Introspection as an Action in Relational Models

Raul Fervari[a], Fernando R. Velázquez-Quesada[b]

[a] *FAMAF - Universidad Nacional de Córdoba & CONICET*
Argentina
[b] *Institute for Logic, Language and Computation, Universiteit van Amsterdam*
The Netherlands

Abstract

This work studies positive and negative introspection not as 'static' properties an agent might or might not have, but rather as epistemic actions that change the agent's knowledge. The proposed actions include not only operations for achieving full introspection (first with respect to all formulas, and then with respect to a particular χ), but also acts for increasing the agent's introspection by only one degree. In all cases, the actions are represented as model update operations, with matching modalities for expressing the operations' effects. Sound and complete axiom systems are provided in most cases, and some properties of the operations are explored.

Keywords: positive introspection, negative introspection, epistemic logic, dynamic epistemic logic.

1. Introduction

One of the reasons of the widespread use of *epistemic logic* (*EL*; [1]. something, she knows that she knows it) and *negative introspection* (if the agent does not know something, she knows that she does not know it). One of the main advantages of the standard *EL* semantic structure, relational models, is that these two properties correspond, at the level of frames, to simple relational properties. In order to deal with an agent with full positive introspection, it is enough for her corresponding indistinguishability relation to be transitive; in order to deal with an agent with full negative introspection, it is enough for such relation to be Euclidean.

As it is the case with most properties of an agent's knowledge in *EL* (e.g., the famous logical omniscience problem; [2, 3]), these two properties have been the subject of an extensive discussion. Positive introspection, also known as *the KK principle*, is supported by [1, page 111]: "*all those circumstances which*

Email addresses: `rfervari@conicet.gov.ar` (Raul Fervari),
`F.R.VelazquezQuesada@uva.nl` (Fernando R. Velázquez-Quesada)

would justify one in saying 'I know' will also justify one in saying 'I know that I know' ". [4, page 173] also considers the principle can be defended *"in the context of the idealizations we are making"*. On the other hand, some authors have provided counter-examples to positive introspection. For instance, [5] asks to consider a question and an agent who knows the question's correct answer, but still fails to answer it at time t; however, she answers the question at time $t + 1$. Because the agent answered the question at $t + 1$, it is obvious that she knows the answer; but she could not answer it at t, which means she did not know she knew it. On a more abstract level, [6] argues that knowledge implies understanding; as a consequence, knowledge of knowledge of p implies understanding of knowledge of p. This can only happen if the agent possesses an adequate theory of knowledge. But, as knowledge of p does not require this, the KK-principle is not valid. (See [7] for an answer to these two arguments.) More recently (and famously), [8] has argued that, in contexts involving inexact knowledge (vagueness or margins of errors), positive introspection should not be valid, because it leads to the conclusion that no agent can have both inexact knowledge and positive introspection.[1] The reader interested in the discussion is also referred to [11], which discusses positive introspection from the point of view of various definitions of knowledge.

Negative introspection has been the subject of an even stronger criticism. It was already rejected in [1], where it is shown that it entails (together with intuitive acceptable principles) that if a given statement is the case, then the agent knows that she considers it possible.[2] The author finds this principle unacceptable, and his reason (found clear and decisive by [4]) is that it implies that, by reflection alone, one could come to see the truth of anything that was compatible with one's knowledge.[3] Some authors, such as [12], have argued that negative introspection cannot be consistently adopted for a strong notion of knowledge which treats knowledge as true conviction. Some others have pointed out that, for example, if such property held for the participants in Socratic dialogues, the whole procedure would be unnecessary.

As this (by no means exhaustive, and in fact rather) brief summary of the literature shows, there are different arguments supporting and/or rejecting positive and negative introspection. The present manuscript is a contribution to this discussion from a different perspective: instead of arguing about whether an agent's knowledge should or should not have these (epistemic) properties, it

[1] See [9] for a proposal introducing a non-standard semantics for epistemic logic, *centered semantics*, which makes compatible a notion of inexact knowledge and the principles of positive and negative introspection. See [10] for its dynamic extension, which can be used to account for the dynamics of reflection on one's margins.

[2] In symbols, $\varphi \to \Box \Diamond \varphi$: the so-called B axiom which corresponds, at the level of frames, to the symmetry of the indistinguishability relation. A reflexive and Euclidean relation is also symmetric, thus the conclusion that B follows from negative introspection and 'intuitive acceptable principles' (the truthfulness of knowledge).

[3] The contrapositive, $\Diamond \Box \varphi \to \varphi$: anything the agent considers possible to know is true.

takes a *dynamic* perspective, discussing the *epistemic actions* an agent would need to perform in order to reach such informational state. Indeed, when neither transitivity nor Euclideanity are assumed, one can build a relational model in which the agent lacks positive and negative introspection, making her knowledge closer to that of 'real life' fallible agents. But, as in real life, not being introspective about one's own knowledge or ignorance should not imply one will never be: the agent can perform introspection *actions* that will make her realize what she knows and what she does not know. Thus, the main goal is not to provide clear and decisive arguments supporting or rejecting these principles. Instead, the aim is to provide formal tools that allow the discussion to incorporate a *dynamic* perspective, involving the properties of the actions an agent would need in order to reach positive and/or negative introspection.

It is worthwhile to emphasise, although the provided actions intend to make the agent reach an introspective state, they might have unintended side-effects that might not conform to our informal ideas of obtaining higher-order knowledge by introspection. A particularly important issue, discussed through the text, is whether such actions preserve the original *propositional* knowledge.

1.1. On dynamic *approaches*

Dynamic approaches on epistemic issues are not new. Such perspective has proved to be useful, for example, in providing explanations for various epistemic paradoxes (Moore's paradox [13]; Fitch's knowability paradox [14]; the Surprise Examination paradox [15]). It has been also useful in the study of epistemic properties by focussing not on whether they should hold, but rather on the different actions the agent can perform to achieve them, as this proposal does. For example, some works have understood the aforementioned logical omniscience not as a 'static' property, but rather as the eventual result of some action. Some of these proposals are based on awareness raising and 'syntactic' inference steps, either in models based on sets of formulas [16, 17, 18, 19], or else in awareness relational models [20, 21]; some others are based on dynamics of evidence or deductive inference in neighbourhood models [22, 23].

Within this dynamic approach, a crucial aspect is the way the involved actions are depicted. One possibility is representing them as *transitions* within a fixed model (e.g., automata theory –[24]– or, in a logical setting, *propositional dynamic logic –PDL* [25]– and *epistemic temporal logic –ETL* [26, 27, 28]–). Another alternative, the one followed here, is the *dynamic epistemic logic* approach (*DEL*; [29, 30]), in which models represent only the concept(s) under study, and then changes in the concept are represented as *model operations* that alter the given structure, thus potentially affecting the given concepts. Of course, relational models can be modified in different ways. One possibility is to change the *domain*, either by adding only one world (e.g., [31]), or else by using the more general 'action-model based' strategy that makes 'restricted copies' of the worlds in the original model (thus allowing also the shrinking of the domain; see, e.g., [32, 33, 34]). This work focuses rather on operations changing the model's *accessibility relation*, with examples including actions for belief revision and/or (reliability based) preference change [35, 36, 37, 38], a more 'abstract' edge-deleting sabotage operation [39], its edge-adding and edge-swapping relatives [40, 41, 42, 43] and the general arrow update operation of [44]. The specific operations used here borrow ideas from most of them.

1.2. Outline

This article extends [45] in the following ways. First, it has already provided a brief review on the philosophical literature on introspection. Then, it elaborates on the discussion on the previously provided operations for achieving *full* introspection. More importantly, it proposes model operations for increasing introspection 'by one degree' (Section 5), complementing the *full* introspection operations. Finally, it includes previously omitted proofs.

The text is organised as follows. Section 2 introduces basic definitions about epistemic logic and the extensions that will be required for axiomatisation purposes. Then, while Section 3 defines model operations to achieve full positive and negative introspection about all formulas (*general* introspection operations), Section 4 focuses on operations to achieve full positive and negative introspection about a given formula (*particular* introspection operations). Section 5 explores similar operations that focus, instead, on increasing the agent's introspection 'by one degree' (thus working locally, with respect to a given evaluation point, rather than globally). In all cases we study important properties of the operations, providing also sound and complete axiomatisations for most of their respective modalities. Finally, Section 6 draws conclusions, also delineating directions for further work.

2. Basic definitions

This section recalls the basic definitions of the basic *EL*. It also recalls extensions of this framework that will be useful when providing axiom systems for modalities representing the introspection operations.[4] Throughout this paper, let $\mathtt{P}$ be a countable set of atomic propositions.

Definition 2.1 (Relational frame, relational model, relational state) A *relational frame* is a tuple $F = \langle W, R \rangle$ with W a non-empty set of possible worlds and $R \subseteq (W \times W)$ a binary relation, the agent's *indistinguishability* relation. A *relational model* is a tuple $M = \langle F, V \rangle$ with F a relational frame and $V : \mathtt{P} \to \wp(W)$ the atomic valuation. A tuple (M, w) with M a relational model and w a world in it (the *evaluation point*) is called a *relational state*. ◀

Note how the accessibility relation is not required to satisfy any property.

[4] Here, the term *modality* is used for operator symbols that require a different relational state to be semantically evaluated. This includes the 'dynamic' modalities of Sections 3 to 5, but also standard modal operators as $\Box$ and $\Diamond$, as they require a change in the evaluation point.

Definition 2.2 (Language $\mathcal{L}_{\Diamond}$) Formulas φ, ψ of $\mathcal{L}_{\Diamond}$ are given by

$$\varphi, \psi ::= p \mid \neg\varphi \mid \varphi \vee \psi \mid \Diamond \varphi,$$

with $p \in \mathsf{P}$. Other Boolean connectives and constants $(\wedge, \rightarrow, \leftrightarrow, \top, \bot)$ as well as the modality $\Box$ are defined as usual ($\Box \varphi := \neg \Diamond \neg\varphi$ for the latter). Formulas of the form $\Box \varphi$ are read as *"the agent knows φ"*.

