

УДК 164.08
DOI: 10.17223/1998863X/47/5

Г.В. Карпов

НЕПРЯМЫЕ ДЕЙСТВИЯ АГЕНТОВ И АВТОРИТЕТНЫЕ ОТНОШЕНИЯ В STIT И PDL

Рассматривается понятие авторитетных отношений между агентами с точки зрения логики действий и с точки зрения пропозициональной динамической логики (PDL). В первом случае наличие авторитетного отношения между агентами означает возможность отдать указание на исполнение действия, результатом которого является данное положение дел. Такое указание моделируется средствами NEXT-семантики, когда влияние одного агента на другого дается как композиция отношений достижимости между мирами. Во втором случае наличие авторитетного отношения между агентами обусловлено возможностью самостоятельного исполнения действия агентом, занимающим авторитетную позицию. Авторитетное отношение здесь предстает как транзитивный переход на процессуальных графах к третьему миру из первого, где действует автор указания, через второй, где действует исполнитель указания. Статья предлагает ряд семантических идей, направленных на объединение сильных сторон логики действий и PDL.

Ключевые слова: логика действий, STIT, NEXT-семантика, PDL, авторитетные отношения.

1. Авторитетные отношения в логике действий на базе NEXT-семантики

1.1. Язык, модельная структура, модель

Для того чтобы иметь возможность говорить об интересующем нас предмете, обозначенном в заглавии статьи, зададим язык L_{aut} логики действий¹. Пусть Φ есть множество пропозициональных переменных, а Γ – множество агентов. Правильно построенная формула φ языка L_{aut} определяется следующим образом:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid X\varphi \mid \Box\varphi \mid [\alpha]\varphi \mid \circ\varphi,$$

где $p \in \Phi$ и $\alpha \in \Gamma$.

Модельная структура F для L_{aut} есть упорядоченная шестерка элементов $\langle M, H, R_X, R_\square, R_\circ, R_A \rangle$. Множество $M = \{m_1, m_2, \dots\}$ есть бесконечное множество моментов времени. Множество $H = \{h_1, h_2, \dots\}$ есть бесконечное множество историй, такое, что каждая история h представляет собой множество упорядоченных элементов из M . Множество пар $\{\langle m/h \rangle, \langle m'/h' \rangle\} \in M \times H$ обозначает множество миров, относительно которых осуществляется означивание формул.

Отношение R_X есть отношение на множестве миров $M \times H$ такое, что если $\langle m/h \rangle R_X \langle m'/h' \rangle$, то $h = h'$. Другими словами, для каждого момента времени

¹ Читатели, знакомые с семантическими основаниями логики действий, могут пропустить эти четыре абзаца.

M отношение R_X определяет последующий, более поздний, момент m' , принадлежащий той же истории, что и m . R_X антирефлексивно, транзитивно и антисимметрично.

Отношение R_{\square} есть отношение на $M \times H$ такое, что если $\langle m/h \rangle R_{\square} \langle m'/h' \rangle$, то $M = m'$. Так отношение R_{\square} для каждой истории h указывает такую историю h' , которая проходит через тот же момент времени, что и история h . R_{\square} рефлексивно, транзитивно и симметрично.

Отношение R_{\circ} есть отношение на $M \times H$ такое, что если $\langle m/h \rangle R_{\circ} \langle m'/h' \rangle$, то $m = m'$ и каждая h' есть «идеальная» или «деонтически совершенная» история, т.е. такая история, которая наиболее предпочтительна для некоторого агента. R_{\circ} серийно (serial), т.е. для каждого мира $\langle m/h \rangle$ оно находит хотя бы один мир $\langle m'/h' \rangle$, такой, где h' – «идеальная» история.

Отношение R_A есть отношение на $M \times H$ такое, что $R_A \subseteq R_X \circ R_{\square}$ и $R_{\circ} \subseteq R_{\square} \circ R_X$, где A – метапеременная для $\{\alpha, \beta, \dots\} \in \Gamma$. Мы также будем называть R_A отношением эффективности для агентов A .

Модель M образуется присоединением к F функции означивания v , которая ставит в соответствие атомарным пропозициям языка L_{aut} множество миров из $M \times H$, где эти пропозиции истинны. Условия истинности формул языка L_{aut} определяются следующими правилами:

$M, \langle m/h \rangle \models p$ е. т.е. $\langle m/h \rangle \in v(p)$;

$M, \langle m/h \rangle \models \neg \phi$ е. т.е. $M, \langle m/h \rangle \not\models \phi$;

$M, \langle m/h \rangle \models \phi \wedge \psi$ е. т.е. $M, \langle m/h \rangle \models \phi$ и $M, \langle m/h \rangle \models \psi$;

$M, \langle m/h \rangle \models X\phi$ е. т.е. существует такой мир $\langle m'/h' \rangle$, что $\langle m'/h' \rangle R_X \langle m/h \rangle$ и $M, \langle m'/h' \rangle \models \phi$;

$M, \langle m/h \rangle \models \square \phi$ е. т.е. для каждого мира $\langle m'/h' \rangle$, такого, что $\langle m'/h' \rangle R_{\square} \langle m/h \rangle$, имеет место $M, \langle m'/h' \rangle \models \phi$;

$M, \langle m/h \rangle \models [A]\phi$ е. т.е. для всех миров $\langle m'/h' \rangle$, таких, что $\langle m'/h' \rangle R_A \langle m/h \rangle$, имеет место $M, \langle m'/h' \rangle \models \phi$;

$M, \langle m/h \rangle \models \circ \phi$ е. т.е. $M, \langle m'/h' \rangle \models \phi$ для каждого мира $\langle m'/h' \rangle$, такого, что $\langle m'/h' \rangle R_{\circ} \langle m/h \rangle$.

1.2. Авторитетное отношение как возможность указывать

Будем считать, что α связан авторитетным отношением с β по ϕ , если и только если α может сделать так, что у β в будущем возникнет соответствующее обязательство. Подобное действие α мы иногда будем называть непрямым действием, так как оно предполагает участие другого агента. На все прочие действия, такого участия не предполагающие, мы будем ссылаться как на прямые. Наша непосредственная задача сводится, таким образом, к моделированию непрямого действия и тех отношений между агентами, которые существуют в связи с возможностью такого действия. Пусть формула $\langle \alpha \rangle \phi$, истинная в некотором мире $\langle m/h \rangle$, в M указывает на то, что существует хотя бы один мир $\langle m'/h' \rangle$, такой, что $\langle m'/h' \rangle R_{\alpha} \langle m/h \rangle$ и $M, \langle m'/h' \rangle \models \phi$. Тогда возможность для α отдать β указание на выполнение ϕ и, следовательно, авторитетное отношение α к β по ϕ можно выразить формулой $\langle \alpha \rangle \circ [\beta] \phi$.

