

doi 10.31063/2073-6517/2018.15-3.3  
УДК 330.88

## ЭКОНОМИЧЕСКАЯ МОДЕЛЬ ЭФФЕКТА ВЫТЕСНЕНИЯ МОТИВАЦИИ К САМОКОНТРОЛЮ<sup>1</sup>

К. А. Паниди

*Недавние экономические исследования привлекли внимание к проблеме динамической несостоятельности предпочтений, которая описывает многие ситуации, связанные с явлением самоконтроля и силы воли. Одним из способов решения проблемы самоконтроля является вознаграждение более “рационального” поведения. Однако исследования в области психологии показывают, что такое вознаграждение, несмотря на краткосрочное улучшение ситуации, часто может приводить к потере мотивации к самоконтролю в будущем. Вытеснение мотивации характерно в основном для ситуаций, где вознаграждение назначается из внешних источников, будь то другой человек или организация, и реже происходит, когда речь идет о вознаграждении, назначаемом человеком самому себе. Данная работа представляет экономическую модель, объясняющую различия в краткосрочных и долгосрочных эффектах внутреннего и внешнего вознаграждения. Модель использует дуальный (dual-self) подход к проблеме самоконтроля и основана на предпосылке о неполноте информации об издержках самоконтроля. Основной вывод модели заключается в том, что наиболее подвержены эффекту вытеснения мотивации будут те индивиды, у которых уровень уверенности в своих способностях к самоконтролю достаточно низок, в то время как на наиболее уверенных в себе индивидов внешнее вознаграждение не окажет такого эффекта, что согласуется с данными эмпирических исследований.*

**Ключевые слова:** проблема самоконтроля, динамическая несостоятельность предпочтений, эффект вытеснения мотивации

### 1. Введение

Недавние экономические исследования обращают внимание на проблему динамической несостоятельности предпочтений, или, иначе говоря, проблему самоконтроля. В классической экономической теории предполагается, что агент дисконтирует будущую полезность экспоненциально. Это означает, что если агенту предложить выбор из двух вариантов, например, получить 250 долларов сегодня или 300 долларов через 1 месяц, и агент выбирает первый вариант, то при сдвиге обеих альтернатив на один и тот же временной промежуток (например, 250 долларов через год и 300 долларов через год и 1 месяц), он не должен менять свой выбор. Это обусловлено предположением о том, что ставка дисконтирования определяется только лишь временным интервалом между опциями, но не тем, насколько эти опции отделены от текущего момента. Эксперимент, проведенный лауреатом премии памяти Альфреда Нобеля по экономике 2017 года Ричардом Талером, показал, что для многих людей это предположение не выполняется, и их предпочтения более точно описываются гиперболическим, а не экспоненциальным дисконтированием.

В реальном поведении гиперболическое дисконтирование проявляется, например, в том, что многие люди систематически предпочитают увеличить сиюминутное потребление, отказавшись от более значительных выгод, отдаленных по времени от текущего момента. Однако когда обе опции предлагаются в будущем, индивиды меняют свой выбор в пользу более отдаленной по времени опции. Эти наблюдения свидетельствуют о том, что потребление в текущий момент имеет непропорционально высокий вес по сравнению с любым будущим моментом, что и порождает проблему динамической несостоятельности предпочтений [2, 12, 15, 23].

Эмпирические исследования показали, что динамическая несостоятельность предпочтений может лежать в основе различных форм нерационального поведения, таких как курение и злоупотребление алкоголем, неспособность соблюдать диету, низкий контроль за уровнем сахара в крови у диабетиков, прокрастинация [9, 21, 25]. Одним из популярных методов борьбы с проблемой самоконтроля является предложение вознаграждения за более выгодное в долгосрочном плане поведение. При этом вознаграждение может быть как внешним, то есть исходить от других людей или организаций, так и внутренним, то есть назначаться человеком самому себе. Однако многие исследования

<sup>1</sup> © Паниди К. А. Текст. 2018.

в области психологии указывают на то, что такое вознаграждение может иметь неоднозначные последствия для самоконтроля. В частности, оно может повышать краткосрочную мотивацию к самоконтролю, однако как только вознаграждение перестает выплачиваться, эта мотивация снижается до уровней иногда даже ниже, чем изначальный. В данной работе приводится теоретическая модель, которая позволяет объяснить это явление с точки зрения поведенческой экономики.

В работе [19] приводится подробный обзор 41 исследования, в которых участникам предлагается денежное вознаграждение в рамках различных программ, стимулирующих здоровый образ жизни. Вознаграждения могут быть в форме прямых денежных выплат, подарочных сертификатов или лотерейных билетов. Также они могут выплачиваться просто за участие в программе или же за достижение определенных целей. Некоторые исследования указывают на то, что наличие вознаграждения повышает шансы на участие в программе [3, 16, 18]. Положительные эффекты от вознаграждения наблюдались в программах по прекращению курения и кокаиновой зависимости. Однако эти же исследования документируют высокий процент участников, вернувшихся к прежнему поведению после окончания программы.

Несколько исследований дают более оптимистичный взгляд на проблему [10, 27, 28]. Результаты в них показывают существенное повышение мотивации для некоторых участников после окончания программ по снижению веса и отказу от курения по сравнению с контрольной группой, которая не получала материального вознаграждения. Интересный эксперимент представлен в работе [24]. В нем школьникам выплачивалось вознаграждение при условии получения хороших оценок по математике по итогам года. Затем их успехи отслеживались в течение следующих трех лет, но уже без вознаграждения. Оказалось, что вытеснение мотивации присутствовало у учеников с низкими математическими способностями, в то время как для учеников с высокими способностями наблюдался обратный эффект.

Эмпирические исследования эффектов «внутреннего» вознаграждения, то есть того, которое индивид назначает сам себе, показывают, что эффект от них может быть более устойчивым в долгосрочной перспективе по сравнению с внешним вознаграждением. Исследование [26] показало, что результаты людей, которые прилагали усилия для отказа от алкоголя и имели при этом высокую внутрен-

нюю мотивацию, оказывались более устойчивыми, чем тех, чья мотивация была преимущественно внешней. Аналогичные результаты приводятся в работах [29, 30 и 7].

Несмотря на то, что существуют экономические модели, позволяющие объяснить эффект вытеснения мотивации, сравнение внутреннего и внешнего типов вознаграждения в них как правило отсутствует. Работа [4] предлагает одно из объяснений того, почему предложение более высокого уровня вознаграждения может приводить к снижению мотивации. В этой работе модель вида «принципал — агент» основана на предположении о том, что принципал обладает частной информацией о характеристиках агента, например, о его способностях, которая не доступна самому агенту. Исходя из этой информации принципал назначает вознаграждение, которое в результате служит сигналом о способностях агента. Если агент наблюдает высокое вознаграждение, он делает вывод о том, что вероятно его способности к выполнению работы низки (так как только работникам с низким уровнем способностей требуется высокое вознаграждение), что снижает его мотивацию к выполнению этой работы. В модели [17] предлагается альтернативное объяснение: здесь наличие мотивации является переменной, которую агент выбирает оптимальным для себя образом. В результате при определенном виде контракта работнику становится выгодно иметь низкую внутреннюю мотивацию, так как она компенсируется денежной.

Работа [22] анализирует механизм, в котором самостоятельно назначаемое вознаграждение за достижение цели служит сигналом о взятых на себя обязательствах по достижению этой цели. Делается предположение о том, что негативные отклонения от цели человек воспринимает более остро, чем позитивные, что стимулирует человека прилагать больше усилий.