Formulas of $\mathcal{L}_{\Diamond}$ are interpreted in relational states in the standard way; here we just make explicit the cases of atomic propositions and the 'diamond' modality. Let (M, w) be a relational state with $M = \langle W, R, V \rangle$; then,

$$\begin{array}{lll} (M, w) \Vdash p & \text{iff}_{def} & w \in V(p), \\ (M, w) \Vdash \Diamond \varphi & \text{iff}_{def} & \text{there is } u \in W \text{ such that } Rwu \text{ and } (M, u) \Vdash \varphi. \end{array}$$

A formula φ is *true* at w in M when $(M, w) \Vdash \varphi$, and is *valid* (notation: $\Vdash \varphi$) when it is true in every world w of every model M. The function $[\![\varphi]\!]^M := \{w \in W \mid (M, w) \Vdash \varphi\}$, returning the set of worlds of a given model M in which a given formula φ is the case (the *truth-set* of φ at M) will be useful. ◀

Theorem 1 (Axiom system for $\mathcal{L}_{\Diamond}$) *As it is well-known (see [46, 47] for details), the axiom schemes and rules of Table 1, denoted by $\mathsf{L}_{\Diamond}$, form a sound and strongly complete axiom system for formulas of the language $\mathcal{L}_{\Diamond}$, characterising those that are valid with respect to relational model.*[5] ■

Prop	$\vdash \varphi$ for φ a propositional tautology	*MP*	If $\vdash \varphi \rightarrow \psi$ and $\vdash \varphi$, then $\vdash \psi$
K	$\vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi)$	*Nec*	If $\vdash \varphi$, then $\vdash \Box \varphi$
Dual	$\vdash \Diamond \varphi \leftrightarrow \neg \Box \neg\varphi$		

Table 1: Axiom system $\mathsf{L}_{\Diamond}$, for formulas in $\mathcal{L}_{\Diamond}$ w.r.t. relational models.

Extended languages The following sections will study languages with additional modalities standing for model-update operations representing introspection actions. In order to introduce their corresponding axiom systems, two extensions of the basic epistemic language will be useful. The first one adds a transitive closure modality.

Definition 2.3 (Language $\mathcal{L}_{\Diamond,\Diamond^+}$) The language $\mathcal{L}_{\Diamond,\Diamond^+}$ extends $\mathcal{L}_{\Diamond}$ with the modality $\Diamond^+$. For its semantic interpretation, let (M, w) be a relational state with $M = \langle W, R, V \rangle$, and recall that the relation R^+, the transitive closure of R (i.e., the smallest transitive relation containing R), is defined as

$$R^+ := \bigcup_{n \geq 1} R^n,$$

[5] Although the *Dual* axiom might look superfluous, it cannot be omitted since $\mathcal{L}_{\Diamond}$ is presented with $\Diamond$ as primitive and $\Box$ as an abbreviation. For further discussions, see [47, 48].

with $R^1 := R$, $R^{k+1} := R^k \circ R$ and '$\circ$' the relational composition.[6] Then,

$$(M, w) \Vdash \oplus\varphi \quad \text{iff}_{def} \quad \text{there is } u \in W \text{ with } R^{+}wu \text{ and } (M, u) \Vdash \varphi.$$

The dual modality $\boxplus$ is defined in the usual way ($\boxplus\varphi := \neg\oplus\neg\varphi$). ◀

For a sound and complete axiom system for $\mathcal{L}_{\Diamond,\oplus}$, the system $\mathsf{L}_{\Diamond}$ is not enough anymore. The following theorem provides the 'missing' pieces.

Theorem 2 (Axiom system for $\mathcal{L}_{\Diamond,\oplus}$ [27]) *The axioms and rules of Table 1 and Table 2, denoted by $\mathsf{L}_{\Diamond,\oplus}$, form sound and weakly complete axiom system for formulas of $\mathcal{L}_{\Diamond,\oplus}$ valid in relational models.* ■

Note how, while $FP_{\boxplus}$ indicates that $\oplus$'s associated relation is a fixed point, $Ind_{\boxplus}$ indicates that such relation is the smallest one.

$Dual_{\boxplus}$	$\vdash \oplus\varphi \leftrightarrow \neg\boxplus\neg\varphi$		
$FP_{\oplus}$	$\vdash \oplus\varphi \;\leftrightarrow\; \Diamond(\varphi \vee \oplus\varphi)$	$Ind_{\oplus}$	If $\vdash \varphi \to \Box(\psi \wedge \varphi)$, then $\vdash \varphi \to \boxplus\psi$

Table 2: Axioms and rule for the modality $\oplus$.

The second extension is $\mathcal{L}_{PDL^{\triangleleft,?}}$, adding not only *regular expressions* for building complex relations from the basic one, but also a converse modality.[7]

Definition 2.4 (Language $\mathcal{L}_{PDL^{\triangleleft,?}}$) Formulas φ, ψ and program expressions α, β in $\mathcal{L}_{PDL^{\triangleleft,?}}$ are given, respectively, by

$$\varphi, \psi ::= p \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\alpha\rangle\varphi \qquad \alpha, \beta ::= \triangleright \mid \triangleleft \mid \alpha \cup \beta \mid \alpha ; \beta \mid \alpha^{*} \mid ?\varphi,$$

with $p \in \mathsf{P}$. The fragment of $\mathcal{L}_{PDL^{\triangleleft,?}}$ without ? is called $\mathcal{L}_{PDL^{\triangleleft}}$. Given $M = \langle W, R, V\rangle$, the semantic interpretation of the new modality $\langle\alpha\rangle\varphi$ is defined as

$$(M, w) \Vdash \langle\alpha\rangle\varphi \quad \text{iff}_{def} \quad \text{there is } u \in W \text{ such that } R_{\alpha}wu \text{ and } (M, u) \Vdash \varphi.$$

Note how such semantic interpretation relies on the relation R_α which is defined inductively, for every program expression α, as

$$\begin{array}{lll} R_{\triangleright} := R, & R_{\alpha\cup\beta} := R_\alpha \cup R_\beta, & R_{\alpha^*} := (R_\alpha)^*, \\ R_{\triangleleft} := \breve{R}, & R_{\alpha;\beta} := R_\alpha \circ R_\beta, & R_{?\varphi} := \mathrm{Id}^M_\varphi, \end{array}$$

[6] Given $R_1, R_2 \subseteq (W \times W)$, their composition $R_1 \circ R_2 \subseteq (W \times W)$ (note the parameters' order) is the relation given by $\{(w, v) \mid \text{there is } u \in W \text{ such that } R_1wu \text{ and } R_2uv\}$.

[7] In classical presentations of $\mathcal{L}_{PDL^{\triangleleft,?}}$, the converse modality works on complex expressions, unlike here where we introduce it at atomic level. In [49] it is shown that both presentations are equivalent.

for $\reflectbox{R}$ the converse of R (also called its inverse), $\mathrm{Id}_{\varphi}^{M}$ the identity relation over the set of worlds satisfying φ in M, and $(\cdot)^{*}$ the reflexive and transitive closure operation.[8] ◀

Theorem 3 (Axiom system for $\mathcal{L}_{PDL^{\triangleleft,?}}$ [50, 51, 52, 25]) *The axioms and rules of Table 3 form sound and weakly complete axiom system (denoted by $\mathsf{L}_{PDL^{\triangleleft,?}}$) for formulas of $\mathcal{L}_{PDL^{\triangleleft,?}}$ with respect to relational models.* ■

Note how, while *Prop* and *MP* are just as in Table 1, K_α and Nec_α are the generalisation of the former K and Nec to any program expression α. The rest of the axioms take care of the converse modality and the program operators. The axiom system for the fragment $\mathcal{L}_{PDL^{\triangleleft}}$, given by $\mathsf{L}_{PDL^{\triangleleft,?}}$ minus axiom ?, is denoted by $\mathsf{L}_{PDL^{\triangleleft}}$.

Prop	$\vdash \varphi$ for φ a propositional tautology	*MP*	If $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$
K_α	$\vdash [\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$	Nec_α	If $\vdash \varphi$, then $\vdash [\alpha]\varphi$
$Dual_\alpha$	$\vdash \langle\alpha\rangle\varphi \leftrightarrow \neg[\alpha]\neg\varphi$	?	$\vdash \langle ?\varphi\rangle\psi \leftrightarrow (\varphi \wedge \psi)$
$\triangleleft_1$	$\vdash \varphi \to [\triangleright]\langle\triangleleft\rangle\varphi$	$\triangleleft_2$	$\vdash \varphi \to [\triangleleft]\langle\triangleright\rangle\varphi$
$\cup$	$\vdash \langle\alpha \cup \beta\rangle\varphi \leftrightarrow (\langle\alpha\rangle\varphi \vee \langle\beta\rangle\varphi)$	;	$\vdash \langle\alpha\,;\beta\rangle\varphi \leftrightarrow \langle\alpha\rangle\langle\beta\rangle\varphi$
$*_1$	$\vdash \langle\alpha^*\rangle\varphi \leftrightarrow (\varphi \vee \langle\alpha\rangle\langle\alpha^*\rangle\varphi)$	$*_2$	$\vdash [\alpha^*](\varphi \to [\alpha]\varphi) \to (\varphi \to [\alpha^*]\varphi)$

Table 3: Axiom system $\mathsf{L}_{PDL^{\triangleleft,?}}$, for formulas in $\mathcal{L}_{PDL^{\triangleleft,?}}$ w.r.t. relational models.

The function $[\![\cdot]\!]^{M}$ is defined for both extensions in the obvious way.

3. General introspection

This section studies operations to achieve *general* positive and negative introspection (that is, *full* positive and negative introspection about *all* formulas). The goal is, thus, to obtain global properties, which can be achieved by making the original accessibility relation satisfy the conditions yielding such properties.

3.1. General positive introspection

When looking for a model operation depicting an action of full positive introspection, the first idea is the following: transitivity makes the positive introspection axiom $\Box\varphi \to \Box\Box\varphi$ valid, so make the accessibility relation transitive.