Условия истинности для $\langle \alpha \rangle \circ [\beta] \phi$ определены на модели, представленной ниже (рис. 1), где отражена ситуация наличия авторитетного отношения между агентами α и β по ϕ , когда в некотором начальном моменте α отдает

указание β на выполнение ϕ , а в последующем моменте β , связанный теперь обязательством, принимает решение о том, исполнять это указание или нет.

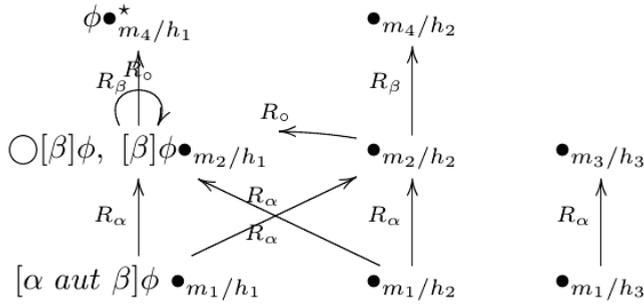


Рис. 1. Модель авторитетного отношения α к β по ϕ

Как видим, наличие авторитетного отношения по ϕ между α и β в мире $\langle m_1/h_1 \rangle$ порождает обязательство β (перед α) в одном из миров, связанных отношением R_α с данным миром. Действительное побуждение агентом α агента β к выполнению ϕ дает обязательство для β сделать так, что ϕ во всех мирах, связанных с данным миром отношением R_α . И в том и в другом случае у β сохраняется возможность уклониться от выполнения ϕ , выбрав мир $\langle m_2/h_2 \rangle$, где допустимо $\neg\phi$. Последнее не снимает $\circ[\beta]\phi$ в $\langle m_2/h_1 \rangle$, так как идеальной историей в данной модели назначена история h_1 (она отмечена звездочкой).

На основании изложенного нетрудно дополнить список определений истинности формул языка L_{aut} следующим производным определением:

$M, \langle m/h \rangle \models \langle \alpha \rangle \circ [\beta]\phi$ е. т.е $M, \langle m'/h' \rangle \models \circ[\beta]\phi$ для некоторого мира $\langle m'/h' \rangle$, такого, что $\langle m/h \rangle R_\alpha \langle m'/h' \rangle$. Или: $M, \langle m/h \rangle \models \langle \alpha \rangle \circ [\beta]\phi$ е. т.е $M, \langle m'/h' \rangle \models [\beta]\phi$, для некоторого мира $\langle m'/h' \rangle$, такого, что $\langle m/h \rangle R_\alpha \circ R_\beta \langle m'/h' \rangle$ (где история h' – «идеальная» история).

Композиция $R_\alpha \circ R_\beta$ отношений определяет миры для α , в которых действует его авторитет по ϕ в связи с β . Смоделируем теперь несколько случаев использования агентами авторитетных отношений.

Случай 1: вложенные авторитетные отношения. Пусть α находится в авторитетном отношении по ϕ к β , когда тот сам связан авторитетным отношением по ϕ к γ ; можно ли утверждать в этом случае, что в авторитетном отношении по ϕ к γ находится α ? И если да, то относительно какого именно момента это верно?

Формула, отражающая случай вложенных авторитетных отношений, имеет вид: $\langle \alpha \rangle \circ \langle \beta \rangle \circ [\gamma]\phi$. Модель для этой формулы дана на рис. 2.

Предположим, что α не связан авторитетным отношением по ϕ с γ ; т.е. во всех мирах, связанных с миром $\langle m_1/h_1 \rangle$ отношением R_α , имеет место $\neg \circ[\gamma]\phi$. Такая ситуация не является неприемлемой. Однако авторитетное отношение α к γ опосредовано авторитетным отношением α к β и β к γ , т.е. возможным действием β , для которого в модели должен найтись подходящий момент. Это действие и момент для него, следовательно, нужно учитывать в формуле, которая выражает отношение α к γ . Предположение, что авторитетного отношения по ϕ между α и γ нет, с учетом сказанного лучше всего выражает формула $\neg \langle \alpha \rangle X \circ [\gamma]\phi$ (неверно, что у α есть возможность сделать так, что в

некотором последующем моменте будет так, что γ должен сделать так, что ϕ . Тогда, $\neg X \circ [\gamma] \phi$ истинна во всех мирах, достижимых из мира $\langle m_1/h_1 \rangle$ по отношению R_α . Тогда, в каждом мире, связанном отношением R_X с данными мирами (т.е. в мирах $\langle m_4/h_1 \rangle$, $\langle m_4/h_2 \rangle$ и $\langle m_5/h_3 \rangle$), будет истинна $\neg[\gamma] \phi$. Это противоречит тому, что мы показали ранее: $\circ[\gamma] \phi$ истинна в $\langle m_4/h_1 \rangle$. Мы заключаем отсюда, что если $\langle \alpha \rangle \circ (\beta) \circ [\gamma] \phi$, то $\langle \alpha \rangle X \circ [\gamma] \phi$.

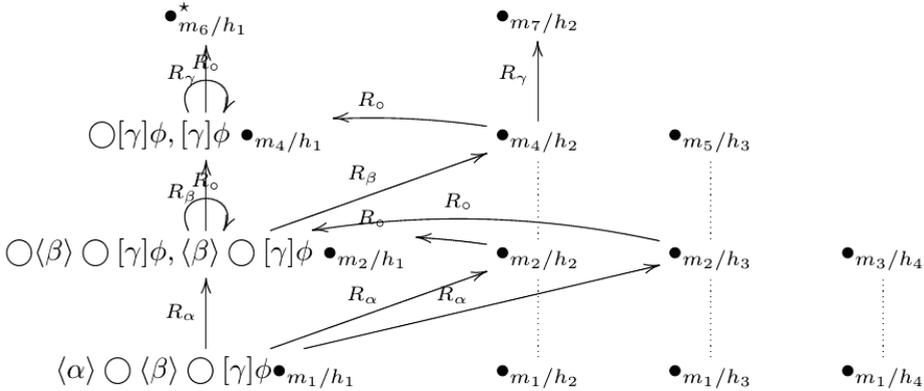


Рис. 2. Вложенные авторитетные отношения

Случай 2: подразумеваемое агентами. Пусть α обладает таким авторитетом, который позволяет ему обязать β сделать так, что при условии ϕ агент β будет делать так, что ψ . Пусть далее α способен сделать так, что у β возникает обязательство по выполнению ϕ . Означает ли это вместе, что α тем самым как бы обязывает β к выполнению ψ ? В языке L_{aut} подобная ситуация может быть воспроизведена, например, с помощью формул

$$\langle \alpha \rangle \circ ([\beta](\phi \rightarrow X[\beta]\psi)) \tag{1}$$

и уже привычной

$$\langle \alpha \rangle \circ [\beta] \phi. \tag{2}$$

Ниже мы покажем, что эти две формулы, взятые вместе, ведут к тому, что истинной оказывается также и формула

$$\langle \alpha \rangle X X \circ [\beta] \psi. \tag{3}$$

Для этого нам потребуется модельная структура¹, аналогичная той, что дана на рис. 3: здесь m_1 – момент выбора агентом α формулы (1); m_2 – момент, в котором виден результат решения, принятого α в m_1 , и одновременно момент, когда β принимает решение о выполнении или невыполнении назначенного ему действия. Неагентный переход (т.е. переход без ветвления) от m_4 к m_6 необходим, так как этого требует формула $X[\beta]\psi$, которая получает значение в моменте, связанном с текущим моментом отношением R_X .

