при $k = 0$ и при трех значениях переменной x_3: $x_3 = 0$ (а), $x_3 = 0{,}5$ (б) и $x_3 = 1$ (в)

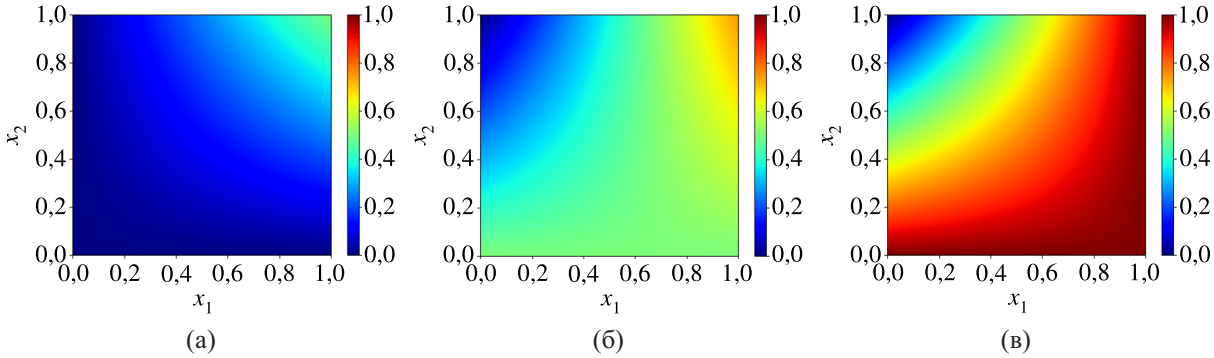


Рис. 2. Значения величины X как функции переменных x_1 и x_2 в соответствии с формулой (32) при $k = 0{,}5$ и при трех значениях переменной x_3: $x_3 = 0$ (а), $x_3 = 0{,}5$ (б) и $x_3 = 1$ (в)

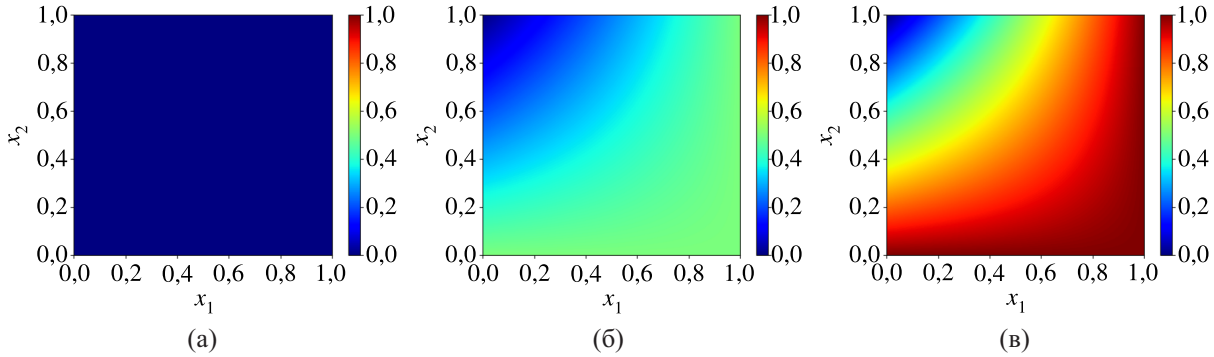


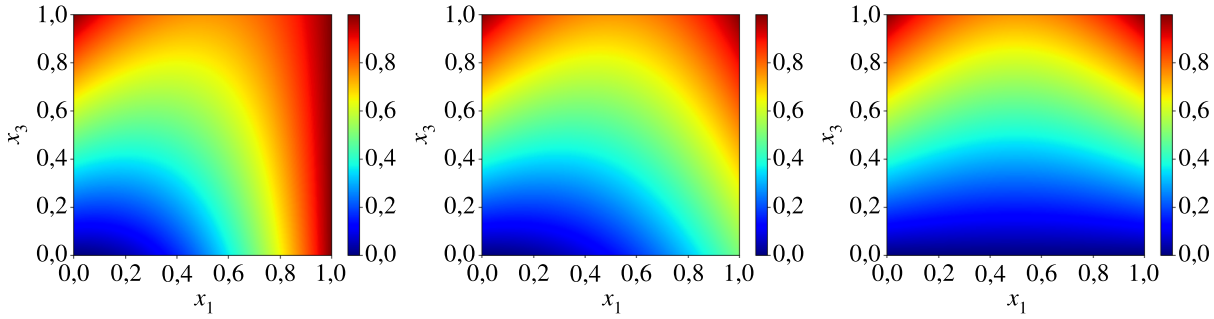
Рис. 3. Значения величины X как функции переменных x_1 и x_2 в соответствии с формулой (32) при $k = 1$ и при трех значениях переменной x_3: $x_3 = 0$ (а), $x_3 = 0{,}5$ (б) и $x_3 = 1$ (в)