В данной работе приводится модель, которая позволяет сравнить краткосрочные и долгосрочные эффекты вознаграждения на мотивацию к самоконтролю. Модель основана на предположении о том, что агент не знает своего уровня издержек, который ассоциирован с усилиями самоконтроля, однако он может их выявить, если эти усилия приложит. Показано, что эффект вытеснения и повышения мотивации может наблюдаться при вознаграждении обоих типов, но будет зависеть от степени уверенности человека в своих способностях к самоконтролю. Модель таким образом помогает

объяснить противоречивые эмпирические результаты, описанные выше.

Модель основана на дуальном подходе к проблеме самоконтроля (*dual-self*), предложенном в работе [11]. Согласно этому подходу, действия индивида в каждом периоде являются результатом равновесия в игре двух его «субличностей», одна из которых, долгосрочная, является принципалом, а другая, краткосрочная, — агентом. Краткосрочная субличность заинтересована только лишь в текущем потреблении и не заботится о возможных долгосрочных негативных последствиях этого потребления. Каждый период индивид характеризуется новой краткосрочной субличностью. Долгосрочный принципал, напротив, в каждом периоде заинтересован в том, чтобы ограничить потребление вредного продукта, и максимизирует суммарную полезность от всех периодов. Как отмечается в [1], такой подход к моделированию проблемы самоконтроля удобен в первую очередь с математической точки зрения, так как часто позволяет получить равновесие в явном виде.

В оригинальной модели [11] принципал добивается снижения потребления тем, что ограничивает доступ краткосрочного агента к продукту (например, если целью принципала является снизить текущие траты в пользу более высокого потребления в будущем, то он может ограничить количество доступных агенту наличных денег). В данной же работе предполагается, что принципал действует иначе — он может стимулировать более «рациональное» поведение агента, предложив ему вознаграждение за успешный отказ от потребления. В результате в некоторых случаях агенту будет выгодно отказаться от потребления в текущем периоде.

Работа построена следующим образом. В разделе 2 приводится общая структура модели и описание основных ее предпосылок. В разделе 3 представлен анализ равновесий модели и описываются условия, при которых может наблюдаться эффект вытеснения мотивации. Основные выводы работы суммируются в разделе 4.

## 2. Теоретическая модель

### 2.1. Общая структура модели

В работе рассматривается двухпериодная дуальная модель самоконтроля. Индивид живет в течение двух периодов. Его поведение в каждом периоде описывается как результат равновесия в игре долгосрочного принципала, целью которого является максимизация полез-

ности, суммарной за два периода, и двух краткосрочных агентов. Каждый из краткосрочных агентов существует только в одном периоде и заинтересован только лишь в максимизации текущей полезности. Каждый из агентов может выбрать потребить вредный продукт и получить от этого потребления некоторую полезность, либо отказаться от потребления, приложив для этого усилия, сопровождающиеся издержками. Долгосрочный принципал имеет возможность предложить агенту бонус за отказ от потребления вредного продукта. Поскольку целью исследования является сравнение эффектов внутренней и внешней мотивации агента, будут введены два типа долгосрочного принципала — внутренний и внешний. Различия между ними описаны ниже.

#### 2.1.1 Действия игроков

Для того чтобы стимулировать краткосрочного агента отказаться от потребления вредного продукта, принципал может предложить ему вознаграждение. В каждом периоде внутренний принципал принимает решение о размере этого вознаграждения  $b^{SP} \geq 0$ .

Также, при наличии внешней мотивации в дополнение к внутренней, индивид может получить вознаграждение и от внешнего принципала  $b^{EP} \geq 0$ . Оба принципала заинтересованы в том, чтобы стимулировать более высокий уровень усилий агента, однако разница между ними заключается в том, что внешний принципал имеет возможность предложить бонус только в первом периоде игры, тогда как внутренний может это делать в обоих периодах. Это различие обусловлено тем, что в эмпирических исследованиях внешняя мотивация, как правило, ограничена во времени (например, предоставляется индивиду только тогда, когда он участвует в программе снижения веса или отказа от курения), в то время как внутренняя доступна индивиду в любой момент. Другое важное различие принципалов заключается в том, что внутренний принципал имеет не только материальную, но и эмоциональную компоненту в своей функции полезности, что будет описано ниже.

В каждом периоде краткосрочный агент принимает решение о том, какой уровень усилий стоит приложить исходя из предложенных принципалами бонусов. При этом уровень усилий  $e_t$ ,  $t \in \{1, 2\}$  является бинарным, то есть может принимать только значения из множества  $\{0, 1\}$ .

Таблица суммирует переменные, которые могут выбирать агенты и принципалы в модели.

Выбор переменных в модели

	Период 1	Период 2
Внутренний принципал	Бонус в первом периоде $b_1^{SP} \geq 0$	Бонус во втором периоде $b_2^{SP} \geq 0$
Внешний принципал	Бонус $b^{EP} \geq 0$	—
Агент 1	Уровень усилий $e_1 \in \{0, 1\}$	—
Агент 2	—	Уровень усилий $e_2 \in \{0, 1\}$

### 2.1.2. Типы игроков

Если агент прилагает нулевые усилия, то издержки от них также равны нулю. Если же агент выбирает положительный уровень усилий, то он несет издержки  $k \geq 0$ . В модели предполагается, что в первом периоде агент не знает в точности, с каким уровнем издержек от усилий ему придется столкнуться. Это предположение обусловлено тем, что в реальной жизни индивид может не знать, какой уровень издержек он понесет до того, как совершит то или иное действие. Например, ему может быть неизвестно в точности, какой уровень дискомфорта доставят попытки отказаться от курения, если индивид никогда не пробовал это делать. Агенты могут принадлежать к одному из двух типов. «Сильный» тип сталкивается с низким уровнем издержек  $k = k_L$ , тогда как «слабый» тип, напротив, имеет высокий уровень издержек  $k = k_H > k_L$ . Без потери общности можно предположить, что  $k_L = 0$ . Также предположим, что  $k_H > 1$ . Как будет показано далее, если это предположение выполнено, то агент, обнаруживший после первого периода, что его издержки слишком высоки, не будет прикладывать усилий для отказа от вредного продукта в будущем. Модель предполагает, что тип, к которому принадлежит агент, определяется природой до начала игры и не меняется между периодами.

### 2.1.3. Веры игроков

В начале первого периода типы агентов неизвестны, однако внутренний и внешний принципалы имеют априорные веры относительно распределения этих типов. В частности, предполагается, что агенты сильного типа встречаются в популяции с вероятностью  $(1 - p)$ , тогда как агенты слабого типа встречаются с вероятностью  $p \in [0, 1]$ . Поскольку внутренний принципал и каждый из краткосрочных агентов по сути являются субличностью одного и того же индивида, логично предположить, что они имеют одинаковые веры относительно того, к какому типу относится агент.

Если агент приложил усилия в первом периоде, то он полностью выявляет уровень издержек, с которым он сталкивается, и таким образом в точности узнает, к какому типу он

принадлежит. Эта информация может влиять на его поведение во втором периоде. Если же агент выбирает не прикладывать усилий в первом периоде, то он не получает новой информации об издержках и во втором периоде по-прежнему не знает, к какому типу он относится.

### 2.1.4. Предпочтения игроков

Необходимо описать предпочтения агентов и принципалов в модели.

Ситуацию, когда индивид имеет как внутреннюю, так и внешнюю мотивацию к самоконтролю, можно формализовать, предположив, что агент первого периода получает некоторый положительный бонус от внутреннего и внешнего принципалов. Таким образом, функция полезности агента первого периода будет иметь следующий вид:

$$U_1^A = (b_1^{SP} + \gamma b^{EP} - k)e_1, \quad (1)$$

где параметр  $k \in \{k_L, k_H\}$  означает издержки самоконтроля,  $e_1 \in \{0, 1\}$  представляет собой уровень усилий, выбранный в первом периоде, а коэффициент  $\gamma \geq 0$  показывает, насколько агент чувствителен к внешней мотивации по сравнению с внутренней.