¹ Из-за громоздкости формул мы не будем давать модель полностью и ограничимся лишь ее наброском. Мы предполагаем, что читатель может свободно видеть в древовидных структурах структуры NEXТ-семантики, и наоборот. Если это не так, то в качестве справочной литературы по этому вопросу мы рекомендуем обратиться к [1] и [2].

Означивание формул (1) и (2) на модели, в основе которой лежит модельная структура, представленная на рис. 3, идет обыкновенным образом – так, что в мирах $\langle m_2/h_1 \rangle$, $\langle m_2/h_2 \rangle$, $\langle m_2/h_3 \rangle$ верно, что $\circ([\beta](\varphi \rightarrow X[\beta]\psi))$ и $\circ[\beta]\varphi$; в мире $\langle m_2/h_1 \rangle$ (где h_1 – идеальная история) верно, что $[\beta](\varphi \rightarrow X[\beta]\psi)$ и $[\beta]\varphi$; в мире $\langle m_4/h_1 \rangle$ верно, что $\varphi \rightarrow X[\beta]\psi$ и Φ (и, следовательно, $X[\beta]\psi$) (в мире $\langle m_6/h_1 \rangle$ верно, что $[\beta]\psi$); наконец, в мире $\langle m_7/h_1 \rangle$ верно, что ψ .

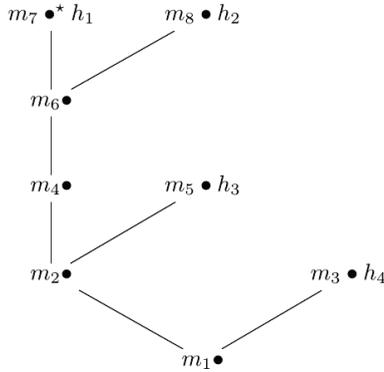


Рис. 3. Модельная структура для подразумеваемых действий

Модель, описываемая здесь, состоит из пяти временных отрезков, сменяющих друг друга один за другим, каждый из которых представлен набором моментов: $\{m_1\}$, $\{m_2, m_3\}$, $\{m_4, m_5\}$, $\{m_6\}$ и $\{m_7, m_8\}$. Формула (3), истинная в текущем моменте, дает $XX\circ[\beta]\psi$ в последующем (в m_2)¹, $-X\circ[\beta]\psi$ в m_4 ; $\circ[\beta]\psi$ и $[\beta]\psi$ – в m_6 (точнее, в мире $\langle m_6/h_1 \rangle$, так как h_1 , напомним, идеальная история); ψ – в m_7 (в мире $\langle m_7/h_1 \rangle$). Напротив, предположение о том, что в m_1 истинна $\neg(\alpha)XX\circ[\beta]\psi$ влечет в конечном итоге противоречие, так как в $\langle m_7/h_1 \rangle$ оказывается в таком случае истинна $\neg\psi$.

Промежуточные итоги

1. Агентные формулы допускают суперпозицию, через которую выражается влияние, воздействие агентов друг на друга. Тогда один агент вместе с его возможными действиями в будущем предстает в качестве объекта действия другого агента.

2. Одни агенты обладают авторитетом по отношению к другим агентам в связи с некоторым положением дел: первые могут обязать вторых к выполнению некоторых действий; вторые могут исполнить обязательство или уклониться от него.

3. Синтаксически авторитетное отношение между агентами по φ выражается как возможность отдать указание на выполнение φ – сделать так, что выполнение φ становится обязательным; семантически авторитетное отношение агента α предстает как композиция $R_\alpha \circ R_\circ$, где R_α есть отношение эффективности для α , а R_\circ – отношение, указывающее для каждого возможного мира, достижимого для α , наилучший.

¹ Так как, очевидно, временной отрезок $\{m_2, m_3\}$ разбивается R_α на два класса эквивалентности, один из которых представлен множеством миров $\{\langle m_2/h_1 \rangle, \langle m_2/h_2 \rangle, \langle m_2/h_3 \rangle\}$, или, для краткости, моментом m_2 и всеми проходящими через него историями.

2. Авторитетные отношения, данные средствами PDL

2.1. О сходстве и отличии логики действий и PDL

Если мы дадим себе труд оторваться от текущей работы и оглядеться вокруг, то сразу же увидим те множества связей, которыми окружена логика действий, те нити, что ведут из нашей области в области, смежные с ней. Потянуть за одну из них означает обнаружить на том конце, к примеру, пропозициональную динамическую логику (PDL) и её семантическую основу – систему отмеченных переходов (labeled-transitions system, LTS). Чем обусловлена эта связь? Определенным сходством и определенным различием. Первым – так как и логика действий, и PDL семантически трактуют действия как бинарные отношения на множестве возможных миров. И там и тут совершить действие означает перейти из одного мира, с одним набором истинностных значений для пропозиций, в другой мир с, возможно, другим набором истинностных значений. Вторым – так как логика действий в такого рода переходе делает акцент на результате, на том наборе истинностных значений, с которым агенту придется иметь дело после того, как действие будет совершено. Для PDL, напротив, наиболее существенным оказывается именно процесс, посредством которого осуществляется переход между мирами. Логика действий и PDL здесь различны так же, как различны для нас выполнение рутинных обязанностей (поскорей бы с этим расправиться!) и игра на музыкальном инструменте или танец (которые доставляют удовольствие сами по себе). Одновременное существование сходств и различий между логикой действий и PDL говорит о том, что они нуждаются друг в друге, так как подходят друг другу как две разные части одной и той же головоломки. Отметим и то, что на многие версии логики действий (в частности, на *STIT* и *XSTIT*, *NEXT*-семантику которой мы использовали для построения логики L_{aut}) можно смотреть как на сокращенные версии PDL. Это сокращение касается множества возможных действий, которое представлено в них не классическим для пропозициональной динамической логики набором, включающим атомарное действие, тест, последовательность, выбор и повтор, а только первым элементом этого набора. Возможность перемещения акцента в исследовании действия с его результата на его процесс предоставит в наше распоряжение семантический инструмент несколько более компактный и удобный в обращении по сравнению с *NEXT*-структурами. Там, где в действии удобно видеть прежде всего процесс, мы будем использовать для моделирования процессуальные графы. На всё это стали обращать внимание ещё совсем недавно. Так, в частности, ван Бентем и Пакуит (Pacuit) в статье [3] наметили пути слияния этих логик в одну логику, трактующую действие наиболее совершенным и всеобъемлющим образом. В этом отделе по уже намеченному пути двинемся и мы.