идентичны, то есть если субъект изначально стремится делать добро ($x_3 = 1$), то его профиль X не зависит от значения k (такому субъекту можно дать условное название «*праведник*»).

2. В рамках модели можно рассматривать ситуации, когда переменные x_1, x_2, x_3, k каким-либо образом взаимосвязаны. Например, можно моделировать поведение *объективного* субъекта, то есть такого, который воспринимает давление внешнего мира неискаженно: $x_2 = x_1$. В этом случае из (32) следует

$$X = x_1^2 \cdot (1 - k) + x_3 \cdot \left(1 - x_1 + x_1^2 \cdot k\right). \tag{35}$$

Соответствующие профили X для разных значений k приведены на рис. 4.



(а) (б) (в)

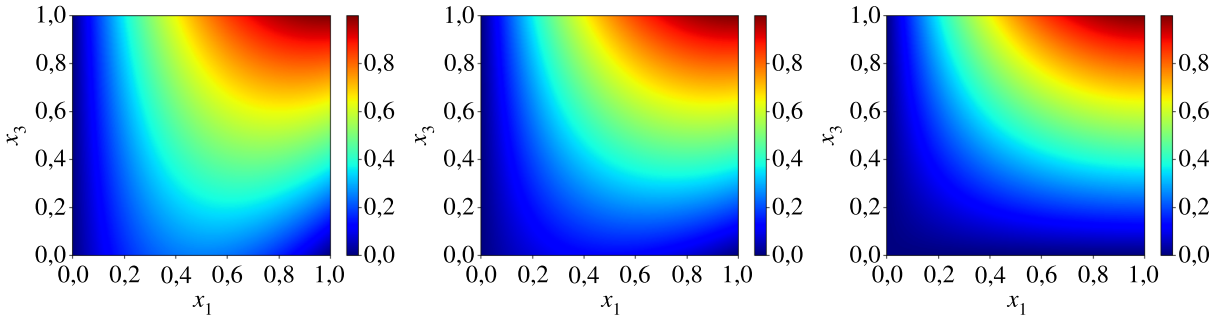
Рис. 4. Значения величины X как функции переменных x_1 и x_3 в соответствии с формулой (35) при разных значениях k: $k = 0$ (а), $k = 0{,}5$ (б) и $k = 1$ (в)

Видно, что если субъект является *объективным* оппортунистом ($x_2 = x_1$, $k = 1$), то действия этого субъекта практически всегда соответствуют его первоначальной интенции вне зависимости от давления внешнего мира (см. рис. 4, *в*). Однако если субъект склонен учитывать то, к чему его склоняет внешний мир ($k < 1$), то его действия в большей степени, чем при $k = 1$, начинают склоняться к положительному полюсу.

3. Также модель позволяет моделировать результаты *манипулирования* сознанием субъектов. Представляет интерес ситуация, когда субъектам целенаправленно внушают информацию о внешнем мире, обратную той, которая имеет место в действительности. На языке модели это означает, что $x_2 = (1 - x_1)$. В этом случае из (32) следует

$$X = x_1 \cdot (1 - x_1) \cdot (1 - k) + x_3 \cdot x_1 \cdot (1 + k - x_1 \cdot k). \tag{36}$$

Соответствующие профили X для разных значений k приведены на рис. 5.