Как было указано ранее, во втором периоде внешний принципал отсутствует, поэтому полезность агента второго периода определяется только вознаграждением, которое предлагает внутренний принципал, и его уровнем усилий:

$$U_2^A = (b_2^{SP} - k)e_2. \quad (2)$$

Определим функции полезности принципалов. Внешний принципал заботится о том, чтобы снизить потребление вредного продукта как сейчас, так и в будущем, однако имеет возможность влиять на поведение агента только в первом периоде. Если агент отказался от потребления, принципал получает полезность, равную усилиям агента, за вычетом бонуса:

$$U^{EP} = (1 - b^{EP})e_1 + e_2. \quad (3)$$

Рассмотрим функцию полезности внутреннего принципала. Обозначим исходы, которые получает внутренний принципал в первом и втором периодах, как  $x_1$  и  $x_2$  соответственно.

Тогда математическое ожидание исхода второго периода с позиции первого периода обозначим как  $Ex_2$ . Исход, который получает внутренний принципал в первом периоде, составляет либо 0 (если агенту не удалось воздержаться от потребления), либо  $(1 - b_1^{SP})$ . Кроме того, принципал понимает, что если агент приложит положительные усилия, это позволит ему выявить и уровень издержек от этих усилий, что может сказаться на мотивации прикладывать усилия в будущем. Если эти издержки окажутся высокими, это может демотивировать агента второго периода и тем самым разочаровать принципала. Таким образом, внутренний принципал имеет не только материальную составляющую в своей полезности, но и эмоциональную, а именно, он испытывает чувство потери, если полученный им исход отклоняется от ожидаемого. Это предположение является стандартным для многих моделей поведенческой экономики (см. например, [14]). Оно основывается на работах когнитивных психологов Дэниела Канемана и Амоса Тверски [20], которые в ряде экспериментов показали, что свойство избегания потерь характерно для целого ряда экономических решений людей. Д. Канеман и А. Тверски продемонстрировали, что эмоциональная реакция людей на потери примерно в 2–2,5 раза сильнее, чем на аналогичные по абсолютной величине выигрыши<sup>1</sup>.

С учетом этого предположения полезность внутреннего принципала от обоих периодов можно записать следующим образом:

$$\begin{aligned} U^{SP} &= x_1 + E[x_2 + v(x_2)], \\ x_1 &= (1 - b_1^{SP}) e_1, \\ x_2 &= (1 - b_2^{SP}) e_2, \\ v(x_2) &= \begin{cases} (x_2 - E \cdot x_2), & \text{if } x_2 \geq E \cdot x_2, \\ \lambda(x_2 - E \cdot x_2), & \text{if } x_2 < E \cdot x_2, \end{cases} \end{aligned} \quad (5)$$

где  $\lambda \geq 1$  является коэффициентом избегания потерь. Первое слагаемое представляет собой материальный исход первого периода, тогда как второе слагаемое является математическим ожиданием исхода второго периода, принимающим во внимание материальный исход  $x_2$  и эмоциональный исход  $v(x_2)$ . Последний определяется положительным или отрицательным отклонением материального исхода от его математического ожидания. Если усилия первого

периода показали, что агент принадлежит к сильному типу, принципалу становится понятно, что во втором периоде агенту будет легко воздержаться от потребления вредного продукта и большой бонус не потребуются. Однако если агент окажется слабого типа, принципалу придется понести существенные потери в будущем, так как такого агента может мотивировать только значительный бонус. Таким образом, принципал сталкивается со следующим выбором. С одной стороны, мотивировать агента в первом периоде ему выгодно, так как это повышает шансы на положительный исход уже сейчас. С другой стороны, получение информации о реальных издержках может привести к разочарованию и демотивации в будущем.

### 3. Анализ равновесий

Для начала проанализируем игру только с внутренним принципалом. Рассмотрим задачу, которую решает агент в первом периоде. Если агент не знает, к какому типу он принадлежит, то единственное доступное ему предположение о его уровне издержек — это его математическое ожидание, равное  $pk_H$ . В первом периоде агент максимизирует ожидаемую полезность, заданную формулой (1), где  $b^{EP} = 0$ , так как внешний принципал пока отсутствует:

$$EU_1^A = b_1^{SP} - \hat{k}, \quad (6)$$

где  $\hat{k} = pk_H$ . Оптимальный для агента уровень усилий описывается следующим образом:

$$e_1 = \begin{cases} 0, & \text{if } b_1^{SP} < \hat{k}, \\ 1, & \text{if } b_1^{SP} \geq \hat{k}. \end{cases} \quad (7)$$

Если после первого периода новой информации об уровне усилий у агента нет, то оптимальный выбор усилий во втором периоде аналогичен. Если же в первом периоде был выбран положительный уровень усилий, то во втором периоде оптимальные усилия описываются этой же формулой с заменой  $\hat{k}$  на тот уровень издержек, который был выявлен.

Рассмотрим теперь задачу принципала. Поскольку принципал существует два периода, необходимо решить его задачу методом обратной индукции. Оптимальный бонус во втором периоде будет зависеть от двух факторов: от того, были ли выявлены истинные издержки в первом периоде и, если не были, то от уровня самоуверенности, описываемого параметром  $p$ . Если агент оказался слабого типа, то во втором периоде только более высокий бонус ( $b^{EP} > 1$ ) может стимулировать его прикладывать усилия, что невыгодно принципалу,

<sup>1</sup> В дальнейшем это предположение было использовано для объяснения многих экономических феноменов, таких как «загадка премии за риск» (*equity premium puzzle*), эффект владения, избыточное избегание риска и другие (см., например, [5, 13, 20]).

так как его максимальный выигрыш от усилий составляет 1. В этом случае агент второго периода выбирает нулевые усилия, а принципал выбирает нулевое вознаграждение. Если же обнаружилось, что агент сильного типа, тогда во втором периоде он выберет положительные усилия и при нулевом бонусе, а принципалу нет смысла платить больше, чем ноль.

Если после первого периода издержки от усилий остались неизвестны, тогда во втором периоде принципал будет стимулировать агента, только если он достаточно оптимистичен относительно его уровня издержек, то есть если вероятность иметь высокие издержки ниже определенного порогового значения  $p < \bar{p} = 1/k_H$ . В противном случае агент потребует слишком высокий для принципала бонус в размере  $pk_H > 1$ .

Для того чтобы определить равновесие в первом периоде, нам необходимо сравнить ожидаемую полезность (согласно формуле (5)) внутреннего принципала от решения платить минимально необходимый для усилий бонус и решения не платить его. Предположим, агент выбрал нулевые усилия в первом периоде. Если  $p < \bar{p} = 1/k_H$ , во втором периоде принципал выберет положительный бонус и агент приложит положительные усилия. В равновесии этот бонус составит  $\hat{k}$ , поэтому полезность принципала второго периода составит  $U_2^{SP}(e_1 = 0) = 1 - b_1^{SP} = 1 - \hat{k}$ . Если же  $p \in (\bar{p}, 1]$ , принципал не может позволить себе платить необходимый бонус и получает нулевую полезность.

Предположим теперь, что агент выбрал  $e_1 = 1$ . Это создает у принципала ожидания относительно полезности второго периода. Если окажется, что агент слабого типа, принципал получит полезность 0. Если же окажется, что он сильного типа, то принципал может рассчитывать на полезность 1. Таким образом, математическое ожидание его полезности составит  $0 \cdot p + 1 \cdot (1-p) = (1-p)$ . Суммируя все вышесказанное и принимая во внимание выражение (5), запишем ожидаемую полезность внутреннего принципала в первом периоде:

$$EU_1^{SP}(e_1 = 1) = (1 - \hat{k}) + (-\lambda p(1-p) + (1-p)(1+p)). \quad (8)$$

**Утверждение 1.** Существует такое значение  $p^* \in (0, 1)$ , что в равновесии верно следующее:

(1) Внутренний принципал предлагает бонус  $b_1^{SP} = pb_H$  для любых  $p \leq p^*$  и платит ноль для любых  $p > p^*$ .