Definition 3.1 (General positive introspection operation) Let M be a relational model $\langle W, R, V\rangle$. The *general positive introspection* operation yields the model $M^{+} = \langle W, R^{+}, V\rangle$, with R^{+} being R's transitive closure. ◀

The following modality allows to describe this operation's effects.

[8] Formally, $\reflectbox{R} := \{(v,u) \mid Ruv\}$, $\mathrm{Id}_{\varphi}^{M} := \{(u,u) \mid u \in [\![\varphi]\!]^{M}\}$ and $R^{*} := \mathrm{Id}_{\top}^{M} \cup R^{+}$.

Definition 3.2 The language $\mathcal{L}_{\Diamond,+}$ extends $\mathcal{L}_{\Diamond}$ with the modality $\langle + \rangle$. For its semantic interpretation, let (M, w) be a relational state. Then,

$$(M, w) \Vdash \langle + \rangle \varphi \quad \text{iff}_{def} \quad (M^{+}, w) \Vdash \varphi.$$

In words, the agent can perform an act of general positive introspection after which φ is the case, $(M, w) \Vdash \langle + \rangle \varphi$, if and only if, after the action, φ is the case, $(M^{+}, w) \Vdash \varphi$. Note that the model operation is functional and its associated modality lacks a precondition; hence, the dual modality $[+] \varphi := \neg \langle + \rangle \neg\varphi$ is equivalent to $\langle + \rangle$. In other words, $\Vdash [+] \varphi \leftrightarrow \langle + \rangle \varphi$ (self-duality). ◀

Note the difference between the just defined model-change modality $\langle + \rangle$ and the transitive closure modality $\oplus$ of before. When evaluated on a given (M, w), the latter is semantically interpreted via a further relation R^{+}, but in the same model M. Nevertheless, in the same situation, the former is semantically interpreted via the basic relation R, but in a model M^{+} in which R has changed. It is the shift to a model in which the relation is different what accounts for the change in the agent's knowledge.

Axiom system When providing an axiom system for a modality representing a model operation (a *dynamic* modality), a useful strategy is to provide *reduction axioms*: valid formulas and validity-preserving rules indicating how to translate a formula with occurrences of this model-changing modality (a formula in the 'dynamic' language) into a provably equivalent one without them (a formula in the 'basic' language).[9] Then, while soundness follows from the validity and validity-preserving properties of the new axioms and rules, completeness follows from the completeness of the axiom system for the basic language, as the reduction axioms define a validity-preserving translation from the 'dynamic' language into the 'basic' one. The reader is referred to Chapter 7 of [29] (cf. [53]) for an extensive explanation of this technique.

Note how this strategy relies on the expressivity of the basic language: the existence of reduction axioms indicates that such language is expressive enough to describe the changes the model operation induces. In this case, $\mathcal{L}_{\Diamond}$ is not expressive enough to deal with the changes the general positive introspection operation brings about: it can describe what holds in worlds that can be reached from the evaluation point in a given *fixed* number of R-steps, but it cannot describe what holds in worlds that can be reached by an *undetermined* (finite non-zero) number of them (i.e., a single R^{+}-step). Thus, in order to provide reduction axioms for $\langle + \rangle$, the basic language will be rather $\mathcal{L}_{\Diamond,\oplus}$ (Definition 2.3). The extra modality $\oplus$ indeed increases the language's expressive power.

Fact 3.1 *$\mathcal{L}_{\Diamond,\oplus}$ is more expressive than $\mathcal{L}_{\Diamond}$.*

[9] It has been suggested [30] that the term *recursion axioms* is more appropriate, as it describes the recursive nature of the translation the axioms define. Still, this text will use the more standard term *reduction axioms*.

Proof. It is well-known that $\mathcal{L}_{\Diamond}$ can be translated into first-order logic, which makes it compact. However, $\mathcal{L}_{\Diamond,\oplus}$ is not. Consider the infinite set $\Gamma = \{\oplus\, p\} \cup \{\Box^n \neg p \mid n \geq 1\}$ with each $\Box^n$ a standard modality for the relation R^n. Every finite subset of Γ is satisfiable, but Γ is not. ■

The language $\mathcal{L}_{\Diamond}$ cannot express what holds in worlds that are reached via the transitive closure of the basic relation R, so there are no reduction axioms translating $\mathcal{L}_{\Diamond,+}$ into $\mathcal{L}_{\Diamond}$. However, $\mathcal{L}_{\Diamond,\oplus}$ extends $\mathcal{L}_{\Diamond}$'s expressivity in the 'right direction', allowing a translation from $\mathcal{L}_{\Diamond,\oplus,+}$ into $\mathcal{L}_{\Diamond,\oplus}$, and hence the formulation of reduction axioms for $\langle + \rangle$.

$+_p$	$\vdash \langle + \rangle\, p \leftrightarrow p$	$+_{\Diamond}$	$\vdash \langle + \rangle \Diamond \varphi \leftrightarrow \oplus \langle + \rangle \varphi$
$+_{\neg}$	$\vdash \langle + \rangle \neg\varphi \leftrightarrow \neg \langle + \rangle \varphi$	$+_{\oplus}$	$\vdash \langle + \rangle \oplus \varphi \leftrightarrow \oplus \langle + \rangle \varphi$
$+_{\vee}$	$\vdash \langle + \rangle (\varphi \vee \psi) \leftrightarrow (\langle + \rangle \varphi \vee \langle + \rangle \psi)$		
SE	If $\vdash \psi_1 \leftrightarrow \psi_2$ then $\vdash \varphi \leftrightarrow \varphi[\psi_2/\psi_1]$, with $\varphi[\psi_2/\psi_1]$ any formula obtained by replacing one or more occurrences of ψ_1 in φ with ψ_2.		

Table 4: Axioms and rule for the modality $\langle + \rangle$.

Theorem 4 (Axiom system for $\mathcal{L}_{\Diamond,\oplus,+}$) *The axioms and rules of Table 4, together with $\mathsf{L}_{\Diamond,\oplus}$ (Tables 1 and 2), form a sound and weakly complete axiom system for formulas of $\mathcal{L}_{\Diamond,\oplus,+}$ w.r.t. relational models.*

Proof. Soundness follows from the validity and validity-preserving properties of the reduction axioms (Table 4). We only show the validity of $+_{\Diamond}$ and $+_{\oplus}$. Take any relational state (M, w) with $M = \langle W, R, V \rangle$. For $+_{\Diamond}$, $(M, w) \Vdash \langle + \rangle \Diamond \varphi$ if and only if $(M^+, w) \Vdash \Diamond \varphi$, that is, if and only if there is $u \in W$ such that $R^+ wu$ and $(M^+, u) \Vdash \varphi$. But the latter is equivalent to $(M, u) \Vdash \langle + \rangle \varphi$, so there is $u \in W$ such that $R^+ wu$ and $(M, u) \Vdash \langle + \rangle \varphi$, i.e., $(M, w) \Vdash \oplus \langle + \rangle \varphi$. For $+_{\oplus}$, $(M, w) \Vdash \langle + \rangle \oplus \varphi$ if and only if $(M^+, w) \Vdash \oplus \varphi$, that is, if and only if there is $u \in W$ such that $(R^+)^+ wu$ and $(M^+, u) \Vdash \varphi$. But $(R^+)^+ = R^+$ and $(M^+, u) \Vdash \varphi$ if and only if $(M, u) \Vdash \langle + \rangle \varphi$, so there is $u \in W$ such that $R^+ wu$ and $(M, u) \Vdash \langle + \rangle \varphi$, i.e., $(M, w) \Vdash \oplus \langle + \rangle \varphi$.

Completeness follows from the completeness of $\mathsf{L}_{\Diamond,\oplus}$, as these extra axioms define a validity-preserving translation from $\mathcal{L}_{\Diamond,\oplus,+}$ to $\mathcal{L}_{\Diamond,\oplus}$, with the rule of substitution of logical equivalents *SE* taking care of formulas with more than one occurrence of $\langle + \rangle$ (one can work with the deepest occurrence of such modality and, once it is eliminated, proceed with the following one). Here is the formal definition of the translation function $t : \mathcal{L}_{\Diamond,\oplus,+} \to \mathcal{L}_{\Diamond,\oplus}$ induced by the axiomatisation.

$$
\begin{array}{rclcrcl}
t(p) &:=& p & \qquad & t(\langle + \rangle\, p) &:=& t(p) \\
t(\neg\varphi) &:=& \neg t(\varphi) & & t(\langle + \rangle\, \neg\varphi) &:=& t(\neg\, \langle + \rangle\, \varphi) \\
t(\varphi \vee \psi) &:=& t(\varphi) \vee t(\psi) & & t(\langle + \rangle\, (\varphi \vee \psi)) &:=& t(\langle + \rangle\, \varphi \vee \langle + \rangle\, \psi) \\
t(\Diamond\, \varphi) &:=& \Diamond\, t(\varphi) & & t(\langle + \rangle\, \Diamond\, \varphi) &:=& t(\diamondplus\, \langle + \rangle\, \varphi) \\
t(\diamondplus\, \varphi) &:=& \diamondplus\, t(\varphi) & & t(\langle + \rangle\, \diamondplus\, \varphi) &:=& t(\diamondplus\, \langle + \rangle\, \varphi) \\
 & & & & t(\langle + \rangle\, \langle + \rangle\, \varphi) &:=& t(\langle + \rangle\, t(\langle + \rangle\, \varphi)).
\end{array}
$$

The definitions on the right column will be applied, substituting recursively each subformula involving $\langle + \rangle$ with a logically equivalent one, reducing each time the complexity of the formula under the scope of the dynamic modality until such modality disappears (first case on right column). Thus, it is clear not only that the translation preserves validity (only validity preserving substitutions are applied), but also that every formula in $\mathcal{L}_{\Diamond,\diamondplus,+}$ is provably equivalent to its translation, as the formulas defining the translation belong to the axiom system.