(а) (б) (в)

Рис. 5. Значения величины X как функции переменных x_1 и x_3 в соответствии с формулой (36) при разных значениях k: $k = 0$ (а), $k = 0{,}5$ (б) и $k = 1$ (в)

Видно, что в случае манипулирования сознанием (или искаженного восприятия действительности) действия субъекта существенно сильнее отклоняются к негативному полюсу, чем в случае, когда субъект является *объективным* ($x_2 = x_1$, см. рис. 4). Интересно, что важен сам факт манипулирования (искаженного восприятия) безотносительно к тому, направлено оно в сторону отрицательного или положительного полюса. Также обращает на себя внимание то, что результаты манипулирования очень слабо зависят от степени *оппортунизма* субъекта (то есть от величины параметра k, см. рис. 5).

Обсуждение модели и результатов моделирования

Ниже представлен ряд комментариев к модели и к полученным на ее основе результатам.

1. В приведенном выше описании модели понятия добра и зла (позитивного и негативного полюсов) не были конкретизированы. В реальных жизненных ситуациях добро и зло конкретны, поэтому использование модели применительно к этим ситуациям требует доопределения указанных моральных понятий.

Так, если рассматривается ситуация, когда субъект вынужден сделать выбор: что приоритетнее — прагматизм (материальная выгода) или следование моральным принципам, то утилитарность можно считать негативным полюсом, а деонтологичнось — положительным.

Если рассматривается ситуация морального выбора в идеологизированном/религиозном обществе, то в нем понятия добра и зла определены нормативными/священными книгами и проблема выбора для субъекта заключается в том, поддаться ли искушению, нарушить моральные/религиозные требования (негативный полюс) или быть морально стойким (позитивный полюс).

2. Модель позволяет переводить на математический язык и интерпретировать явления, ранее подвергавшиеся лишь гуманитарному анализу.

Например, широко известна фраза, приписываемая Бисмарку: «Революции задумывают гении, осуществляют фанатики, а пользуются результатами отпетые негодяи». На языке модели гении — это те, кто в условиях социального кризиса предлагают обществу новую систему ценностей (новые понятия о добре и зле); фанатики — это те, кто в новой системе ценностей характеризуются интенцией $x_3 = 1$ и показателем *оппортунизма* $k = 1$; отпетые негодяи — это те, кто в новой системе ценностей характеризуются интенцией $x_3 = 0$ и показателем *оппортунизма* $k = 1$ (то есть это циники, которые поступают аморально, с выгодой для себя, пользуясь тем, что окружающие действуют морально, накладывая на себя самоограничения[1]).

3. Анализ рис. 1–3 показывает, что при $k < 1$ реальное поведение может быть моральнее интенции, то есть давление среды может делать поступки субъекта более моральными, чем это было первоначально в его интенции.

При этом для $k = 1$ из (33) следует $X \leqslant x_3$, то есть при $k = 1$ реальное поведение не может быть моральнее интенции. Другими словами, при $k = 1$ давление среды не может повысить моральный уровень поступков, однако в определенных случаях может его снизить.

4. Сравнение рис. 5 с рис. 4 показывает, что в случае манипулирования сознанием (или искаженного восприятия действительности) действия субъекта существенно сильнее отклоняются к негативному полюсу, чем в случае, когда субъект является *объективным* и неискаженно воспринимает давление внешнего мира ($x_2 = x_1$). Этот результат свидетельствует о вреде манипулирования (пусть даже с благими намерениями) и пользе правды, хоть и горькой.

[1] Такая ситуация в теории игр называется проблемой безбилетника.

5. Результаты моделирования позволяют сформулировать **закон снижения моральности общества**, состоящего из *свободных* субъектов (то есть таких, которые стремятся действовать в соответствии со своими интенциями: $X = x_3$). В этом случае выражение (32) преобразуется в

$$X = x_1 \cdot \frac{1-k}{1-x_1 \cdot k}. \tag{37}$$

Соответствующие графики величины X как функции параметра k для разных значений x_1 приведены на рис. 6.

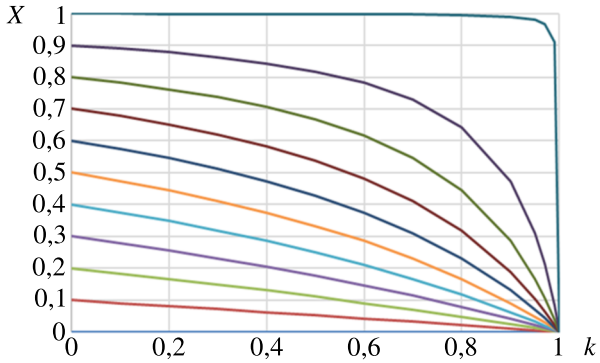


Рис. 6. Значения величины X как функции параметра k в соответствии с формулой (37) при разных значениях x_1

На рис. 6 каждый график соответствует определенному значению x_1. Значения x_1 изменялись от 0 до 1 с шагом 0,1 (в соответствии с формулой (37) при $k = 0$ выполняется равенство $X = x_1$, что позволяет определить, какому значению x_1 соответствует каждый график на рис. 6).