2.2. Авторитетные отношения как возможность делать что-то руками другого

Исследования авторитетных отношений имеют давнюю историю. Обширная содержательная проработка понятия дана, среди прочих, юристом и философом права Иосифом Разом, см., в частности, [4]. Философ и логик Ингмар Пёрн в [5] дал формально-ориентированный анализ авторитетных,

властных отношений и порожденных ими иерархических структур. Для наших целей, однако, нам потребуется совсем немного из этого наследия.

В предыдущем отделе в качестве основного инструмента анализа использовалась *NEXT*-семантика. Это позволило нам увидеть в авторитетном отношении между агентами, прежде всего, возможность дать указание – выполнить некоторое действие, такое, что агент, выступающий объектом этого действия, связывался некоторым обязательством перед тем, кто дал ему указание. Разумеется, это не единственное сущностное свойство авторитетного отношения. Обращение к PDL позволит выявить и смоделировать его другие существенные черты.

Отдавая себе отчет в некоторой искусственности, ограниченности предлагаемого в этом отделе взгляда на понятие авторитетного отношения, скажем тем не менее, что авторитетным в некотором вопросе принято называть того, кто, прежде всего, знает, как следует поступать или думать, или того, кто может сделать что-то сам.

Например: инструктор по гребле, каллиграф, учитель музыки, специалист по компьютерной верстке текста – все они авторитетны в связи со своими занятиями и навыками, так как знают, что нужно делать, чтобы плыть, писать, играть и печатать, и сами могут показать, как нужно.

Будем считать, что α может выступать авторитетом для β в связи с некоторым положением дел, таким, что ϕ , если верно, что α сам в состоянии сделать так, что ϕ .

Процессуальный граф, иллюстрирующий эту идею, дан на рис. 4.

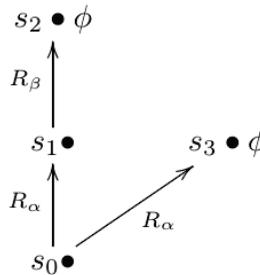


Рис. 4. Авторитетное отношение в PDL

Пусть в s_0 α отдает указание β на выполнение ϕ . В соответствии с основным определением авторитетного отношения предыдущего раздела следствием этого будет наличие авторитетного отношения между этими агентами по ϕ , подкрепленного теперь еще тем, что α может выполнить ϕ самостоятельно. Последнее верно, ведь из базового мира s_0 для α оказывается достижимым мир s_3 , где истинно ϕ (значит, в s_0 истинно $\langle \alpha \rangle \phi$, т.е. действительно α сам оказывается в состоянии сделать так, что ϕ). К процессуальному графу на рис. 4 несложно найти другой, связанный с ним отношением бисимуляции.

Считается, что если между моделями установлено отношение бисимуляции, то они отражают разными способами одни и те же явления, в нашем случае – одни и те же процессы. Мы увидим, что для интерпретации авторитетного отношения эта разница в способе представления процесса, связывающего агентов, наиболее существенна. Бисимуляцией между моделями M и N будем считать бинарное отношение E , связывающее миры s и t , принадле-

жащие соответственно моделям M и N , так, что sEt , и, для любых миров x и y , в том случае, когда xEy , оказывается верным следующее:

– в мирах x и y истинны одни и те же пропозициональные формулы (условие (1));

– если верно, что в M существует $R(x, z)$, то в N существует такой мир u , когда $R(y, u)$ и $R(z, u)$ (условие (2));

– если верно, что в N существует $R(y, u)$, то в M существует такой мир z , когда $R(x, z)$ и $R(z, u)$ (условие (3)).

Подробнее об этом см. [6. Р. 30]. Процесс и соответствующий ему граф, связанные отношением бисимуляции с процессом и графом, данными на рис. 4, изображены на рис. 5.

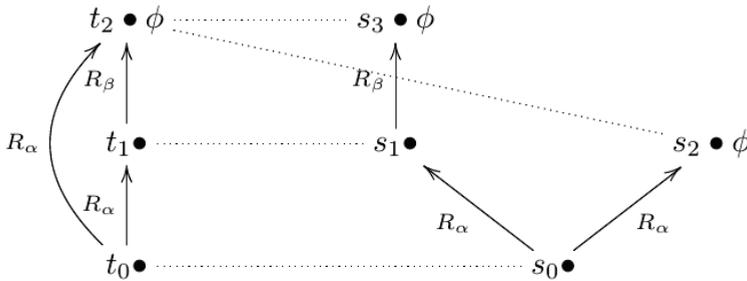


Рис. 5. Принцип *Qui facit per alium* в PDL

Покажем, что отношение, связывающее миры левого (модель M) и правого (модель N) процессуальных графов рис. 5, – это отношение бисимуляции. Процедура, сопоставляющая определение бисимуляции с конкретными процессуальными графами, заимствована нами из [7. Р. 223]. В соответствии с ней для проверки условия (2) определения бисимуляции необходимо выяснить, верно ли, что $\check{E} \circ R_M \subseteq R_N \circ \check{E}$ (где \check{E} – отношение между мирами x и y , обратное E : $\check{E}(x, y) = E(y, x)$); условие (3) проверяется посредством применения к паре моделей следующего утверждения: $E \circ R_N \subseteq R_M \circ E$. Условие (1) специальной упрощающей процедуры не требует. Тогда для отношения между моделями M и N справедливо следующее:

$\check{E} \circ R_{M, \alpha} = \{(s_0, t_1), (s_0, t_2)\}$ и $R_{N, \alpha} \circ \check{E} = \{(s_0, t_1), (s_0, t_2)\}$, т.е. для R_α выполняются условие (2);

$E \circ R_{N, \alpha} = \{(t_0, s_1), (t_0, s_2)\}$ и $R_{M, \alpha} \circ E = \{(t_0, s_1), (t_0, s_2)\}$, т.е. для R_α выполняются условие (3).