Notice that since the basic system $\mathsf{L}_{PDL^{\triangleleft,?}}$ is weakly complete, we also get weak completeness for $\mathsf{L}_{\Diamond,\diamondplus}$. ■

Properties of the operation With the operation defined and its associated modality axiomatised, the important question is: what is the effect of this operation? Obviously, R^{+} is transitive. Then, after applying the operation, the agent has indeed full positive introspection about any formula φ.

Proposition 3.1 *Let φ be a formula in $\mathcal{L}_{\Diamond,+}$. Then,*

$$\Vdash [+]\,(\Box\,\varphi \rightarrow \Box\,\Box\,\varphi). \qquad ■$$

However, the operation does not take the agent from a state in which she knows a given φ without knowing she knows it, $\Box\,\varphi \wedge \neg\,\Box\,\Box\,\varphi$, to a state in which she knows φ and is positively introspective about it, $\Box\,\varphi \wedge \Box\,\Box\,\varphi$.

Fact 3.2 *The formula $\Box\,\varphi \rightarrow [+]\,(\Box\,\varphi \wedge \Box\,\Box\,\varphi)$ is not valid, not even when φ is propositional.*

Proof. In the relational state below on the left (each world containing the atoms true at it, and evaluation point double circled), the agent knows p without knowing that she knows it, $(M, w_1) \Vdash \Box\, p \wedge \neg\,\Box\,\Box\, p$.

M $\qquad\qquad$ M^{+}

Nevertheless, after the operation (relational state on the right) she does not know p anymore: $(M^{+}, w_1) \Vdash \neg\,\Box\, p$, that is, $(M, w_1) \Vdash \langle + \rangle\, \neg\,\Box\, p$. Hence, $(M, w_1) \Vdash \Box\, p \wedge \langle + \rangle\, \neg\,\Box\, p$: the formula $\Box\, p \rightarrow [+]\,(\Box\, p \wedge \Box\,\Box\, p)$ is not valid. ■

Discussion Fact 3.2 shows that the defined operation does not work as one might expect, and it also explains why. The operation makes the agent's knowledge positively introspective (Proposition 3.1), but instead of increasing her knowledge (she knows φ and, afterwards, she also knows that she knows it), it discards what was non-introspective (if she knows φ without knowing that she knows it, afterwards she does not know φ anymore).

Thus, this definition of a positive introspection action can be seen as having a 'pessimist' perspective. If an agent knows φ but still does not know that she knows it, the problem is not the lack of introspective reasoning: the problem is that she actually cannot guarantee (i.e., she does not know) that φ is the case. Though possibly counter-intuitive, this is actually what a relational-model representation of such situation amounts to: $\Box\varphi$ holds, so φ is the case in all worlds reachable from the evaluation point in one step. However, $\Box\Box\varphi$ fails, so there is at least one world reachable from the evaluation point in two steps where φ fails. Under this representation, a non-positively-introspective agent does not lack an introspective reasoning step; instead, she has to realise that "possible to be possible" implies "possible", what the operation actually does. Formally, the following formula is valid:

$$\Vdash \Diamond\Diamond\varphi \rightarrow \langle + \rangle \Diamond\varphi.$$

Going a step further, one might even wonder whether $\Box\varphi$ can be faithfully interpreted as "the agent knows φ" when the accessibility relation is not transitive, as the agent might still consider possible to consider $\neg\varphi$ possible (that is, there might be a $\neg\varphi$-world accessible from the evaluation point in two steps). In other words, maybe $\Box$ can be read as knowledge only when the accessibility relation is transitive, and thus $\Box\varphi$ guarantees that every world the agent 'eventually' consider possible satisfies φ. Readers agreeing with this idea might also consider, alternatively, not to assume transitivity, but rather to place the transitivity requirement inside the definition of the agent's knowledge. In other words, one might use $\boxplus\varphi$ (Definition 2.3) as the formula stating that the agent knows φ. Observe, in any case, how this discussion is different from the traditional one about positive introspection: here the question is not whether the knowledge of real agents has such property, but rather whether $\Box$ should be read as knowledge even when the modality's associated relation is not transitive.

3.2. General negative introspection

Just as in the positive introspection case, the first idea when looking for a model operation representing a negative introspection action is to make the accessibility relation Euclidean, thus making $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$ valid.

Definition 3.3 (General negative introspection operation) Let M be a relational model $\langle W, R, V\rangle$. The *general negative introspection* operation yields the model $M^{-} = \langle W, R^{E}, V\rangle$, with R^{E} the Euclidean closure of R, that is,

$$R^{E} := R \cup (\mathfrak{R} \circ (R \cup \mathfrak{R})^{*} \circ R). \qquad \blacktriangleleft$$

We can show that given any relation R, the relation R^E as defined above is indeed its Euclidean closure.

Lemma 3.1 *For any $R \subseteq (W \times W)$, the relation $R^E := R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^* \circ R)$ is R's Euclidean closure (the smallest Euclidean relation containing R).* ■

Proof. First, $R^E = R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^* \circ R)$ clearly contains R. Moreover, R^E is Euclidean, for suppose there are $w, u, v \in W$ such that $(w, u) \in R^E$ and $(w, v) \in R^E$. There are three possibilities, depending on the origin of such pairs: ***(i)*** if $(w, u) \in R$ and $(w, v) \in R$, then $(u, v) \in (\reflectbox{$R$} \circ R)$, and since $(\reflectbox{$R$} \circ R) \subseteq (\reflectbox{R} \circ (R \cup \reflectbox{R})^* \circ R)$ (just take 0 as the number of iterations of $(R \cup \reflectbox{$R$})^*$), then $(u, v) \in (\reflectbox{$R$} \circ (R \cup \reflectbox{R})^* \circ R)$ and hence $(u, v) \in R^E$; ***(ii)*** if $(w, u) \in R$ and $(w, v) \in (\reflectbox{$R$} \circ (R \cup \reflectbox{R})^\ell \circ R)$ for some $\ell \geq 0$, then $(u, v) \in (\reflectbox{$R$} \circ \reflectbox{$R$} \circ (R \cup \reflectbox{R})^\ell \circ R)$, that is, $(u, v) \in (\reflectbox{$R$} \circ (R \cup \reflectbox{R})^{\ell+1} \circ R)$ and hence $(u, v) \in R^E$; ***(iii)*** if both (w, u) and (w, v) are in $(R \circ (\reflectbox{$R$} \cup \reflectbox{$R$})^* \circ R)$, then so is (u, v), as it can be shown that the relation $(\reflectbox{$R$} \circ (R \cup \reflectbox{R})^* \circ R)$ is indeed Euclidean. Therefore, $(u, v) \in R^E$.

Finally, to prove that R^E is the smallest Euclidean relation containing R, it can be proved that any Euclidean relation $S \subseteq (W \times W)$ is such that $(\reflectbox{$S$} \circ (S \cup \reflectbox{S})^* \circ S) \subseteq S$. Consider now any Euclidean relation R' such that $R \subseteq R'$; it will be shown that $R^E \subseteq R'$, so take any $(w, u) \in R^E$. Then, either $(w, u) \in R$ or else $(w, u) \in (\reflectbox{$R$} \circ (R \cup \reflectbox{R})^* \circ R)$. In the first case, clearly $(w, u) \in R'$. For the second case, $R \subseteq R'$ implies $(\reflectbox{$R$} \circ (R \cup \reflectbox{R})^* \circ R) \subseteq (\reflectbox{R'} \circ (R' \cup \reflectbox{R'})^* \circ R')$, so $(w, u) \in (\reflectbox{$R'$} \circ (R' \cup \reflectbox{R'})^* \circ R')$; thus, by the just stated property and the fact that R' is Euclidean, $(w, u) \in R'$. Thus, $R^E \subseteq R'$. ■

Definition 3.4 The language $\mathcal{L}_{\Diamond,-}$ extends $\mathcal{L}_{\Diamond}$ with the modality $\langle - \rangle$. For its semantic interpretation, let (M, w) be a relational state. Then,

$$(M, w) \Vdash \langle - \rangle \varphi \quad \text{iff}_{def} \quad (M^-, w) \Vdash \varphi.$$

In words, the agent can perform an act of general negative introspection after which φ is the case, $(M, w) \Vdash \langle - \rangle \varphi$, if and only if, after the action, φ is the case, $(M^-, w) \Vdash \varphi$. The dual modality $[-]$ is defined as usual. ◀

Clearly, R^E can be equivalently defined in *PDL* plus the converse operator, thus suggesting that $\mathcal{L}_{PDL^\triangleleft}$ (Definition 2.4) will be useful to provide reduction axioms for this operation.

Axiom system It is not hard to see how $\mathcal{L}_{\Diamond}$ is not expressive enough to describe the effects of this operation (and thus there are no reduction axioms from $\mathcal{L}_{\Diamond,-}$ to $\mathcal{L}_{\Diamond}$): the definition of the Euclidean closure involves the use of relational operations that are not expressible with $\mathcal{L}_{\Diamond}$ (converse, iteration). However, the clearly more expressive $\mathcal{L}_{PDL^\triangleleft}$ allows us to deal with such operations, allowing also the formulation of reduction axioms from $\mathcal{L}_{PDL^\triangleleft,-}$ to $\mathcal{L}_{PDL^\triangleleft}$. Still, $\mathcal{L}_{PDL^\triangleleft}$ involves formulas of the form $\langle \alpha \rangle \varphi$, with the expression α being an arbitrary program expression; thus, an appropriate translation for each α must be provided. The *program transformer* defined below, a simplification of that in [33] for providing reduction axioms for regular *PDL*-expressions after action-model operations, captures this: it takes a program α describing a path in the new model M^{-}, returning its 'matching' path $\mathrm{T}(\alpha)$ in M.