(2) При  $1 < \lambda \leq k_H + 1$ ,

$$p^* = \frac{\lambda + k_H - \sqrt{(\lambda + k_H)^2 - 8(\lambda - 1)}}{2(\lambda - 1)} > \frac{1}{k_H}.$$

$$(3) \text{ При } \lambda > k_H + 1, p^* = \frac{1}{\lambda - 1} < \frac{1}{k_H} = \bar{p}.$$

(4) Пороговое значение  $p^*$  убывает по  $\lambda$ .

Это утверждение означает, что внутренний принципал мотивирует агента только если его уровень самоуверенности достаточно высок (то есть  $p$  ниже, чем  $p^*$ ). Также оно означает, что чем больше принципал избегает потерь, тем более уверен в себе он должен быть для того, чтобы мотивировать агента. Интуитивно этот результат можно понять следующим образом: чем выше коэффициент избегания потерь  $\lambda$ , тем больше принципал заинтересован в сохранении высокой уверенности в своих силах в будущем. Это в свою очередь сужает область значений  $p$ , для которых принципал готов получить информацию об издержках.

Следующее утверждение описывает равновесный уровень усилий в первом и втором периодах. Поскольку реальный уровень издержек заранее неизвестен, мы можем описать лишь ожидаемый уровень усилий второго периода  $\bar{e}_2$ .

**Утверждение 2.**

(i) При  $1 < \lambda \leq k_H + 1$ :

если  $p \in [0, p^*]$ , то  $b_1^{SP} = pk_H, b_2^{SP} = 0, e_1 = 1, \bar{e}_2 = 1 - p$  (вытеснение мотивации);

если  $p \in (p^*, 1]$ , то  $b_1^{SP} = 0, b_2^{SP} = 0, e_1 = 0, \bar{e}_2 = 0$ .

(ii) При  $\lambda > k_H + 1$ :

если  $p \in \{0, p^*\}$ , то  $b_1^{SP} = pk_H, b_2^{SP} = 0, e_1 = 1, \bar{e}_2 = 1 - p$  (вытеснение мотивации);

если  $p \in (p^*, 1/k_H]$ , то  $b_1^{SP} = 0, b_2^{SP} = pk_H, e_1 = 0, \bar{e}_2 = 1$  (повышение мотивации);

если  $p \in (1/k_H, 1]$ , то  $b_1^{SP} = 0, b_2^{SP} = 0, e_1 = 0, \bar{e}_2 = 0$ .

Внутренний принципал сталкивается с выбором. С одной стороны, он может стимулировать усилия в первом периоде с помощью бонуса, что даст ему немедленный положительный выигрыш. С другой стороны, при определенных условиях это может привести к вытеснению мотивации в будущем, если окажется, что издержки от усилий слишком высоки. Утверждение 2 описывает условия, при которых может произойти вытеснение мотивации. В частности, оно происходит именно для тех индивидов, которые изначально имеют достаточно низкий уровень уверенности. Принципал решает стимулировать усилия такого агента с помощью бонуса, что приводит к высоким усилиям в первом периоде, и к потере мотивации во втором периоде, если издержки также оказались высокими.

Теперь рассмотрим поведение игроков, когда в первом периоде присутствует не только внутренний, но и внешний принципал. В этом

случае внутренний принципал решает свою оптимизационную задачу, принимая уровень бонуса внешнего принципала как заданный. Зная оптимальный ответ на свои действия со стороны внутреннего принципала, внешний также решает свою задачу максимизации полезности.

Решив модель в этом случае, мы получаем два возможных равновесия. В первом равновесии внутренний принципал не стимулирует агента в первом периоде, однако высокий уровень уверенности агента позволяет ему прикладывать высокие усилия во втором периоде. Поскольку внешний принципал присутствует только в первом периоде, он не заботится о сохранении высокой уверенности и поэтому может выбрать высокий бонус, который заставит агента первого периода приложить высокие усилия. В этом случае может иметь место вытеснение мотивации для тех агентов, которые принадлежат к слабому типу.

Во втором равновесии уровень уверенности агента низок, поэтому без внешнего вознаграждения агент выбирает нулевые усилия в обоих периодах. Внешний принципал решает предложить бонус, в результате чего агент прикладывает усилия, и для индивидов сильного типа будет наблюдаться вытеснение мотивации, тогда как для индивидов слабого типа, напротив, мотивация во втором периоде возрастет.

Обозначим

$$C = \min \left( \frac{k_H}{2k_H - 1} \left( \frac{\lambda}{k_H} - 1 - \frac{\lambda - 1}{k_H^2} \right), k_H - 1 \right).$$

Утверждение 3 описывает условия, которые должны быть выполнены, чтобы наблюдалось вытеснение или повышение мотивации.

**Утверждение 3.**

i) При  $\lambda < 1 + k_H$ :

если  $\gamma \leq k_H - 1$ , внешний принципал предлагает ненулевой бонус при  $p \in [p^*, 1]$ , агент выбирает усилия, равные 1, ожидаемый уровень усилий второго периода равен  $\bar{e} = 1 - p$ ;

если  $\gamma < k_H - 1$ , то существует такое значение  $p^\gamma$ , при котором для  $p \in [p^*, p^\gamma]$  внешний принципал предлагает ненулевой бонус, агент выбирает усилия, равные 1, ожидаемый уровень усилий во втором периоде составит  $\bar{e} = 1 - p$ . Если  $p \in (p^\gamma, 1]$ , оба принципала предлагают нулевой бонус.

ii) При  $\lambda \leq 1 + k_H$ :

если  $\gamma > 0$ , существует такое значение  $p^\gamma$ , при котором для  $p \in [p^*, p^\gamma]$  внешний принципал предлагает ненулевой бонус, агент выбирает

$e_1 = 1$ , ожидаемый уровень усилий второго периода составляет  $\bar{e}_2 = 1 - p$ ;

если  $\gamma \leq C$ , тогда для любых  $p \in [1/k_H, 1]$  оба принципала предлагают нулевой бонус в периоде 1;

если  $\gamma > C$ , тогда существует интервал внутри  $p \in [1/k_H, 1]$ , в котором внешний принципал предлагает ненулевой бонус, агент выбирает  $e_1 = 1$ , ожидаемый уровень усилий второго периода равен  $\bar{e}_2 = 1 - p$ .

Доказательство этого утверждения приведено в Приложении. Интуитивно это утверждение можно объяснить следующим образом. Решение внешнего принципала о вознаграждении зависит от двух факторов: от того, предлагается ли бонус внутренним принципалом, а также того, насколько агент чувствителен к бонусу внешнего принципала (параметр  $\gamma$ ). Возможны три различных случая:

1) Если уровень уверенности агента высок, то внутренний принципал предлагает бонус в первом периоде, что дает возможность внешнему принципалу сэкономить и не платить вознаграждение. В этом случае присутствие внешнего принципала никак не влияет на мотивацию агента как в первом, так и во втором периоде.

2) Если уровень уверенности агента относительно низок, и одновременно агент достаточно чувствителен к внешней мотивации (например,  $\gamma > C$  или  $\gamma \geq k_H - 1$ ), тогда внешнее вознаграждение существенно влияет на решение агента отказаться от потребления вредного продукта. Оба принципала предлагают вознаграждение.

3) Если чувствительность агента к внешней мотивации  $\gamma$  слишком низка, агент потребует слишком высокое вознаграждение за свои усилия, что невыгодно для внешнего принципала. В результате он не получает бонуса.