Видно, что во всех случаях величина X не превышает значения x_1 и монотонно уменьшается с увеличением значения k, снижаясь до нуля при $k = 1$. То есть *свободные* субъекты с высоким уровнем *оппортунизма* склонны поступать менее морально, чем тот уровень моральности, к которому склоняет их внешний мир. При этом если количество таких субъектов в обществе растет, то они сами начинают влиять на общественное мнение и отношение общества к моральным нормам, постепенно своим поведением снижая общий уровень моральности x_1. Возникает самосогласованный процесс неуклонного снижения моральности общества[1]. Остановить этот процесс можно только путем постоянного и целенаправленного поддержания обществом и государством высокого уровня x_1 (идеологическая работа, пропаганда традиционных ценностей, воспитательная работа в школе и т. п.), в противном случае общество *свободных* субъектов с неизбежностью со временем станет утилитарным, ориентирующимся при принятии решений исключительно на материальные факторы.

Заключение

1. В данной работе приведены результаты исследований по созданию математической модели морального выбора, основанной на развитии подхода, предложенного В. А. Лефевром. В отличие от В. А. Лефевра, который рассматривал ситуацию умозрительного рефлексирующего субъекта, выбирающего между абстрактными добром и злом, осознающего давление на него

[1] Нечто подобное мы наблюдаем в настоящее время в странах Запада в связи с распространением там движений ЛГБТ, БЛМ и т. п.

внешнего мира и рефлексирующего свое субъективное восприятие этого давления, в нашем исследовании рассмотрена более приземленная и практически значимая ситуация. Мы рассматриваем случай, когда субъект при принятии решений ориентируется на свое индивидуальное восприятие внешнего мира (которое может быть искаженным, например, вследствие внешнего целенаправленного информационного воздействия на субъекта и манипулирования его сознанием), а добро и зло не абстрактны, а обусловлены системой ценностей, принятой в конкретном рассматриваемом обществе и привязанной к конкретной идеологии/религии, которые могут быть различными для разных обществ и цивилизаций.

2. В результате проведенных исследований разработана базовая математическая модель, рассмотрены частные случаи ее применения. Выявлены некоторые закономерности, связанные с моральным выбором, приведено их формальное описание. В частности, на языке модели рассмотрена ситуация манипулирования сознанием, сформулирован закон снижения моральности общества, состоящего из так называемых *свободных* субъектов (то есть таких, которые стремятся действовать в соответствии со своими интенциями и соответствовать в своих действиях образу своего «я»).

3. В дальнейшем планируется использовать разработанный инструментарий для анализа конкретных ситуаций и выявления/объяснения макросоциальных закономерностей (например, для анализа закономерностей цивилизационных циклов: взлета и падения Римской империи, СССР, современной Западной цивилизации).

Список литературы (References)

Айзерман М. А., Алескеров Ф. Т. Выбор вариантов (основы теории). — М.: Наука, 1990.
Ajzerman M. A., Aleskerov F. T. Vybor variantov (osnovy teorii) [Choice of options (fundamentals of theory)]. — Moscow: Nauka, 1990 (in Russian).

Лефевр В. А. Алгебра совести. — М.: Когито-Центр, 2003.
Lefebvre V. A. Algebra sovesti [The algebra of conscience]. — Moscow: Kogito-Centr, 2003a (in Russian).

Лефевр В. А. Конфликтующие структуры. — М.: Издательство «Институт психологии РАН», 2000.
Lefebvre V. A. Konfliktuyushchie struktury [Conflicting structures]. — Moscow: Izdatel'stvo "Institut psikhologii RAN", 2000 (in Russian).

Лефевр В. А. Космический субъект. — Изд. 3-е доп. — М.: Когито-Центр, 2005.
Lefebvre V. A. Kosmicheskij sub"ekt [A cosmic subject]. — Moscow: Kogito-Centr, 2005 (in Russian).