Definition 3.5 (Program transformer T) The *program transformer* T, a function from program expressions to program expressions, is defined inductively in the following way.

$$
\begin{array}{ll}
\mathrm{T}(\triangleright) := \triangleright \cup (\triangleleft \,;\, (\triangleright \cup \triangleleft)^{*} \,;\, \triangleright) & \qquad \mathrm{T}(\alpha \cup \beta) := \mathrm{T}(\alpha) \cup \mathrm{T}(\beta) \\
\mathrm{T}(\triangleleft) := \triangleleft \cup (\triangleright \,;\, (\triangleleft \cup \triangleright)^{*} \,;\, \triangleleft) & \qquad \mathrm{T}(\alpha \,;\, \beta) := \mathrm{T}(\alpha) \,;\, \mathrm{T}(\beta) \\
 & \qquad \mathrm{T}(\alpha^{*}) := (T(\alpha))^{*}.
\end{array}
$$

The crucial property of this program transformer is the following one.

Proposition 3.2 *Let $M = \langle W, R, V\rangle$ be a relational model; take $M^{-} = \langle W, R^{E}, V\rangle$. For every program expression α, the relation described by α in the resulting M^{-} is the same as the relation described by $\mathrm{T}(\alpha)$ in the original M:*

$$(R^{E})_{\alpha} = R_{\mathrm{T}(\alpha)}. \qquad \blacksquare$$

Proof. The proof is by structural induction on α. For the base case $(R^{E})_{\triangleright}$ (the case $(R^{E})_{\triangleleft}$ is similar):

$$
\begin{aligned}
(R^{E})_{\triangleright} \;=\; R^{E} \;=\; R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^{*} \circ R) \;&=\; R_{\triangleright} \cup (R_{\triangleleft} \circ (R_{\triangleright} \cup R_{\triangleleft})^{*} \circ R_{\triangleright}) \\
&=\; R_{\triangleright} \cup (R_{\triangleleft} \circ (R_{\triangleright \cup \triangleleft})^{*} \circ R_{\triangleright}) \\
&=\; R_{\triangleright} \cup (R_{\triangleleft} \circ R_{(\triangleright \cup \triangleleft)^{*}} \circ R_{\triangleright}) \\
&=\; R_{\triangleright} \cup R_{\triangleleft ; (\triangleright \cup \triangleleft)^{*} ; \triangleright} \\
&=\; R_{\triangleright \cup (\triangleleft ; (\triangleright \cup \triangleleft)^{*} ; \triangleright)} \\
&=\; R_{\mathrm{T}(\triangleright)}
\end{aligned}
$$

For the inductive cases (with $(R^{E})_{\alpha} = R_{\mathrm{T}(\alpha)}$ and $(R^{E})_{\beta} = R_{\mathrm{T}(\beta)}$ the inductive hypotheses),

$$(R^{E})_{\alpha \cup \beta} \;=\; (R^{E})_{\alpha} \cup (R^{E})_{\beta} \;=\; R_{\mathrm{T}(\alpha)} \cup R_{\mathrm{T}(\beta)} \;=\; R_{\mathrm{T}(\alpha) \cup \mathrm{T}(\beta)} \;=\; R_{\mathrm{T}(\alpha \cup \beta)}$$

$$(R^{E})_{\alpha ; \beta} \;=\; (R^{E})_{\alpha} \circ (R^{E})_{\beta} \;=\; R_{\mathrm{T}(\alpha)} \circ R_{\mathrm{T}(\beta)} \;=\; R_{\mathrm{T}(\alpha) ; \mathrm{T}(\beta)} \;=\; R_{\mathrm{T}(\alpha ; \beta)}$$

$$(R^{E})_{\alpha^{*}} \;=\; ((R^{E})_{\alpha})^{*} \;=\; (R_{\mathrm{T}(\alpha)})^{*} \;=\; R_{(\mathrm{T}(\alpha))^{*}} \;=\; R_{\mathrm{T}(\alpha^{*})}$$

Then, it is possible to introduce the reduction axioms.

Theorem 5 (Axiom system for $\mathcal{L}_{PDL^{\triangleleft},-}$) *The axioms and rules of Table 5, together with the axiom system $\mathsf{L}_{PDL^{\triangleleft}}$ (Table 3), form a sound and weakly complete axiom system for formulas of $\mathcal{L}_{PDL^{\triangleleft},-}$ w.r.t. relational models.*

$-_p$	$\vdash \langle - \rangle p \leftrightarrow p$	$-_{\langle\alpha\rangle}$	$\vdash \langle - \rangle \langle \alpha \rangle \varphi \leftrightarrow \langle \mathrm{T}(\alpha) \rangle \langle - \rangle \varphi$
$-_\neg$	$\vdash \langle - \rangle \neg\varphi \leftrightarrow \neg \langle - \rangle \varphi$		
$-_\vee$	$\vdash \langle - \rangle (\varphi \vee \psi) \leftrightarrow (\langle - \rangle \varphi \vee \langle - \rangle \psi)$		
SE	If $\vdash \psi_1 \leftrightarrow \psi_2$ then $\vdash \varphi \leftrightarrow \varphi[\psi_2/\psi_1]$, with $\varphi[\psi_2/\psi_1]$ any formula obtained by replacing one or more occurrences of ψ_1 in φ with ψ_2.		

Table 5: Axioms and rule for the modality $\langle - \rangle$.

Proof. (Sketch) With respect to soundness, here only the validity of $-_{\langle\alpha\rangle}$ is discussed. Take any relational state (M, w) with $M = \langle W, R, V \rangle$. Then, $(M, w) \Vdash \langle - \rangle \langle \alpha \rangle \varphi$ if and only if $(M^-, w) \Vdash \langle \alpha \rangle \varphi$, that is, if and only if there is $u \in W$ such that $(R^E)_\alpha wu$ and $(M^-, u) \Vdash \varphi$. But while the former is, by Proposition 3.2, equivalent to $R_{\mathrm{T}(\alpha)} wu$, the latter is, by definition, equivalent to $(M, u) \Vdash \langle - \rangle \varphi$; hence, $(M, w) \Vdash \langle \mathrm{T}(\alpha) \rangle \langle - \rangle \varphi$.

With respect to completeness, the argument runs as that of Theorem 4. ■

Properties of the operation As expected, after the operation, the agent has negative introspection.

Proposition 3.3 *Let φ be a formula in $\mathcal{L}_{\Diamond,-}$. Then,*

$$\Vdash [-] (\neg \Box \varphi \rightarrow \Box \neg \Box \varphi). \qquad ■$$

This is a good initial step but, as the general positive introspection case showed, it is not enough. Does the operation take a state where $\neg \Box \varphi \wedge \neg \Box \neg \Box \varphi$ holds, to a state in which $\neg \Box \varphi \wedge \Box \neg \Box \varphi$ holds? Different from the analogous question in the positive introspection case (Fact 3.2), when the involved formula is purely propositional, the answer here is yes.

Proposition 3.4 *Let γ be a propositional formula. Then,*

$$\Vdash \neg \Box \gamma \rightarrow [-] (\neg \Box \gamma \wedge \Box \neg \Box \gamma). \qquad ■$$

Proof. Let (M, w) be a relational state with $M = \langle W, R, V \rangle$, and suppose $(M, w) \Vdash \neg \Box \gamma$; then there is $u \in W$ such that Rwu and $(M, u) \Vdash \neg\gamma$, so $R^E wu$ (as, by definition, $R \subseteq R^E$) and $(M^-, u) \Vdash \neg\gamma$ (as γ is propositional). Thus, first, $(M^-, w) \Vdash \Diamond \neg\gamma$, i.e., $(M^-, w) \Vdash \neg \Box \gamma$. Second, for every $u' \in W$, $R^E wu'$ implies $R^E u'u$ (R^E is Euclidean), and hence $(M^-, u') \Vdash \Diamond \neg\gamma$ so $(M^-, w) \Vdash \Box \Diamond \neg\gamma$, i.e., $(M^-, w) \Vdash \Box \neg \Box \gamma$. Thus, $(M, w) \Vdash [-] (\neg \Box \gamma \wedge \Box \neg \Box \gamma)$. ■

In fact, it is not hard to see that the property stated in the previous proposition holds not only for propositional formulas, but also for those whose *falsity* is preserved by the operation

Proposition 3.5 *Let φ be a be a formula in $\mathcal{L}_{\Diamond,-}$ such that $\Vdash \neg\varphi \to [-]\neg\varphi$. Then,*

$$\Vdash \neg \Box \varphi \to [-](\neg \Box \varphi \wedge \Box \neg \Box \varphi). \qquad \blacksquare$$

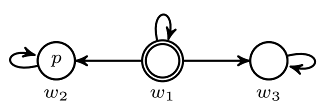
From this validity, it follows that

$$\Vdash \neg \Box \varphi \to [-] \Box \neg \Box \varphi. \tag{1}$$

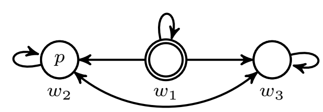
This is nothing but at *dynamic* version of the negative introspection axiom $\neg \Box \varphi \to \Box \neg \Box \varphi$ for formulas φ with particular requirements. Still, the validity does not hold for arbitrary formulas in $\mathcal{L}_{\Diamond,-}$.

Fact 3.3 *The formula $\neg \Box \varphi \to [-] \Box \neg \Box \varphi$ is not valid.*

Proof. Consider $\varphi := \neg \Box p$ and the relational state (M, w_1) shown below on the left. Note how $(M, w_1) \Vdash \neg \Box(\neg \Box p)$, that is, $(M, w_1) \Vdash \Diamond \Box p$.



M



M^-

After the operation (relational state on the right), $(M^-, w_1) \Vdash \Diamond \Box \Diamond \neg p$, that is, $(M^-, w_1) \Vdash \neg \Box \neg \Box(\neg \Box p)$ so $(M, w_1) \Vdash \langle - \rangle \neg \Box \neg \Box(\neg \Box p)$. $\blacksquare$

Thus, the property *"if the agent does not know φ, then after the operation she knows she does not know it"* ($\neg \Box \varphi \to [-] \Box \neg \Box \varphi$) does not hold for all φ. Still, note how the formula φ used as a counterexample, $\neg \Box p$, expresses that *"the agent does not know p"*. The reason why the mentioned property does not work for all φ is because φ itself might be about the agent's lack of knowledge about propositional ($\neg \Box p$) formulas and thus, according to Proposition 3.4, the agent will know φ after the operation ($[-] \Box(\neg \Box p)$). Hence, if the agent's knowledge is truthful (i.e., if the relation is reflexive, as in the given counterexample), the agent's lack of knowledge of φ (the initial $\neg \Box(\neg \Box p)$) will be gone ($\neg \langle - \rangle \neg \Box(\neg \Box p)$), and therefore the agent cannot know it ($[-] \neg \Box \neg \Box(\neg \Box p)$).