В модели предполагается, что внешний принципал имеет возможность наблюдать уровень внутренней мотивации агента (т. е. предлагаемое внутренним принципалом вознаграждение). Очевидно, что в реальности это не всегда так. Например, индивид может участвовать в программе, стимулирующей здоровый образ жизни, в которой вознаграждение фиксировано и не зависит от его уровня внутренней мотивации. В этой ситуации решение модели принципиально не меняется. Если фиксированный бонус будет слишком мал, это аналогично равновесию модели, в котором бонус равен нулю, и он не приведет к вытеснению мотивации. Однако если предлагаемый бонус значителен, он мо-

жет привести к вытеснению мотивации тех участников, которые в процессе программы обнаружат, что их издержки от усилий слишком высоки (т. е. они принадлежат к слабому типу).

#### 4. Заключение

Представленная в работе теоретическая модель позволяет объяснить противоречивые результаты, которые наблюдаются в эмпирической литературе относительно долгосрочных эффектов внешнего вознаграждения на мотивацию к самоконтролю. Механизм, который описан в модели, не основывается на предположении о том, что назначаемое принципалом вознаграждение является сигналом о сложности задачи или о способностях агента (как в работе [4]). Модель описывает выбор между более низкими краткосрочными результатами и сохранением относительно высокого уровня уверенности и мотивации в будущем, и более высокими краткосрочными результатами, которые могут обернуться демотивированностью в будущем.

Данная модель показывает, что эффект вытеснения мотивации может возникнуть даже в отсутствие асимметрии информации между принципалом и агентом. Основной эффект в модели возникает из-за того, что приложение усилий не только увеличивает шансы на успех в текущем периоде, но одновременно выявляет информацию о ненаблюдаемом уровне издержек, которая может повлиять на мотивацию в будущем. Также модель показывает, что поведение людей с достаточно высоким уровнем уверенности слабо зависит от присутствия внешних стимулов. Эффект вытеснения мотивации с большей вероятностью возникнет именно для индивидов с относительно низким уровнем уверенности в своих способностях к самоконтролю.

Модель также помогает объяснить, почему эффект вытеснения мотивации в большей степени проявляется в ситуациях, когда вознаграждение предлагается за достижения, чем когда оно предлагается за сам факт участия в программе [6, 8, 19]. Исходя из результатов модели, это наблюдение можно объяснить следующим образом. Вознаграждение за участие не зависит от успехов индивида и усилий, которые он прилагает в процессе участия. Таким образом, агент имеет возможность подстраивать уровень усилий так, чтобы то количество информации о своем типе, которое он получит в процессе, не снижало его мотивацию в будущем. Если же воз-

награждение предлагается за достижения, эта возможность у агента отсутствует, так как теперь получение вознаграждения напрямую от них зависит.

## 5. Приложение

### 5.0.1. Доказательство Утверждения 1.

Выпишем функцию полезности внутреннего принципала согласно формуле (8). Бонус, который выплачивается агенту в первом периоде, равен  $\hat{k}$ . Во втором периоде бонус равен нулю независимо от его издержек. Ожидаемый исход составляет  $p \times 0 + (1 - p) \times 1 = (1 - p)$  и, следовательно, из (5) получаем:

$$\begin{aligned} EU_1^{SP}(e_1 = 1) &= (1 - \hat{k}) + (1 - p) \times \\ &\times (1 + 1 - (1 - p)) + \lambda p(0 - (1 - p)) = \\ &= (1 - \hat{k}) + (-\lambda p(1 - p) + (1 - p)(1 + p)). \end{aligned}$$

В начале периода 1 внутренний принципал сравнивает полезности  $EU_1^{SP}(e_1 = 1)$  и  $EU_1^{SP}(e_1 = 0) = 1 - \hat{k}$ :

$$\begin{aligned} EU_1^{SP}(e_1 = 1) - EU_1^{SP}(e_1 = 0) &= \\ &= \begin{cases} (-\lambda p(1 - p) + (1 - p)(1 + p)), & \text{if } p < \bar{p}, \\ (1 - \hat{k}) + (-\lambda p(1 - p) + (1 - p)(1 + p)), & \text{if } p < \bar{p}. \end{cases} \end{aligned} \quad (9)$$

Следовательно, полезность внутреннего принципала выше при  $e_1 = 1$ , если  $1 + p > \lambda p$ ,  $p < \frac{1}{\lambda - 1}$ .

Возможны следующие два случая:

$$1) \frac{1}{\lambda - 1} \geq \bar{p} = \frac{1}{k_H} \text{ или } \lambda \leq 1 + k_H. \text{ В этом случае}$$

пороговое значение  $\frac{1}{\lambda - 1}$  будет выше значения  $\bar{p}$ , после которого внутренний принципал не платит бонус во втором периоде. Это означает, что для всех  $p < \bar{p}$ ,  $EU_1^{SP}(e_1 = 1) - EU_1^{SP}(e_1 = 0) > 0$ .

Рассмотрим интервал  $p \in [\bar{p}, \frac{1}{\lambda - 1}]$ .

Функция (9) принимает следующий вид:

$$\begin{aligned} EU_1^{SP}(e_1 = 1) - EU_1^{SP}(e_1 = 0, p \geq \bar{p}) &= \\ &= p^2(\lambda - 1) - p(k_H + \lambda) + 2. \end{aligned} \quad (10)$$

Заметим, что минимум этой функции достигается при  $p = \frac{k_H + \lambda}{2(\lambda - 1)} \geq 1$ , так как  $2 + k_H > 1 + k_H \geq \lambda$ . Более того, функция положительна при  $p = \bar{p}$  (функция принимает значения больше  $1 - \hat{k}$  в интервале  $[0, \bar{p}]$ ) и отрицательна при  $p = 1$ , так как  $1 - k_H < 0$ . Следовательно, функция положительна в интервале  $p \in [\bar{p}, \frac{1}{\lambda - 1}]$ . Существует единственное значение аргумента, при котором функция равна нулю в интервале  $[\bar{p}, 1]$ .

Решив уравнение  $p^2(\lambda - 1) - p(k_H + \lambda) + 2 = 0$  получаем 
$$p^* = \frac{k_H + \lambda - \sqrt{(k_H + \lambda)^2 - 8(\lambda - 1)}}{4(\lambda - 1)}.$$

Заметим, что при  $\lambda \leq 1 + k_H$ :

$$\begin{aligned} (k_H + \lambda)^2 - 8(\lambda - 1) &\geq (2\lambda - 1)^2 - 8(\lambda - 1) = \\ &= 4\lambda^2 - 4\lambda + 1 - 8\lambda + 8 = \\ &= 4\lambda^2 - 12\lambda + 9 = (2\lambda - 3)^2 \geq 0. \end{aligned}$$

Таким образом корень уравнения существует. Значение второго корня

$$\frac{k_H + \lambda + \sqrt{(k_H + \lambda)^2 - 8(\lambda - 1)}}{4(\lambda - 1)}$$

больше первого

корня и больше 1. При  $p \geq p^*$  внутренний принципал платит нулевой бонус в первом периоде.

2)  $\frac{1}{\lambda - 1} < \bar{p} = \frac{1}{k_H}$  или  $\lambda > 1 + k_H$ . В этом случае внутренний принципал в первом периоде платит бонус  $\hat{k}$  при  $p < \frac{1}{\lambda - 1} < \bar{p}$ , поскольку функция (9) положительна в этом интервале. Если  $p \geq \frac{1}{\lambda - 1}$ , функция (9) становится отрицательной, внутренний принципал предлагает нулевой бонус. Следовательно,  $p^* = \frac{1}{\lambda - 1}$ . Q.E.D.