Условие (1) для M и N выполняется наиболее очевидным образом. Аналогично можно проверить выполнение условий (2) и (3) для отношения R_β .

Таким образом, между моделями M и N , т.е. между левым и правым процессуальными графами рис. 5, установлено отношение бисимуляции. Заметим, что M и N не были бы связаны отношением бисимуляции при другом обозначении переходов между мирами. Например, обозначить переход от t_0 к t_2 как R_β означает нарушить условие (3) для R_α (ведь в таком случае $E \circ R_N = \{(t_0, s_1), (t_0, s_2)\}$, а $R_M \circ E = \{(t_0, s_1)\}$); обозначить переход от s_0 к s_2 как R_β – значит нарушить для R_α условие (2).

Процесс, стоящий за понятием авторитетного отношения, за возможностью отдать указание, в PDL можно представить как транзитивный пере-

ход одного агента из первого мира в третий через второй мир, в котором действует другой агент. Авторитетное отношение выступает в PDL наиболее явным образом как возможность для одного агента добиться чего-то руками другого агента, воспользовавшись тем как инструментом. При этом полученное им воспринимается как такое положение дел, которого он добился так, как если бы действовал сам.

Подобная трактовка перекликается с известным в логике действий принципом **Q**, или *Qui facit per alium*: $[\alpha][\beta]\varphi \rightarrow [\alpha]\varphi$. Естественно, что оба процессуальных графа рис. 5 подтверждают этот принцип. В самом деле, невозможно представить ситуацию, когда $[\alpha][\beta]\varphi$ истинна в t_0 и при этом в том же мире истинна $\neg[\alpha]\varphi$: переход от t_0 к t_2 по отношению R_α гарантирует в t_0 истинность $[\alpha]\varphi$.

Ещё раз отметим в скобках компактность процессуальных графов по сравнению с другими модельными структурами, использующимися в логике действий. Челлас, при построении доказательства принципа **Q** (используя для этого модели BTS-семантики, которые гораздо ближе *NEXT*-семантике и её моделям, чем PDL и процессуальным графам) оговаривается, что рассуждение оказывается чрезмерно сложным (и чрезмерно длинным, заметим мы). Бисимуляция, таким образом, использовалась нами не только как средство, позволяющее установить сходство между моделями, выстроенными на основе *NEXT*-семантики, и процессуальными графами, но и по ее прямому назначению – для сокращения моделей. В самом деле, получившийся процессуальный граф компактен; он более наглядно представляет процессы взаимодействия агентов.

Отношения, подобные транзитивным отношениям на процессуальных графах, раскрывают существенные свойства агентов, обладающих авторитетом. Можно сказать, что здесь речь идет об агентах особого типа – тех, авторитет которых по некоторой φ сохраняется лишь в силу того, что они могут исполнить Φ самостоятельно. Можно ли представить существование агентов другого типа, таких, авторитет которых не зависит от их возможностей, как агентов, совершающих непосредственные действия? Этот вопрос намечает пути дальнейшего исследования, которое сегодня выходит за рамки статьи. Ниже мы рассмотрим лишь несколько случаев, где авторитетное отношение, представленное средствами PDL, проявляет себя особым, на наш взгляд, весьма любопытным образом.

До настоящего момента в этом разделе мы обходились без строгих определений. Того, что было в нашем распоряжении (и в распоряжении читателей), оказывалось достаточно для схватывания основной интуиции об авторитетном отношении. Но дальнейшее продвижение, безусловно, сопряжено с формальной работой в области, возникающей на стыке логики действий (в её STIT-редакции) и PDL. Необходимость в этом возникает в тот момент, когда на нашей логической сцене появляется третий агент γ .

2.3. Авторитетные отношения на тройке агентов

Пусть авторитетные отношения агентов по φ образуют такую цепочку, когда α выступает авторитетом для β , а β , в свою очередь, для γ . В этом случае, если мы попытаемся смоделировать авторитетные отношения на тройке агентов так, как мы делали это раньше, модельная структура, отражающая

процесс их взаимодействия, будет выглядеть так, как представлено на рис. 6 (левый процессуальный граф).

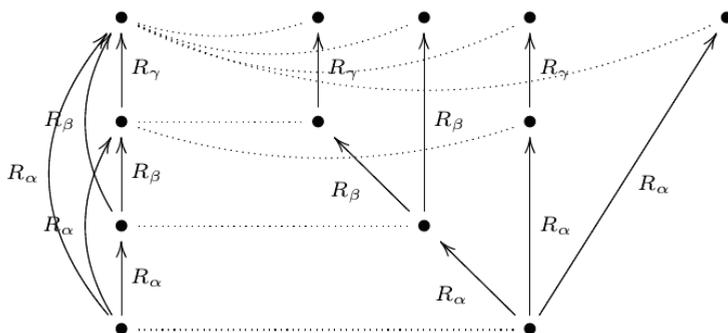


Рис. 6. Авторитетные отношения между агентами α , β и γ

Данный граф составлен из двух графов, связывающих соответственно агентов α и β , β и γ . Еще раз объявим те соображения, на которых основывается наличие переходов, указывающих на авторитетные отношения между агентами. Первое, неформальное, исходит из того содержания, которое обычно вкладывается в понятие авторитета. Второе, формализованное, представлено принципом **Q** и рождается из согласия синтаксического и семантического фрагментов выстраиваемой логики: принцип находит подтверждение на модельной структуре; та, в свою очередь, служит выражением принципа.

Принцип **Q**, приложенный к тройке агентов, даёт такие переходы: $[\alpha][\beta][\gamma]\phi \rightarrow [\alpha][\beta]\phi$ и $[\alpha][\beta]\phi \rightarrow [\alpha]\phi$. Каждая из формул – одно из действий α , множество которых, сосредоточенных в начальной точке процессуального графа, раскрывает перед ним как бы веер возможностей: выполнить ϕ самостоятельно, поручить выполнение β или γ ¹. Однако если мы попытаемся построить модель, используя для этого процессуальный граф рис. 6 (левый) или правый процессуальный граф указанного рисунка (оба они, связанные отношением бисимуляции, представляют одни и те же процессы; правый – в несколько более наглядной форме), то здесь применение указанных соображений вызовет ряд сложностей.