So, what is the knowledge the agent has after this general negative introspection operation? Here, the provided reduction axioms are useful to get an answer, as the effect of the general negative introspection operation can be fully described within the language. It should be noted, however, that some of the tools needed for the axiomatisation (as, e.g., the converse relation $\mathfrak{R}$) might not have a natural epistemic interpretation, and thus formulas in $\mathcal{L}_{PDL^{\triangleleft},-}$ might not have a direct epistemic reading.

Here is the characterisation of the knowledge of the agent after the general negative introspection operation (observe that, in this extended language $\mathcal{L}_{PDL^{\triangleleft},-}$, the modality $\Box \varphi$ corresponds to $[\triangleright] \varphi$).

Proposition 3.6 *Let φ be a formula in $\mathcal{L}_{PDL^{\triangleleft},-}$. The agent can perform a general negative introspection step after which she will know φ ($\langle - \rangle [\triangleright] \varphi$) if and only if she knows that after the operation φ will be the case ($[\triangleright][-]\varphi$) and also she 'conversely knows'* [10] *that it is 'common among her knowledge and her converse knowledge'* [11] *that she knows that after the operation φ will be the case ($[\triangleright][-]\varphi$). More precisely,*

$$\Vdash \langle - \rangle [\triangleright] \varphi \leftrightarrow \left([\triangleright][-]\varphi \wedge [\triangleleft][(\triangleright \cup \triangleleft)^*][\triangleright][-]\varphi \right). \qquad \blacksquare$$

Proof. The validity can be proved either semantically or else syntactically. The latter is simpler, as a derivation from $\langle - \rangle [\triangleright] \varphi$ leads, by a successive application of the definition of $[\alpha]$, $-_{\neg}$, $-_{\langle \alpha \rangle}$, $-_{\neg}$ (Table 5) and the definition of $[\alpha]$, again, to $[\mathrm{T}(\triangleright)] \langle - \rangle \varphi$. Then, after substituting $\mathrm{T}(\triangleright)$ by "$\triangleleft \cup (\triangleright ; (\triangleleft \cup \triangleright)^* ; \triangleleft$", a successive application of the 'box' versions of the axioms $\cup$ and $;$ (Table 3) lead to the desired formula $[\triangleright][-]\varphi \wedge [\triangleleft][(\triangleright \cup \triangleleft)^*][\triangleright][-]\varphi$. $\blacksquare$

Naturally, the axiom system (in particular, the reduction axioms) can be used to find other interesting validities. For example, a simple substitution of φ for $\neg [\triangleright] \varphi$ in the result of the previous proposition produces a validity characterising the knowledge of the lack of knowledge of the given φ (i.e., characterising those formulas φ the agent will be negatively introspective about) after the operation:

$$\Vdash \langle - \rangle [\triangleright] \neg [\triangleright] \varphi \leftrightarrow \left([\triangleright][-] \neg [\triangleright] \varphi \wedge [\triangleleft][(\triangleright \cup \triangleleft)^*][\triangleright][-] \neg [\triangleright] \varphi \right).$$

This validity can be refined by a further use of recursion axioms.

Discussion The obvious difference between these 'natural' attempts to define a positive and a negative introspection operation is that, while the positive introspection case does not behave as expected, the one for negative introspection does. This is worth of notice because both operations follow the same strategy: give the accessibility relation the property that makes the correspondent introspection property valid. Nevertheless, both operations work by *adding* edges (Definitions 3.1 and 3.3), and this might not be what is needed in both cases. Indeed, while positive introspection attempts to reach knowledge of knowledge of φ, negative introspection attempts to achieve knowledge of *lack of* knowledge of φ. Thus, in the case of propositional formulas γ, while adding edges might fail for positive introspection because the original "knowledge of γ" (all R-reachable worlds are γ-worlds) may be lost (a $\neg\gamma$-world might become R-reachable), adding edges works for the second because the original "lack of knowledge of γ" (there is an R-reachable $\neg\gamma$-world) will not be lost (such world will still be R-reachable). Moreover, "knowledge of lack of knowledge of γ" will

[10] A very informal reading of $[\triangleleft]$.

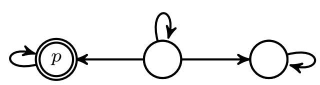
[11] A reading for $[(\triangleright \cup \triangleleft)^*]$, analogous to that of $[(i \cup j)^*]$ as common knowledge among i and j. See [33] for an epistemic reading of *PDL* operators.

be reached (the relation becomes Euclidean, so all R-reachable worlds can see the 'problematic' $\neg\gamma$-world).

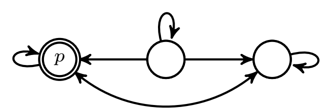
Still, this negative introspection operation also has unintended side-effects: while it makes the agent negatively introspective on her lack of propositional knowledge (Proposition 3.4), it might also cause propositional knowledge disappear.

Fact 3.4 *The formula* $\Box\, p \rightarrow [-]\, \Box\, p$ *is not valid.*

Proof. In the relational state below on the left, $\Box\, p$ holds; however, this is not the case in the relational state on the right, the one resulting from a negative introspection operation.



M



M^{-}

This emphasises that, while the negative introspection action allows the agent to reach an introspective state, it might not correspond to our intuitive understanding of what an introspection act is.

4. Particular introspection with respect to χ

Section 3 introduced operations to achieve *full* positive and negative introspection about *all* formulas (so called *general* introspection operations). By making the accessibility relation transitive (resp., Euclidean), such strategy produces agents whose knowledge has full positive (resp., negative) introspection. Still, although this general negative introspection operation behaves as expected with respect to (lack of) knowledge about propositional formulas (Proposition 3.5; but recall how the agent might lose knowledge too, Fact 3.4), the same is not the case for the general positive introspection operation (Fact 3.2)

This section explores another alternative for introspection operations. Instead of focussing on achieving full introspection for all formulas, it focuses on achieving full introspection with respect to a particular one.

4.1. Particular positive introspection

The operation of Definition 3.1 allows the agent to have full positive introspection at the cost of losing the 'non-introspective' knowledge. This does not follow the intuition of what an actual positive introspection reasoning step should do. Such a reasoning step would rather take the agent from knowing χ without knowing she knows it, to knowing χ and knowing she knows it. But then the operation representing such epistemic action within relational models should be radically different. When (M, w) is a relational state in which the agent knows a given χ without having full positive introspection about it, every world R-reachable from w *in one step* satisfies χ, but there is at least one world R-reachable from w in two or more steps where χ fails. Then, to have full positive introspection about χ, such $\neg\chi$-worlds should not be R-reachable anymore in any number of steps. In other words, the operation should remove edges.

Definition 4.1 (U-disconnecting operation) Let M be a relational model $\langle W, R, V\rangle$, with $U \subseteq W$ a subset of its domain. The U*-disconnecting* operation yields the model $M_{+U} = \langle W, R', V\rangle$, with R' given by

$$R' := R \setminus (U \times \overline{U}),$$

for $\overline{U} := W \setminus U$. In words, the operation drops edges from worlds in U to worlds not in U. ◀

When the parameter of this model operation, the set U, is given by the truth-set of a formula, say χ, then the U-disconnecting operation can be understood as a *particular positive* χ*-introspection operation*, as it removes edges from worlds satisfying χ to worlds not satisfying χ. In such case, the relation of the resulting model can be equivalently defined, using *PDL* notation, as $R' := (?\neg\chi \,;\, R) \cup (R \,;\, ?\chi)$. This states, then, that there will be an R'-edge from w to u if and only if there is already an R-edge from w to u and either w is a $\neg\chi$-world or else u is a χ-world.

The modality for this positive χ-introspection operation will be introduced in two stages, the first one being the definition of an auxiliary modality.

Definition 4.2 The language $\mathcal{L}_{\Diamond,+'\chi}$ extends $\mathcal{L}_{\Diamond}$ with a modality $\langle +'\chi\rangle$ for each formula χ. For the semantic interpretation, let (M, w) be a relational state with $M = \langle W, R, V\rangle$ and recall that, for any formula which can be interpreted in relational models, $[\![\chi]\!]^M$ is the set of worlds of the model M in which the formula holds. Then,

$$(M, w) \Vdash \langle +'\chi\rangle\, \varphi \quad \text{iff}_{def} \quad (M_{+[\![\chi]\!]^M}, w) \Vdash \varphi.$$

Note how there is no circularity: the truth-value of $\langle +'\chi\rangle\,\varphi$ on (M, w) depends on both the set of worlds in M in which χ is the case and the truth-value of φ on $(M_{+[\![\chi]\!]^M}, w)$, but both cases deal with a strict subformula of the original $\langle +'\chi\rangle\,\varphi$. Note also how, similar to the previous cases, the operation is deterministic and the truth condition for its modality does not have a precondition; hence the modality $[+']$, defined as $[+'\chi]\,\varphi := \neg\,\langle +'\chi\rangle\,\neg\varphi$, is equivalent to $\langle +'\rangle$.

This auxiliary modality differs from what one might expect in one crucial way: its semantic interpretation has no precondition, thus indicating that the epistemic action it represents, an introspective reasoning step for χ, can take place in any situation (even in those in which the agent does not know χ). This issue can be solved in a second stage by introducing another modality with the corresponding precondition:

$$\langle +\chi\rangle\,\varphi := \Box\,\chi \wedge \langle +'\chi\rangle\,\varphi.$$

The reader familiar with *DEL* might notice here a departure from the traditional approach: why using an auxiliary 'preconditionless' modality? The reason is, as it will be discussed below, that the former simplifies the formulation of reduction axioms in order to provide a sound and complete axiom system.