**5.0.2. Доказательство Утверждения 2**

i) Из доказательства Утверждения 1 мы знаем, что в первом периоде внутренний принципал платит бонус  $\hat{k}$  при  $p \leq p^*$  и платит нулевой бонус в противном случае. Это означает, что  $e_1 = 1$  при  $p \leq p^*$ , и во втором периоде только агенты с низкими издержками будут прилагать положительные усилия. Следовательно, ожидаемый уровень усилий второго периода равен  $\bar{e}_2 = 1 - p$ .

Если  $p > p^*$ , то  $e_1 = 0$  и реальный уровень издержек не будет наблюдаться. Более того, так как  $p^* > \bar{p}$ , внутренний принципал будет платить нулевой бонус во втором периоде. Следовательно,  $\bar{e}_2 = 0$ .

ii) Из доказательства Утверждения 1 мы знаем, что в первом периоде внутренний принципал платит бонус  $\hat{k}$  при  $p \leq p^*$ . Следовательно  $e_1 = 1$ , реальный уровень издержек раскрывается и во втором периоде только агенты с низкими издержками прикладывают усилия. Поэтому ожидаемый уровень усилий составит  $\bar{e}_2 = 1 - p$ .

Если  $p \in (p^*, 1/k_H]$ , тогда внутренний принципал платит ноль в первом периоде и платит  $\hat{k}$  во втором периоде. Это означает, что агент первого периода не прикладывает усилий, а во втором периоде усилия равны 1, то есть  $e_1 = 0$  и  $\bar{e}_2 = 1$ .

Если  $p > 1/k_H$ , внутренний принципал платит нулевой бонус в обоих периодах, поэтому  $e_1 = 0, \bar{e}_2 = 0$ .

**5.0.3. Доказательство Утверждения 3**

Чтобы доказать Утверждение 3, необходимо сначала описать оптимальное поведение внешнего принципала для различных значений параметров. Как показано в доказательстве Утверждения 1, существуют два основных случая:  $\lambda > 1 + k_H$  и  $1 < \lambda \leq 1 + k_H$ .

Случай 1:  $1 < \lambda \leq 1 + k_H$ .

Как показано в доказательстве Утверждения 1, внутренний принципал предлагает положительный бонус в первом периоде,  $e_1 = 1$  при  $p \leq p^*$ , где  $p^* > 1/k_H$ . Это означает, что внешний принципал предлагает нулевой бонус. Если  $p \in (p^*, 1]$ , внутренний принципал не предлагает бонуса в первом периоде, так как это приводит к отрицательной полезности для него. Рассмотрим функцию полезности внутреннего принципала в этом интервале с учетом бонуса внешнего принципала  $b^{EP}$ :

$$\begin{aligned} EU_1^{SP}(e_1 = 1) &= \gamma b^{EP} + (1 - \hat{k}) + \\ &+ (-\lambda p(1 - p) + (1 - p)(1 + p)). \end{aligned}$$

Таким образом, бонус внешнего принципала уменьшает бонус внутреннего принципала на величину  $\gamma b^{EP}$ . Внешний принципал предлагает бонус, который позволяет внутреннему принципалу получать нулевую полезность. Это означает, что:

$$\gamma b^{EP} = -1 + \hat{k} + \lambda p(1 - p) - (1 - p)(1 + p). \quad (11)$$

Заметим, что полезность внешнего принципала в случае  $e_1 = 0$  равна нулю, в то время как при положительных усилиях реальный уровень издержек выявляется и с вероятностью  $1 - p$  агент второго периода также приложит положительные усилия. Следовательно,  $b^{EP} \leq 1 + 1 - p = 2 - p$ , и

$$\gamma(2 - p) \geq -1 + \hat{k} + \lambda p(1 - p) - (1 - p)(1 + p). \quad (12)$$

Обозначим

$$\begin{aligned} f(p) &= -1 + \hat{k} + \lambda p(1 - p) - (1 - p)(1 + p) = \\ &= -p^2(\lambda - p) + p(\lambda + k_H) - 2. \end{aligned}$$

Это квадратичная функция от  $p$ , и она пересекает ноль при  $p = p^*$ . Вычислим значение аргумента, при котором эта функция принимает максимальное значение:

$$f'(p) = -2(\lambda - 1)p + (\lambda + k_H) = 0, p = \frac{\lambda + k_H}{2(\lambda - 1)}.$$

Заметим, что поскольку  $\lambda \leq 1 + k_H$ , верно следующее:

$$p_1 = \frac{\lambda + k_H}{2(\lambda - 1)} \geq \frac{\lambda + \lambda - 1}{2\lambda - 2} > 1.$$

Это означает, что максимум функции достигается за пределами интервала  $[p^*, 1]$ , поэтому максимальное значение внутри интервала достигается при  $p = 1$ , так как  $f(p^*) = 0$  и  $p_1 > 1$ . Вычислим  $f_1(1) = -(\lambda - 1) + \lambda + k_H - 2 = k_H - 1$ . Эта функция возрастает в интервале  $[p^*, 1]$ .

Функция  $g(p) = \gamma(2 - p)$  убывает в интервале  $[p^*, 1]$ . Ее значение при  $p = p^*$  положительно ( $g(p^*) = \gamma(2 - p^*)$ ), и ее значение при  $p = 1$  равно  $g(1) = \gamma$ .

В случае  $\gamma \geq k_H - 1$  единственной точкой пересечения функций  $f(p)$  и  $g(p)$  в интервале  $[p^*, 1]$  является  $p = 1$ . Это означает, что внешний принципал всегда может предложить бонус, который будет стимулировать агента в первом периоде прикладывать усилия  $e_1 = 1$ .

В случае  $\gamma < k_H - 1$  функции  $f(p)$  и  $g(p)$  пересекаются в точке, которая является решением уравнения:  $p^2(\lambda - 1) + p(\lambda + k_H) - 2 = \gamma(2 - p)$ ,

$$p^{**} = \frac{\lambda + k_H + \gamma - \sqrt{(\lambda + k_H + \gamma)^2 - 4(2 + 2\gamma)(\lambda - 1)}}{2(\lambda - 1)}.$$

(Второй корень уравнения принимает значение больше 1, так как  $\lambda + k_H > 2(\lambda - 1)$ ). Внешний принципал компенсирует бонус внутреннего принципала до точки  $p^{**}$ , и оба принципала платят нулевой бонус, если  $p \in [p^*, 1]$ , поэтому в интервале  $p \in [p^*, 1]$  решение выглядит следующим образом:

i. Если  $\gamma \geq k_H - 1$  внешний принципал предлагает бонус:

$$b_{EP} = \frac{-p^2(\lambda - 1) + p(\lambda + k_H) - 2}{\gamma}, \quad (13)$$

агент выбирает усилия  $e_1 = 1$  в первом периоде, внутренний принципал предлагает бонус  $k - b^{EP}$ .

ii. Если  $\gamma < k_H - 1$  и  $p \in [p^*, p^{**}]$ , тогда внешний принципал предлагает бонус (13), агент выбирает усилия  $e_1 = 1$ , внутренний принципал предлагает бонус  $k - b^{EP}$ . Если  $\gamma < k_H - 1$  и  $p \in (p^{**}, 1]$ , тогда оба принципала платят нулевой бонус и агент выбирает нулевые усилия в первом периоде.

Случай 2:  $\lambda > 1 + k_H$ .

Как показывает доказательство Утверждения 1,  $p^* < 1/k_H$  и внутренний принципал предлагает бонус при  $p \leq p^*$ . Тогда внешний принципал предлагает нулевой бонус в этом интервале. Если  $p > p^*$ , внутренний принципал предлагает нулевой бонус.