Наше построение снова выражает лишь интуицию авторитетных отношений на тройке агентов и в отличие от построения с использованием *NEXT*-семантики в строгом смысле не является моделью. Мы можем сопоставить с мирами процессуального графа формулы, однако такое сопоставление будет следствием интуитивного представления о процессах, связывающих тройку агентов. Оно не будет отражать главного – выбора, который делают, каждый в своем мире, α , β и γ . Процессуальный граф позволяет увидеть лишь возможные действия агентов в каждом из миров, но не то, каким образом соотносятся друг с другом их классы эквивалентности. Очевидно,

¹ Последнее справедливо, так как в противном случае пропадает одно из наших допущений. Синтаксически это можно выразить так: $\alpha[\beta]\phi \wedge [\beta][\gamma]\phi \rightarrow \alpha[\gamma]\phi$. Легко убедиться в том, что наличие в базовом мире $\neg\alpha[\gamma]\phi$ даёт $\neg\alpha[\beta]\phi$ или $\neg[\beta][\gamma]\phi$, что противоречит исходным предположениям о наличии авторитетных отношений между α и β или между β и γ .

что, например, в случае с α переходы от s_0 к s_1 и от s_0 к s_7 – это разные действия, не подпадающие, следовательно, под один и тот же класс эквивалентности. Мы сможем построить PDL-модель для тройки агентов и их авторитетных отношений тогда, когда найдем инструмент, позволяющий различать классы эквивалентности на процессуальных графах.

Еще одна сложность, которой мы не касались до настоящего момента, – это статус того, кому отдают указание: насколько он является агентом в связи с назначенной на исполнение ф? Сущность свойства быть агентом в связи с некоторым положением дел раскрывается в возможности реализовывать данное положение дел или, напротив, отказываться от действий, которые бы к нему привели. В построениях не следует надолго терять эту сущность «вторых», «третьих» и всех последующих агентов в цепочке из виду: каждый из них должен иметь возможность побыть немного Бартлби и в ответ на указание «Делай ф!» произнести: «Я бы предпочел отказаться».

2.4. Классы эквивалентности на процессуальных графах

Вопрос о представлении классов эквивалентности на процессуальных графах имеет философское основание. Найти способ их обозначения – это не то же, что решить исключительно техническую задачу. В основе этого вопроса лежит то, как мы понимаем ветвление на модельных структурах. Здесь допустимы два варианта: ветвление указывает на непредсказуемые последствия действий агента; ветвление есть само действие с его возможными последствиями, тот выбор, который осуществляет агент в некотором действительном мире. До того как мы научимся различать классы эквивалентности, эти две интерпретации формально развести не удастся. А значит, не удастся и выработать взгляд, объединяющий их. Тем не менее он необходим, так как, соединив эти два варианта толкования ветвления, мы наметим путь объединения логики действия и PDL, ведь первый из них принадлежит, собственно, PDL, в то время как второй касается в большей степени логики действия (STIT-логики в частности и преимущественно).

Отметим также, что в вопросах, касающихся ветвления, возможно, как нигде больше, проявляет себя ничем не связанная воля исследователя. Пусть агент получил приглашение на вечеринку и стоит перед выбором: идти или нет. Это, как кажется, выбор из двух, но так ли это на самом деле? Предположим, агент воздерживается от принятия решения, идти или нет, причем воздерживается сознательно, т.е. принимает выбор не выбирать. Тогда выбор из двух оказывается выбором из трех, или серией выборов из двух: решение о решении сначала и решение идти или нет потом. Исследователь вправе моделировать эту ситуацию так, как ему заблагорассудится, лишь бы выстраиваемая модель согласовывалась с ранее принятыми определениями. Мы будем поступать аналогично: наш набросок базовой модели (модельная структура) для трех агентов, данный на рис. 6, будет изменен так, чтобы учесть те особенности агентных действий, на которые мы указывали выше, и прежде всего – на возможность для агента «побыть Бартлби». Опишем инструмент, с помощью которого надеемся различать классы эквивалентности на процессуальных графах.

Пусть $w \sim_\alpha w'$ указывает на отношение, которое связывает миры w и w' так, что миром, где α делает выбор, является мир w , а миром, где содержится

описания всех последствий действия α , является мир w' . Тогда если $w \sim_\alpha w'$, то $[w]_\alpha = \{w, w'\}$.

Пусть $w \approx_\alpha w'$ указывает на отношение, которое связывает миры w и w' так, что снова миром, где α делает выбор, является мир w , а миром, который отсекается действием α в w , мир w' . Тогда если $w \approx_\alpha w'$, то $[w]_\alpha = \{w\}$ и $[w']_\alpha = \{w'\}$.

Например, если в мире w агент α принимает решение о том, пойти ему на вечеринку или не пойти, и если $w \sim_\alpha w'$ и $w \approx_\alpha w''$, то мир w' – это мир, в котором α идёт на вечеринку, а w'' – мир, отсекаемый выбором α , т.е. тот, в котором α предпочитает остаться дома.

Отношение \approx указывает на возможные альтернативы действию агента, не входящие в класс эквивалентности, определенный через отношение \sim . Средствами \sim и \approx можно благополучно указать на различия в толковании ветвления, о которых мы говорили выше: см. рис 7.



Рис. 7. Ветвление как непредсказуемые последствия выбора или как выбор

Здесь переходы, предоставленные в распоряжение α на левом процессуальном графе, попадают в разные классы эквивалентности, что означает, что в мире w α стоит перед выбором между ϕ и ψ . На правом процессуальном графе рис. 7, напротив, оба действия α подпадают под один и тот же класс эквивалентности. Можно сказать, что в этом случае α совершает одно действие, такое, что в мирах w' и w'' , составляющих один класс эквивалентности, заданный \sim_α , имеет место ϕ . Формула ψ появляется в w'' как возможное непредсказуемое последствие действия α или просто как возможное обстоятельство его действия, такого, что ϕ . Применительно к понятию авторитетных отношений, прямых и не прямых действий, осуществляемых в связи с ними, отношения \sim и \approx будут работать так.

Пусть α – агент в связи с некоторой ϕ , и пусть α обладает авторитетным отношением по ϕ к β . Будем считать, что отношения \sim и \approx вместе исчерпывают все возможности выбора агента в некоторый момент времени. Тогда наше допущение об α означает, что он может или исполнить ϕ самостоятельно, или обязать β сделать так, что ϕ . (До времени опустим возможность для α отказаться от осуществления выбора вообще.) Тогда данную ситуацию можно смоделировать так:

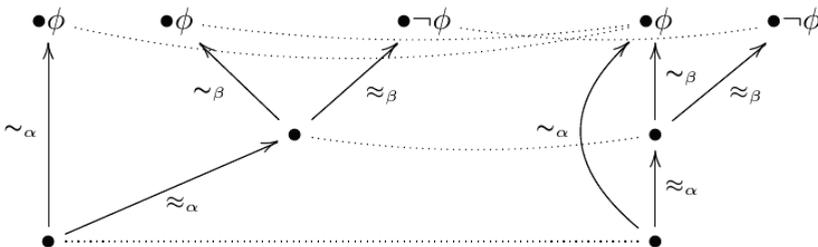


Рис. 8. Распределение классов эквивалентности при наличии авторитетного отношения между агентами

Начальный мир левого процессуального графа показывает, что перед α открывается выбор, как поступить: выполнить ϕ самостоятельно (т.е. осуществить переход к миру, где истинна ϕ , воспользовавшись для этого отношением \sim_α) или отказаться от самостоятельного выполнения, перепоручив осуществление ϕ агенту β (и осуществить переход в мир, связанный с начальным отношением \approx_α). Во втором случае уже β принимает решение о ϕ : выполнить (переход \sim_β) или отказаться (переход \approx_β).