Axiom system The first step for providing an axiom system for the modality $\langle +\chi \rangle$ is to provide reduction axioms for its 'preconditionless' counterpart $\langle +'\chi \rangle$.

$+'\chi_p$	$\vdash \langle +'\chi \rangle\, p \;\leftrightarrow\; p$
$+'\chi_\neg$	$\vdash \langle +'\chi \rangle \neg\varphi \;\leftrightarrow\; \neg \langle +'\chi \rangle\, \varphi$
$+'\chi_\vee$	$\vdash \langle +'\chi \rangle (\varphi \vee \psi) \;\leftrightarrow\; \langle +'\chi \rangle\, \varphi \vee \langle +'\chi \rangle\, \psi$
$+'\chi_\Diamond$	$\vdash \langle +'\chi \rangle \Diamond\, \varphi \;\leftrightarrow\; (\neg\chi \wedge \Diamond \langle +'\chi \rangle\, \varphi) \vee \Diamond(\chi \wedge \langle +'\chi \rangle\, \varphi)$
SE	If $\vdash \psi_1 \leftrightarrow \psi_2$ then $\vdash \varphi \leftrightarrow \varphi[\psi_2/\psi_1]$, with $\varphi[\psi_2/\psi_1]$ any formula obtained by replacing one or more *non-modality* occurrences of ψ_1 in φ with ψ_2, with the non-modality occurrences of ψ_1 being those which are *not* within the brackets of the 'dynamic' modality $\langle +' \;\rangle$.[12]

Table 6: Axioms and rule for the modality $+'\chi$.

Theorem 6 (Axiom system for $\mathcal{L}_{\Diamond,+'\chi}$) *The axioms and rules of Table 6, together with the axiom system* $\mathsf{L}_\Diamond$ *(Table 1), form a sound and strongly complete axiom system for formulas of* $\mathcal{L}_{\Diamond,+'\chi}$ *w.r.t. relational models.* ■

Theorem 6 provides a sound and strongly complete axiom system for $\langle +'\chi \rangle$. For axiomatising $\langle +\chi \rangle$, recall that such modality is just a straightforward abbreviation; thus, the clearly valid $\langle +\chi \rangle\, \varphi \;\leftrightarrow\; (\Box\, \chi \wedge \langle +'\chi \rangle\, \varphi)$ is enough.

Now, for the reason of using auxiliary 'preconditionless' modalities. Suppose $\langle +\chi \rangle$ were defined directly with its precondition,

$$(M, w) \Vdash \langle +\chi \rangle\, \varphi \quad \text{iff} \quad (M, w) \Vdash \Box\, \chi \text{ and } (M_{+[\![\chi]\!]^M}, w) \Vdash \varphi.$$

How its reduction axiom for $\Diamond$ would look like? A first attempt would be to simply 'plug' the precondition in axiom $+'\chi_\Diamond$, that is,

$$\langle +\chi \rangle \Diamond\, \varphi \;\leftrightarrow\; \Box\, \chi \wedge ((\neg\chi \wedge \Diamond \langle +\chi \rangle\, \varphi) \vee \Diamond(\chi \wedge \langle +\chi \rangle\, \varphi)).$$

But such formula is not valid: in the left-to-right direction, one needs to show that there is a R-reachable world, say u, satisfying $\langle +\chi \rangle\, \varphi$ (the subformulas $\Diamond \langle +\chi \rangle\, \varphi$ and $\Diamond(\_\_ \wedge \langle +\chi \rangle\, \varphi)$). This amounts for u not only to satisfy φ after

[12] The side condition of the *SE* rule is not really needed; still, it provides exactly what is needed to make the system complete. This side condition makes our *SE* analogous to the rule *RE* used in [53] (table on page 106) for axiomatising a public announcement modality.

the operation, $(M_{+\llbracket\chi\rrbracket^M}, u) \Vdash \varphi$, but also to satisfy the modality's precondition, $(M, u) \Vdash \Box\,\chi$. But while the first part is indeed the case, the second cannot be guaranteed, as the relation R does not need to satisfy any particular property and hence the operation's precondition cannot be 'pushed' through R-edges. The best one can do is to ask for such u to satisfy φ *conditionally* after the operation, i.e., to satisfy not $\langle +\chi\rangle\,\varphi$ but rather $[+\chi]\,\varphi$, thus yielding the formula

$$\langle +\chi\rangle \Diamond \varphi \;\leftrightarrow\; \Box\,\chi \wedge ((\neg\chi \wedge \Diamond\,[+\chi]\,\varphi) \vee \Diamond(\chi \wedge [+\chi]\,\varphi)).$$

But then it is now the right-to-left direction which fails, as the right-hand side does not guarantee that indeed there is a R'-reachable world satisfying φ.

On the other hand, the 'preconditionless' modality allows a simple reduction axiom: a simple substitution using axiom $+'\chi_{\Diamond}$ and $\langle +\chi\rangle$'s definition yields

$$\Vdash \langle +\chi\rangle \Diamond \varphi \;\leftrightarrow\; \Box\,\chi \wedge ((\neg\chi \wedge \Diamond\,\langle +'\chi\rangle\,\varphi) \vee \Diamond(\chi \wedge \langle +'\chi\rangle\,\varphi)).$$

This validity is very similar to the first attempt above, but differs from it in that, on the right-hand side, the 'dynamic' modalities are the 'preconditionless' ones. It also makes clear the role of the precondition: it is required for the *epistemic action* to take place, but not for the *operation* to be applied.

From a general perspective, the reason for the indirect definition of the required modality is the 'mismatch' between the action's precondition ($\Box\,\chi$) and what worlds need to satisfy in the original pointed model in order to be reachable in the new one (χ). In the well-known case of public announcement logic (PAL; [54, 55]), these two are the same: the precondition for the action is for the announced formula to be the case (χ), and a world needs to satisfy such formula (χ) in order to be R'-reachable. However, in the case of positive χ-introspection they are different, with none of them implying the other without further requirements (e.g., reflexivity, so satisfying the precondition implies being reachable in the new model). Hence the need of an auxiliary 'preconditionless' modality.

Properties of the operation It is now time to explore the operation's behaviour. First, here it is a validity characterizing the knowledge of the agent after the operation.

Proposition 4.1 *Let χ and φ be formulas in $\mathcal{L}_{\Diamond,+'\chi}$. Then,*

$$\Vdash \langle +\chi\rangle\,\Box\,\varphi \;\leftrightarrow\; \Box(\chi \wedge [+'\chi]\,\varphi). \qquad \blacksquare$$

Proof. Let (M, w) be a relational state, $M = \langle W, R, V\rangle$. From left to right, $(M, w) \Vdash \langle +\chi\rangle\,\Box\,\varphi$ is, by definition, $(M, w) \Vdash \Box\,\chi$ and $(M, w) \Vdash \langle +'\chi\rangle\,\Box\,\varphi$. From the first, Rwu implies $(M, u) \Vdash \chi$; from the latter, $(M_{+\llbracket\chi\rrbracket^M}, w) \Vdash \Box\,\varphi$, that is, $R'wu$ implies $(M_{+\llbracket\chi\rrbracket^M}, u) \Vdash \varphi$. Consider now any $u \in W$ such that Rwu: then, $(M, u) \Vdash \chi$ and hence, from the definition of the accessibility relation R' in $M_{+\llbracket\chi\rrbracket^M}$, it follows that $R'wu$, so $(M_{+\llbracket\chi\rrbracket^M}, u) \Vdash \varphi$ which is, by definition, $(M, u) \Vdash [+'\chi]\,\varphi$. Thus, Rwu implies $(M, u) \Vdash \chi \wedge [+'\chi]\,\varphi$, and therefore $(M, w) \Vdash \Box(\chi \wedge [+'\chi]\,\varphi)$.

From right to left, $(M, w) \Vdash \Box(\chi \wedge [+'\chi]\,\varphi)$ implies, first, $(M, w) \Vdash \Box\,\chi$, and second, $(M, w) \Vdash \Box\,[+'\chi]\,\varphi$, with the latter stating that Rwu implies $(M, u) \Vdash [+'\chi]\,\varphi$. Take now any $u \in W$ such that $R'wu$: since $R' \subseteq R$, then Rwu so $(M, u) \Vdash [+'\chi]\,\varphi$, i.e., $(M_{+[\![\chi]\!]^M}, u) \Vdash \varphi$. Thus, $(M_{+[\![\chi]\!]^M}, w) \Vdash \Box\,\varphi$, and therefore $(M, w) \Vdash \langle +'\chi \rangle\,\Box\,\varphi$. But recall the first: $(M, w) \Vdash \Box\,\chi$. Hence, $(M, w) \Vdash \Box\,\chi \wedge \langle +'\chi \rangle\,\Box\,\varphi$ and thus, by definition, $(M, w) \Vdash \langle +\chi \rangle\,\Box\,\varphi$. ■

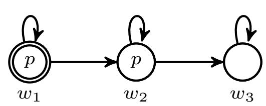
In order to show how the operation behaves as expected, consider the instance of the previous validity with φ replaced by $\Box\,\chi$:

$$\Vdash \langle +\chi \rangle\,\Box\,\Box\,\chi \leftrightarrow \Box(\chi \wedge [+'\chi]\,\Box\,\chi).$$

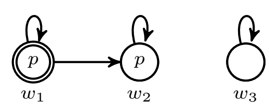
The formula states what is needed for the agent to have a one-level positive introspection about χ ($\Box\,\Box\,\chi$) after the operation: the agent should know *"χ and, after the 'preconditionless' operation, she will know χ"* ($\Box(\chi \wedge [+'\chi]\,\Box\,\chi)$). One might expect for the second conjunct inside the scope of $\Box$ in the right-side, $[+'\chi]\,\Box\,\chi$ (*"after the 'preconditionless' operation, the agent knows χ"*), to collapse to $\top$, so that the necessary and sufficient condition for the agent to reach one-level positive introspection about χ after the action is for her to simply know χ. Nevertheless, this is not the case.