Лефевр В. А. Рефлексия. — М.: Когито-Центр, 2003.
Lefebvre V. A. Refleksiya [Reflection]. — Moscow: Kogito-Centr, 2003b (in Russian).

Лефевр В. А. Формула человека: контуры фундаментальной психологии / пер. с англ. — М.: Прогресс, 1991.
Lefebvre V. A. Formula cheloveka: kontury fundamental'noj psikhologii [The human formula: contours of fundamental psychology]. — Moscow: Progress, 1991 (in Russian).

Малков С. Ю. Модель принятия решения в конфликтных ситуациях // Информационные войны. — 2008. — № 1. — С. 17–23.
Malkov S. Yu. Model' prinyatiya resheniya v konfliktnykh situaciyakh [The model of decision-making in conflict situations] // Information wars. — 2008. — No. 1. — P. 17–23 (in Russian).

Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. — М.: Физматлит, 2007.
Podinovskij V. V., Nogin V. D. Pareto-optimal'nye resheniya mnogokriterial'nykh zadach [Pareto-optimal solutions to multi-criteria problems]. — Moscow: Fizmatlit, 2007 (in Russian).

Подиновский В. В., Потапов М. А. Методы анализа и системы поддержки принятия решений: учебное пособие. — М.: Компания «Спутник плюс», 2003.
Podinovskij V. V., Potapov M. A. Metody analiza i sistemy podderzhki prinyatiya reshenij: uchebnoe posobie [Methods of analysis and decision support systems]. — Moscow: Kompaniya "Sputnik plyus", 2003 (in Russian).

Розен В. В. Математические модели принятия решений в экономике. — М.: Книжный дом «Университет», Высшая школа, 2002.
Rozen V. V. Matematicheskie modeli prinyatiya reshenij v ehkonomike [Mathematical models of decision-making in economics]. — Moscow: Knizhnyj dom "Universitet", Vysshaya shkola, 2002 (in Russian).

Соболь И. М., Статников Р. Б. Выбор оптимальных параметров в задачах со многими критериями. — М.: Дрофа, 2006.
Sobol' I. M., Statnikov R. B. Vybor optimal'nykh parametrov v zadachakh so mnogimi kriteriyami [Choosing optimal parameters in tasks with many criteria]. — Moscow: Drofa, 2006 (in Russian).

Соколов А. В., Токарев В. В. Методы оптимальных решений. — Т. 1 и 2. — 2-е изд., испр. — М.: Физматлит, 2011.
Sokolov A. V., Tokarev V. V. Metody optimal'nykh reshenij. — Vol. 1 and 2 [Methods of optimal solutions]. — Moscow: Fizmatlit, 2011 (in Russian).

Aleskerov F., Bouyssou D., Monjardet B. Utility maximization, choice and preference. — Berlin: Springer Verlag, 2007.

Baird A., Schuller B. Considerations for a more ethical approach to data in AI: on data representation and infrastructure // Frontiers in Artificial Intelligence. — 2020. — Vol. 3. — DOI: 10.3389/fdata.2020.00025

Brams S. J., Taylor A. Fair division. — New York: Cambridge University Press, 1996.

Clemen R. Making hard decisions: an introduction to decision analysis. — 2nd edition. — Belmont CA: Duxbury Press, 1996.

Keeney R. L., Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. — New York: Wiley, 1976.

Raiffa H. Decision analysis: introductory readings on choices under uncertainty. — McGraw Hill, 1997.

Roth A., Sotomayor M. O. Two-sided matching. — Cambridge: Cambridge University Press, 1990.

Saaty T. L. The analytic hierarchy process. — New York: McGraw Hill International, 1980.

Schramowski P., Turan C., Jentzsch S., Rothkopf C., Kersting K. The moral choice machine // Frontiers in Artificial Intelligence. — 2020. — Vol. 3. — https://doi.org/10.3389/frai.2020.00036

Smith J. Q. Decision analysis: a Bayesian approach. — Chapman and Hall, 1988.

Strimling P., Vartanova I., Jansson F. et al. The connection between moral positions and moral arguments drives opinion change // Nature Human Behaviour. — 2019. — No. 3. — P. 922–930.