Рассмотрим интервал  $p \in [1/k_H, 1]$ . В этом интервале разница между полезностью внутрен-

него принципала при положительных и нулевых усилиях в первом периоде описывается формулой (9). Обозначим:

$$f_1(p) = -(-\lambda p(1 - p) + (1 - p)(1 + p)) = -p^2(\lambda - 1) + \lambda p - 1.$$

Внешний принципал получает  $2 - p$  при  $e_1 = 1$ , поскольку только с вероятностью  $1 - p$  агент прикладывает усилия во втором периоде. Если  $e_1 = 0$ , внутренний принципал мотивирует агента второго периода и внешний принципал получает 1. Таким образом, максимальный бонус, который может быть предложен внешним принципалом, составляет  $b^{EP} = 1 - p$ .

В интервале  $p \in [p^*, 1/k_H]$  агент выберет  $e_1 = 1$  только в случае, когда:

$$\gamma b^{EP} = f_1(p), \gamma(1 - p) \geq f_1(p). \quad (14)$$

Это означает, что бонус, который предложит внешний принципал, составит:

$$b^{EP} = \frac{-p^2(\lambda - 1) + \lambda p - 1}{\gamma}. \quad (15)$$

Рассмотрим теперь интервал  $p \in [1/k_H, 1]$ . В этом интервале разница между полезностями внутреннего принципала при усилиях 1 и 0 описывается выражением (10). Обозначим:

$$f_2(p) = -(p^2(\lambda - 1) - p(k_H + \lambda) + 2) = -p^2(\lambda - 1) + p(\lambda + k_H) - 2.$$

Внешний принципал получает  $2 - p$ , если  $e_1 = 1$ , поскольку только с вероятностью  $1 - p$  агент приложит усилия во втором периоде. Если  $e_1 = 0$ , внутренний принципал не может мотивировать агента работать во втором периоде, поэтому внешний получает 0. Это означает, что максимальный бонус составляет  $b^{EP} = 2 - p$ . В интервале  $p \in [1/k_H, 1]$  агент приложит усилия  $e_1 = 1$ , только если:

$$\gamma b^{EP} = f_2(p), \gamma(2 - p) \geq f_2(p). \quad (16)$$

Обе функции,  $f_1(p)$  и  $f_2(p)$ , являются квадратичными. Сравним  $f_1$  с  $h_1(p) = \gamma(1 - p)$  и  $f_2$  с  $h_2(p) = \gamma(2 - p)$ . Обе функции  $h_1$  и  $h_2$  являются линейными и убывающими. Рассмотрим сначала  $f_1, h_1$ .

Заметим, что  $f_1(1) = h_1(1) = 0$ , поэтому единственной касательной линией к  $f_1$  вида  $\gamma(1 - p)$  является та, которая имеет единственную точку пересечения в  $p = 1$ . Любая другая функция  $h_1$  пересекает функцию  $f_1$  дважды в интервале  $[p^*, 1]$  — в точке  $p = 1$  и в точке  $p_{L1}^\gamma$ :

$$-p^2(\lambda - 1) + \lambda p - 1 = \gamma(1 - p);$$

$$-p^2(\lambda - 1) + (\lambda + \gamma)p - 1 - \gamma = 0;$$

$$(p - 1)(-p(\lambda - 1) + 1 + \gamma) = 0;$$

$$p_{L1}^\gamma = \frac{1 + \gamma}{\lambda - 1}.$$

Эта точка больше, чем точка  $p^* = \frac{1}{\lambda - 1}$  и должна быть меньше, чем  $1/k_H$  или:

$$\frac{1 + \gamma}{\lambda - 1} \leq \frac{1}{k_H}, \gamma \leq \frac{\lambda - 1}{k_H} - 1.$$

Поэтому при малых  $\gamma$ ,  $f_1$  и  $h_1$  пересекаются внутри интервала  $[p^*, 1/k_H]$ . Функции пересекаются в точке выше, чем  $1/k_H$ , если  $\gamma$  достаточно велико.

Верен следующий вывод: если  $0 < \gamma \leq \frac{\lambda - 1}{k_H} - 1$ ,

тогда внешний принципал предлагает бонус (15) в интервале  $[p^*, p_{L1}^\gamma]$ , агент выбирает усилия  $e_1 = 1$  и  $Ee_2 = 1 - p$ ; в интервале  $(p^*, p_{L1}^\gamma]$  внешний принципал предлагает нулевой бонус, агент выбирает  $e_1 = 0$  и  $e_2 = 1$ .

Если  $0 < \gamma > \frac{\lambda - 1}{k_H} - 1$ , тогда внешний принципал платит бонус в интервале  $[p^*, 1/k_H]$ , агент выбирает  $e_1 = 1$  и  $Ee_2 = 1 - p$ .

Функции  $f_2$ ,  $h_2$  представляются более сложными для анализа, поэтому сосредоточимся только на тех случаях, которые предоставляют условия, описанные в доказываемом утверждении. Заметим, что функция  $f_2$  достигает максимума в точке  $p_2 = \frac{\lambda + k_H}{2(\lambda - 1)}$ . Ее значение в этой точке составляет:

$$f_2^{\max} = \frac{(\lambda + k_H)^2}{2(\lambda - 1)} - \frac{(\lambda + k_H)^2}{4(\lambda - 1)} - 2 = \frac{(\lambda + k_H)^2}{4(\lambda - 1)} - 2.$$

В одном из случаев прямая  $h_2$  лежит выше  $f_2$ . В этой ситуации внешний принципал предлагает бонус (13) в первом периоде для любых  $p \in [1/k_H, 1]$ . Достаточное условие:  $h_2(1) = \gamma > f_2^{\max}$ . Тогда  $e_1 = 1$  и только агент с низкими издержками выберет высокий уровень усилий во втором периоде,  $Ee_2 = 1 - p$ .

Возможен также другой случай, когда  $h_2(1/k_H) = \gamma(2 - 1/k_H) > f_2(1/k_H)$ ,  $h_2(1/k_H) = \gamma < f_2(1) = k_H - 1$ . В этой ситуации внешний принципал предлагает положительный бонус при  $p \in [1/k_H, p_{R1}^\gamma]$  и предлагает ноль при  $p \in (p_{R1}^\gamma, 1]$ . Здесь  $p_{R1}^\gamma$  определяется как корень уравнения:

$$-p^2(\lambda - 1) + p(\lambda + k_H) - 2 = \gamma(2 - p);$$

$$-p^2(\lambda - 1) + p(\lambda + k_H + \gamma) - 2 - 2\gamma = 0;$$

$$\frac{\lambda + k_H + \gamma - \sqrt{(\lambda + k_H + \gamma)^2 - 4(\lambda - 1)(2 + 2\gamma)}}{2(\lambda - 1)}$$

В интервале  $p \in [1/k_H, p_{R1}^\gamma]$  получаем  $e_1 = 1$ ,  $Ee_2 = 1 - p$ , и для  $p \in (p_{R1}^\gamma, 1]$  получаем  $e_1 = 0$ ,  $e_2 = 0$ .

Еще один важный случай возникает, когда бонус равен нулю в интервале  $[1/k_H, 1]$ . Это происходит, если  $h_2(1/k_H) = \gamma(2 - 1/k_H) < f_2(1/k_H)$ ,  $h_2(1) = \gamma < f_2(1) = k_H - 1$ , например, при достаточно малых  $\gamma$ . Внешний принципал предлагает нулевой бонус, поэтому усилия агента  $e_1 = e_2 = 0$ .