Мы построили правый процессуальный граф рис. 8, связанный отношением бисимуляции с левым процессуальным графом, для демонстрации того, что обращение к отношениям \sim и \approx (и соответствующим модальностям $[\sim]_\alpha$, $[\sim]_\beta$, ..., $[\approx]_\alpha$, $[\approx]_\beta$, ...) сохраняет базовую интуицию данного раздела об авторитетном отношении по ϕ как о транзитивном переходе к ϕ через возможное действие другого агента (при обязательном условии возможности самостоятельного исполнения ϕ). Она, теперь выраженная средствами выстраиваемого на основе PDL фрагмента логической системы, примет следующий вид: $[\approx]_\alpha[\sim]_\beta\phi \rightarrow [\sim]_\alpha\phi$.

Будем считать, что модель M_{PDL} для представления авторитетных отношений между агентами в PDL дана набором $\langle W, \{\sim\}_\Gamma, \{\approx\}_\Gamma \rangle$, где $\{\sim\}_\Gamma$ и $\{\approx\}_\Gamma$ есть множества отношений на W для агентов из $\Gamma = \{\alpha, \beta, \dots\}$, в каждое из которых вкладывается указанное выше содержание. Тогда:

$M_{PDL}, w \models [\sim]_\alpha\phi$ е.т.е. для всех w' , таких, что $w \sim_\alpha w'$, верно, что $w' \models \phi$ (выбор α в w) и

$M_{PDL}, w \models [\approx]_\alpha\phi$ е.т.е. для всех w' , таких, что $w \approx_\alpha w'$, верно, что $w' \models \phi$ (альтернативный выбор α в w).

Эти определения позволяют показать то, как соотносятся классы эквивалентности для ситуации авторитетных отношений между тройкой агентов α, β и γ .

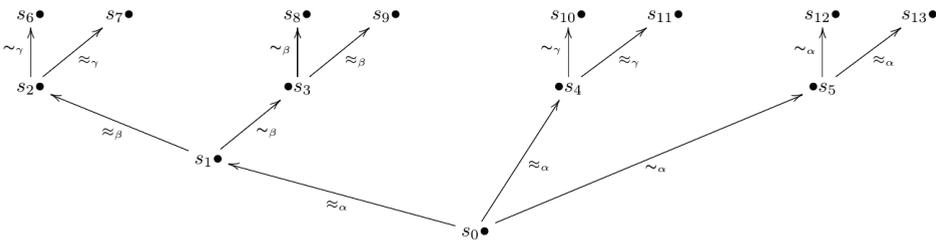


Рис. 9. Авторитетные отношения на тройке агентов данные средствами \sim и \approx

Процессуальный граф на рис. 9 дает описание следующих вариантов развития событий в ситуации наличия авторитетного отношения по ϕ между тройкой агентов α, β и γ :

$[\approx]_\alpha[\approx]_\beta[\sim]_\gamma\phi$ – α отдаёт указание β , β отдаёт указание γ , γ выполняет указание;

$[\approx]_\alpha[\sim]_\beta[\sim]_\beta\phi$ – α отдаёт указание β , β отдаёт указание β (так), β выполняет указание;

$[\approx]_\alpha[\sim]_\gamma\phi$ – α отдаёт указание γ , γ выполняет указание;

$[\sim]_\alpha[\sim]_\alpha\phi$ – α отдаёт указание α , α выполняет указание;

$[\approx]_\alpha[\approx]_\beta[\approx]_\gamma$ – α отдаёт указание β , β отдаёт указание γ , γ отказывается выполнить указание;

$[\approx]_\alpha[\sim]_\beta[\approx]_\beta$ – α отдаёт указание β , β отдаёт указание β , β отказывается выполнить указание;

$[\approx]_\alpha[\approx]_\gamma$ – α отдаёт указание γ , γ отказывается выполнить указание;

$[\sim]_\alpha[\approx]_\alpha$ – α отдаёт указание α , α отказывается выполнить указание.

Обратим внимание читателя на несколько интересных моментов. Во-первых, мы ввели в модель «ветки Бартлби» для агентов: теперь α , β и γ могут отказаться от выполнения ϕ , что делает их агентами в связи с ϕ в еще большей степени, чем раньше, когда таких веток не было. Во-вторых, на нашей модели агенты обладают возможностью отдавать указание самим себе. Это делают β и α тогда, когда решают выполнить ϕ самостоятельно и не назначать к выполнению ϕ другого агента, связанного с ними авторитетным отношением по ϕ . Агент γ не отдаёт указание самому себе на выполнение ϕ , так как он не принимает такого решения, которое бы предполагало отдачу указания на выполнение ϕ другому агенту. Из второго замечания следует третье: мы интерпретируем отношение \approx двояким образом, в зависимости от того, связан ли некоторый агент, осуществляющий по нему переходы, авторитетным отношением с другим агентом. Если это так, то для такого агента переход по \approx означает отказ от самостоятельного исполнения и отдачу указания на исполнение другому агенту. Если авторитетного отношения нет, то переход по \approx означает отказ от выполнения некоторого (прямого) действия. В-четвертых, заметим, что теоретически мы можем отказаться от ограничения на возможные действия агента, связанного авторитетным отношением, и уклониться от дихотомии «делай сам или отдай указание другому». В самом деле, агент, связанный авторитетным отношением по ϕ с некоторым другим агентом, может отказаться от выполнения ϕ , но при этом не поручать выполнение ϕ другому агенту. Аналогично тот же агент может отказаться от отдачи поручения другому агенту, но при этом и не браться за выполнение ϕ самостоятельно. Пусть α – агент, который связан авторитетным отношением с β по ϕ . Тогда веер возможностей для α раскрывается так:

$[\sim]_\alpha\phi$ – выполнить ϕ самостоятельно;

$[\approx]_\alpha[\sim]_\alpha\phi \wedge \neg[\approx]_\alpha[\sim]_\beta\phi$ – отказаться от выполнения ϕ , при этом не поручая ϕ другому;

$[\approx]_\alpha[\sim]_\beta\phi$ – поручить выполнение ϕ агенту β ;

$[\approx]_\alpha[\approx]_\alpha\phi \wedge \neg[\sim]_\alpha\phi$ – отказаться от поручения ϕ агенту β , но при этом не выполнять ϕ самостоятельно.

В-пятых, предположим, что на модели (см. рис. 9) в мире s_0 истинна формула $[\approx]_\alpha[\approx]_\beta[\sim]_\gamma\phi$. Тогда по определению в s_1 и в s_4 истинна $[\approx]_\beta[\sim]_\gamma\phi$. Обращает на себя внимание второе: по определению $[\approx]_\beta[\sim]_\gamma\phi$ истинна в s_4 е.т.е. для всех миров, связанных с миром s_4 отношением \approx_β , верно, что в них истинна $[\sim]_\gamma\phi$. Но таких миров на модели нет! Тогда s_4 оказывается миром, который обычно называют *dead-end*. Но это своё качество s_4 проявляет не безусловным образом, а в связи с отношениями \approx_β и \sim_β (то же справедливо и для отношений \approx_α и \sim_α). Мир s_4 – это *dead-end* для агентов β и α . Поэтому в s_4 истинными оказываются любые формулы, начинающиеся с $[\approx]_\beta$ или с $[\sim]_\beta$ (снова, то же для α), подобно тому как в *dead-end*-мирах оказываются истинными любые формулы, начинающиеся с оператора \square . Подобное положение дел да-

ет два следствия: во-первых, наличие *dead-end*-миров для некоторого набора отношений или агентов позволяет сохранить истинность базовых формул, как мы только что убедились на примере формулы $[\approx]_{\alpha}[\approx]_{\beta}[\sim]_{\gamma}\varphi$; во-вторых, такие миры, и бесконечное множество формул, которое они порождают, приводят к ситуации, когда на моделях оказывается истинным слишком многое, что может сказаться самым пагубным образом на истинности базовых формул, непосредственно выражающих непрямые действия и авторитетные отношения.

Заключительные замечания. Дальнейшее исследование может развиваться по самым разным направлениям или оставаться в русле указанного и сосредоточиться на частностях (которые, в свою очередь, могут дать новые направления). Мы укажем на пару таких сюжетов.

Первый из них связан с типами агентов, в частности с таким, который сохранял бы авторитетное отношение по некоторой формуле φ даже после того, как он утратил возможность выполнять φ самостоятельно. Это тем более интересно для случаев, когда авторитетными отношениями связано достаточно большое количество агентов. Продолжает ли и в этом случае действовать принцип **Q**, и всегда ли в таких случаях действие последнего агента в цепочке равносильно действию первого, когда тот совершает некоторое φ как бы его или её руками?

В рамках второго возможного сюжета интересно найти способ показать на процессуальных графах различные действия агентов, связанные с уклонением от выполнения назначенного. До настоящего времени мы трактовали такое уклонение слишком узко: как выполнение $\neg\varphi$. Однако очевидно, что спектр подобных действий значительно шире. Возможно, уклонение некоторого β следует понимать как обратный переход между мирами по отношению к тому переходу, который делает агент α , связанный с β авторитетным отношением. Наложение обязательств, понуждение к некоторому действию тогда не предстают как одномоментные действия. Они, напротив, являются результатом целой серии конкурентных действий агентов и подобны игре в теннис: отскочивший от поля мяч (возможное обязательство) всегда можно выбить обратно – на поле противника.

Литература

1. Broersen J. A complete stit logic for knowledge and action, and some of its applications // Declarative Agent Languages and Technologies VI, p. 47–59. Springer, 2009.
2. Broersen J. Making a start with the stit logic analysis of intentional action // Journal of philosophical logic. 2011. 40(4). P. 499–530.
3. Benthem J. van, Pacuit E. Connecting Logics of Choice and Change // Nuel Belnap on Indeterminism and Free Action. Springer. 2014. P. 291–314.
4. Raz J. Practical Reason and Norms. Oxford, 1999. (Первое издание 1975.)
5. Pörn I. The Logic of Power. Oxford, 1970.
6. Handbook of Modal Logic / ed. by P. Blackburn, J. van Benthem, F. Wolter. Elsevier, 2007.
7. Benthem J. van, Ditmarsch H. van, Eijck J. van, Jaspars J. Logic in Action. 2012.

Gleb V. Karpov, Saint Petersburg University (Saint Petersburg, Russian Federation).
Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science. 2019. 47. pp. 47–62.

DOI: 10.17223/1998863X/47/5

INDIRECT ACTIONS AND AUTHORITY RELATIONS IN STIT AND PDL

Keywords: logic of actions; STIT; NEXT-semantics; PDL; authority relations.

The article investigates the notion of an authority relation between agents from two perspectives. Its first part argues that the existence of an authority relation between agents, with respect to some state of affairs, means that an agent who is in authority has an ability to command to act in order to see to it that such state of affairs is true. A command of that kind is modelled in NEXT-semantics. Here an impact of one agent on another is pictured as a composition of accessibility relations between the worlds. A case of an embedded authority relation and a case of presupposed conditions are investigated. The second part of the article argues that an authority relation between agents, with respect to some state of affairs, exists due to a personal and immediate ability of an agent, who is in authority, to make that state of affairs true. Here an authority relation is interpreted on process graphs as a transitive move to the third world from the first one (where an agent who is in authority acts) via the second world (where an agent who is under authority acts) or as an ability of one agent to use others as a proxy for his or her actions. Such an ability is modelled for three agents by means of a special relation that can distinguish equivalent classes on process graphs. Finally, the article introduces some semantic ideas which are to make further synthesis of the logic of action and PDL fruitful.

References

1. Broersen, J. (2009) A complete STIT logic for knowledge and action, and some of its applications. *Declarative Agent Languages and Technologies*. VI. pp. 47–59. DOI: 10.1007/978-3-540-93920-7_4.
2. Broersen, J. (2011) Making a start with the stit logic analysis of intentional action. *Journal of Philosophical Logic*. 40(4). pp. 499–530.
3. Benthem, J. van & Pacuit, E. (2014) Connecting Logics of Choice and Change. In: Müller, T. (ed.) *Nuel Belnap on Indeterminism and Free Action*. Springer. pp. 291–314. DOI: 10.1007/978-3-319-01754-9_14
4. Raz, J. (1999) *Practical Reason and Norms*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198268345.001.0001
5. Porn, I. (1970) *The Logic of Power*. Oxford: Barnes & Noble.
6. Blackburn, P., Benthem, J. van & Wolter, F. (eds) (2007) *Handbook of Modal Logic*. Elsevier.
7. Benthem, J. van, Ditmarsch H. van, Eijck J. van & Jaspars, J. (2012) *Logic in Action*. [s.l., s.n.].