Наконец, обозначим

$$C = \min \left( \frac{k_H}{2k_H - 1} \left( \frac{\lambda}{k_H} - 1 - \frac{\lambda - 1}{k_H^2} \right), k_H - 1 \right).$$

Тогда при  $\gamma \leq C$  пересечение между  $h_2$  и  $f_2$  отсутствует. Оба принципала предлагают нулевой бонус, и агент выбирает нулевые усилия в обоих периодах. Если  $\gamma \geq C$ , тогда либо  $h_2(1/k_H) > f_2(1/k_H)$ ,  $h_2(1) > f_2(1)$ , либо выполнены оба условия. Тогда либо  $h_2 > f_2$  на всем интервале, либо эти функции пересекаются внутри интервала. Следовательно, существует хотя бы один интервал внутри  $[1/k_H, 1]$ , где внешний принципал предлагает положительный бонус и агент выбирает  $e_1 = 1$  и  $Ee_2 = 1 - p$ .

Доказательство Утверждения 3 завершено.

### Список источников

1. Ali N. Learning Self-Control // Quarterly Journal of Economics. — 2011. — Vol. 126. — No. 2. — P. 857–893.
2. Angeletos G.-M., Laibson D., A. Repetto A., Tobacman J., Weinberg S. The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation // Journal of Economic Perspectives. — 2001. — Vol. 15. — No. 3. — P. 47–68.
3. Bains N., Pickett W., Hoey J. The use and impact of incentives in population-based smoking cessation programs : a review // American Journal of Health Promotion. — 1998. — Vol. 12. — No. 5. — P. 307–320.
4. Bénabou R., Tirole J. Intrinsic and Extrinsic Motivation // Review of Economic Studies. — 2003. — Vol. 70. — P. 489–520.
5. Benartzi Sh., Richard H. Thaler Myopic Loss Aversion and the Equity Premium Puzzle // The Quarterly Journal of Economics. — 1995. — Vol. 110 (1). — P. 73–92.
6. Charness G., Gneezy U. Incentives to Exercise. Working paper. 2008.
7. Curry S.J., Wagner E.H., Grothaus L.C. Intrinsic and Extrinsic Motivation for Smoking Cessation // Journal of Consulting and Clinical Psychology. — 1990. — Vol. 58. — P. 310–316.
8. Deci E., Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior // Perspectives in Social Psychology. Springer. 1985.
9. Dodd M. Obesity and Time-Inconsistent Preferences // Obesity Research and Clinical Practice. — 2008. — Vol. 2, Issue 2. — P. 83–89.

10. *Finkelstein E., Linnan L., Tate D., Birken B.* A Pilot Study Testing the Effect of Different Levels of Financial Incentives on Weight Loss Among Overweight Employees // *Journal of Occupational and Environmental Medicine*. — 2007. — Vol. 49, Issue 9. — P. 981–989.
11. *Fudenberg D., Levine D.* A Dual Self Model of Impulse Control // *American Economic Review*. — 2006. — No. 96. — P. 1449–1476.
12. *Frederick S., Loewenstein G., O'Donoghue T.* Time Discounting and Time Preference: A Critical Review // *Journal of Economic Literature*. — 2002. — Vol. XL. — P. 351–401.
13. *Genesove D., Mayer Ch.* Loss Aversion and Seller Behavior: Evidence from the Housing Market // *The Quarterly Journal of Economics*. — 2001. — Vol. 116 (4). — P. 1233–1260.
14. *Gill D., Prowse V.* A Structural Analysis of Disappointment Aversion in a Real Effort Competition // *American Economic Review*. — 2012. — Vol. 102. — No. 1. — P. 469–503.
15. *Gruber J., Köszegi B.* Is Addiction “Rational”? Theory and Evidence // *Quarterly Journal of Economics*. — 2001. — Vol. 116(4). — P. 1261–1303.
16. *Harland J., White M., Drinkwater C., Chinn D., Farr L., Howel D.* The Newcastle Exercise Project: A Randomized Controlled Trial of the Methods to Promote Physical Activity in Primary Care // *British Medical Journal*. — 1999. — Vol. 319. — P. 828–832.
17. *James H. S. Jr.* Why Did You Do That? An Economic Examination of the Effect of Extrinsic Compensation on Intrinsic Motivation and Performance // *Journal of Economic Psychology*. — 2005. — Vol. 26. — P. 549–566.
18. *Hey K., Perera R.* Quit and Win contests for smoking cessation (Review) // *Cochrane Database of Systematic Reviews*. — 2005. — Issue 2, article CD004986.
19. *Jochelson K.* Paying the Patient. Improving Health Using Financial Incentives // *King's Fund paper series*. 2007.
20. *Kahneman D., Tversky A.* Prospect Theory: An Analysis of Decisions under Risk // *Econometrica*. — 1979. — Vol. 47 (2). — P. 263–291.
21. *Kan K.* Cigarette Smoking and Self-Control // *Journal of Health Economics*. — 2007. — Vol. 26, Issue 1. — P. 61–81.
22. *Koch A., Nafziger J., Suvorov A., van de Ven J.* Self-rewards and Personal Motivation // *European Economic Review*. — 2014. — Vol. 68(C). — P.151–167.
23. *Laibson D.* Golden Eggs and Hyperbolic Discounting // *Quarterly Journal of Economics*. — 1997. — Vol. 112, Issue 2. — P. 443–478.
24. *Leuven E., Oosterbeek H., van der Klaauw B.* The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment”. Working paper. — 2006.
25. *O'Donoghue T., Rabin M.* Doing It Now or Later // *The American Economic Review*. — 1999. — Vol. 89. — No. 1. — P. 103–124.
26. *Ryan R., Plant R., O'Malley S.* Initial Motivations for Alcohol Treatment: Relations with Patient Characteristics, Treatment Involvement, and Dropout // *Addictive Behaviors*. — 1995. — Vol. 20. — No. 3. — P. 279–297.
27. *Volpp K., John L., Troxel A., Norton L., Fassbender J., Loewenstein G.* Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial // *JAMA*. — 2008. — Vol. 300(22). — P. 2631–2637.
28. *Volpp K., Troxel A., Pauly M., Glick H., Puig A., Asch D., Galvin R., Zhu J., Wan F., DeGuzman J., Corbett E., Weiner J., Audrain-McGovern J.* A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation // *The New England Journal of Medicine*. — 2009. — Vol. 360 (7).
29. *Williams G., Grow V. M., Freedman Z. R., Ryan R. M., Deci E. L.* Motivational Predictors of Weight Loss Weight-Loss Maintenance // *Journal of Personality and Social Psychology*. — 1996. — Vol. 70. — No. 1. — P. 115–126.
30. *Williams G., Gagné M., Ryan R., Deci E.* Facilitating Autonomous Motivation for Smoking Cessation // *Health Psychology*. — 2002. — Vol. 21. — No. 1. — P. 40–50.

### Информация об авторе

**Паниди Ксения Андреевна** — Ph.D. in Economics, доцент факультета экономических наук, Национальный исследовательский университет — Высшая школа экономики (Москва, Российская Федерация; e-mail: kpanidi@hse.ru)

**Panidi K. A.**

### Economic Model of Motivational Crowding Out and the Self-Control Problem

*Recent economic research has drawn a lot of attention to the problem of dynamically inconsistent preferences, which describes many situations connected with the phenomenon of self-control and willpower. One of the ways to solve the self-control problem is offering a reward in order to stimulate more rational behavior. However, many psychological studies indicate that offering such a reward may undermine long-term self-control motivation in the future, despite a seeming short-term improvement. Motivational crowding-out is mostly observed for external (when a reward is provided by an individual or an organization) in rather than internal rewards (when a person chooses a reward for himself). The author develops an economic model which helps to explain differences in the short-term and long-term effects of internal and external rewards on the self-control motivation. The model is based on the dual-self approach to the dynamic inconsistency and assumes lack of information on the costs of self-control. The model shows that individuals who have a low level of self-control are more prone to motivational crowding out than*

*individuals who have a high level of self-control. The second ones are mostly unaffected by external rewards. These conclusions coincide with the results of empirical studies on self-control and external rewards.*

**Keywords:** self-control problem, time-inconsistent preferences, motivational crowding-out