Fact 4.1 *The formula $\Box\,\chi \to [+'\chi]\,\Box\,\chi$ is not valid (knowing χ does not guarantee the agent will still know it after the 'preconditionless' χ-introspection action), and thus neither is $[+'\chi]\,\Box\,\chi$.*

Proof. Take $\chi := p \wedge \Diamond\,\Diamond\,\neg p$ and the relational state shown below to the left. The formula χ holds at w_1 and w_2, so $(M, w_1) \Vdash \Box\,\chi$.



M



$M_{+[\![\chi]\!]^M}$

Since χ fails at w_3 in M, the operation yields the relational state to the right, with χ false at w_2; then, $(M_{+[\![\chi]\!]^M}, w_1) \Vdash \neg\,\Box\,\chi$ and hence $(M, w_1) \Vdash \Box\,\chi \wedge \langle +'\chi \rangle\,\neg\,\Box\,\chi$, i.e., $(M, w_1) \Vdash \Box\,\chi \wedge \neg\,[+'\chi]\,\Box\,\chi$: the agent knows χ, but she does not know it anymore after a 'preconditionless' positive χ-introspection action. Note how $(M, w_1) \Vdash \neg\,\Box\,\Box\,\chi$, so the introspection action is not redundant. Even more, (M, w_1) satisfies $\Box\,\chi$, the precondition of $\langle +\chi \rangle$; hence, it satisfies $\langle +\chi \rangle\,\neg\,\Box\,\chi$, that is, $\neg\,[+\chi]\,\Box\,\chi$: neither $\chi \to [+\chi]\,\Box\,\chi$ is valid. ■

Fact 4.1 is just one more instance of the well-known Moorean phenomenon in *DEL*. In its best-known incarnation, within *PAL*, it appears as formulas that become false after its truthful public announcement [56, 57], with the paradigmatic example being $p \wedge \neg\,\Box\,p$. Here, it occurs as formulas that are known but, after a particular positive introspection action, are *not* known anymore. Still, although the operation does not behave as expected in the general case, it does so when restricted to formulas whose truth is preserved by the operation.

Proposition 4.2 *Let χ be a formula in $\mathcal{L}_{\Diamond,+'\chi}$ such that $\Vdash \chi \to [+'\chi]\,\chi$. Then,*

$$\Vdash \langle +\chi\rangle \,\Box\,\Box\,\chi \;\leftrightarrow\; \Box\,\chi. \qquad \blacksquare$$

Proof. The direction from left to right is an immediate consequence of the validity that results from replacing φ with $\Box\,\chi$ in the formula of Proposition 4.1 (such validity appears immediately after the mentioned proposition, on page 21). From right to left, take any relational state (M,w) with $M = \langle W,R,V\rangle$, and suppose $(M,w) \Vdash \Box\,\chi$; then Rwu implies $(M,u) \Vdash \chi$. Now consider any $u \in W$ such that $R'wu$ and any $v \in W$ such that $R'uv$. Since $R' \subseteq R$, then Rwu and hence $(M,u) \Vdash \chi$. But $R'uv$ so, from the definition of R', it follows that $(M,v) \Vdash \chi$. Then, by the additional assumption about the operation's effect on χ's truth-value, $(M,v) \Vdash [+'\chi]\,\chi$, that is, $(M_{+[\![\chi]\!]^M}, v) \Vdash \chi$. Since v is an arbitrary R'-successor of u, $(M_{+[\![\chi]\!]^M}, u) \Vdash \Box\,\chi$; since u is an arbitrary R'-successor of w, $(M_{+[\![\chi]\!]^M}, w) \Vdash \Box\,\Box\,\chi$. Hence, $(M,w) \Vdash \langle +'\chi\rangle\,\Box\,\Box\,\chi$ and, since the precondition holds, $(M,w) \Vdash \langle +\chi\rangle\,\Box\,\Box\,\chi$. $\blacksquare$

The right-to-left direction of this validity,

$$\Vdash \Box\,\chi \to \langle +\chi\rangle\,\Box\,\Box\,\chi \tag{2}$$

is the *dynamic* version of the positive introspection axiom $\Box\,\varphi \to \Box\,\Box\,\varphi$, here with the action affecting a specific formula whose truth is preserved by the operation. Indeed, the agent does not need to have (a one-step) positive introspection for such a χ (the indistinguishability relation does not need to be transitive), but she can act on it. In fact, the operation does more: for such formulas χ, the agent will have *full* positive introspection after the operation.

Proposition 4.3 *Let χ be a formula in $\mathcal{L}_{\Diamond,+'\chi}$ such that $\Vdash \chi \to [+'\chi]\,\chi$. Then,*

$$\Vdash \Box\,\chi \to \langle +\chi\rangle\,\Box^n\,\Box\,\chi,$$

for any $n \geq 0$, with $\Box^0\,\varphi := \varphi$ and $\Box^{k+1}\,\varphi := \Box^k\,\Box\,\varphi$. $\blacksquare$

Proof. Let (M,w) be a relational state, $M = \langle W,R,V\rangle$, and suppose $(M,w) \Vdash \Box\,\chi$; then Rwu implies $(M,u) \Vdash \chi$.

The first step is to show, by induction on $n \geq 0$, how $(R')^{n+1}wu$ implies $(M,u) \Vdash \chi$. The base case is immediate: $(R')^1wu$ is $R'wu$, and since $R' \subseteq R$, then Rwu and thus $(M,u) \Vdash \chi$. For the inductive case, suppose $(R')^{n+2}wu$. Then there is $v \in W$ such that $(R')^{n+1}wv$ and $R'vu$, and hence $(M,v) \Vdash \chi$ (from the first and inductive hypothesis) and Rvu (from the second and $R' \subseteq R$). But $R'vu$ so, from the definition of R', it follows that $(M,u) \Vdash \chi$.

Now, to prove the required $(M,w) \Vdash \langle +\chi\rangle\,\Box^n\,\Box\,\chi$, take any $n \geq 0$ and any $u \in W$ such that $(R')^{n+1}wu$. Then $(M,u) \Vdash \chi$ and therefore, by the assumption, $(M,u) \Vdash [+'\chi]\,\chi$, that is, $(M_{+[\![\chi]\!]^M}, u) \Vdash \chi$. Thus, $(R')^{n+1}wu$ implies $(M_{+[\![\chi]\!]^M}, u) \Vdash \chi$, that is, $(M_{+[\![\chi]\!]^M}, w) \Vdash \Box^n\,\Box\,\chi$ so $(M,w) \Vdash \langle +'\chi\rangle\,\Box^n\,\Box\,\chi$. But $\langle +\chi\rangle$'s precondition holds; thus, $(M,w) \Vdash \langle +\chi\rangle\,\Box^n\,\Box\,\chi$, as required. $\blacksquare$

Discussion The results above show how, if the operation does not affect χ's truth, the only requirement for the agent to get full positive χ-introspection (while preserving knowledge!) after the operation is the action's precondition: to know χ. This operation gives then a more 'optimistic' representation of non-positively-introspective situations: the agent knows χ without noticing it, and thus she only needs to make a further 'introspective' effort to realise it, therefore making $\Box\Box\chi$ true. This is the interpretation that comes to mind when one talks about such situations 'in real life'. Indeed, the operation does not produce full positive introspection for all formulas, but it guarantees that, after the update, introspection holds for the given χ if the agent already knew χ before.

There are two points about this operation that are worthwhile to discuss. First, indeed the operation might not work as expected for arbitrary formulas χ, but, again, this should not be seen as a reason for not *interpreting* it as a positive introspection operation. This is not only because the operation does work when the initially-known χ is propositional (i.e., it does not talk about knowledge, the case one typically has in mind), but also because, by the same reasoning, the public announcement operation would not deserve such name, as in general it does not make the announced formula common knowledge among the involved agents (the above discussed Moorean phenomenon). On the other hand, it is also fair to emphasise that this U-disconnecting operation is *not* a positive introspection operation: it is just a model operation which, following a particular representation of an agent's knowledge, has similar effects to a positive introspection reasoning step.

The second point is the similarities between this positive χ-introspection operation and the public announcement operation of the already mentioned *PAL*. The operations act over different components of the model (the operation defined here removes edges, the typical public announcement removes worlds), but they are indeed similar in the sense that, in the new model, former χ-worlds can only reach former χ-worlds.[13] Thus, when the evaluation point w is a χ-world, the resulting models are bisimilar.[14]

Despite the technical similarities with the already mentioned *PAL*, the two operations represent epistemic actions of a very different nature: while a public announcement represents the result of interaction with the environment (external communication, observation), an act of introspection is typically understood as an act of self-reflection. It is then remarkable how their representations are so similar within relational models. One can argue that the introspection action presented here is too drastic, as it removes any 'eventual' (i.e., possibility of having a possibility) uncertainty the agent might have about the given formula, and this is indeed the case. However, it is both the definition of knowledge by means of $\Box$ and this modality's semantic interpretation (in relational models) what leaves no other choice in order to represent this specific epistemic action:

[13] A public announcement can be also defined in terms of edge elimination by preserving only pairs in R whose target is a χ-world.

[14] Another minor difference is the precondition of their modalities, as discussed above.

the semantic interpretation of $\Box$ is the same, regardless of whether it appears under the scope of another $\Box$ modality or not.

4.2. Particular negative introspection

In contrast with the general positive introspection, the general negative introspection operation of the previous section already behaves as expected: it preserves the agent's (propositional) lack of knowledge while giving her negative introspection (Proposition 3.4).[15] This section explores negative introspection over a given χ in order to get an uniform presentation, and also for recognising similarities and differences with respect to its positive counterpart.

An operation for full negative introspection about χ should make sure that all worlds R-reachable from the evaluation point (in zero or more steps, so the original lack of knowledge is preserved and full introspection is reached) can see a $\neg\chi$-world. Assuming that initially the agent does not know χ, this property can be achieved by using a particular instance of the Euclidean closure operation of Definition 3.3 in which the new edges point only to $\neg\chi$-worlds.

Definition 4.3 (U-connecting operation) Let M be a relational model $\langle W, R, V\rangle$, with $U \subseteq W$ a subset of its domain. The *U-connecting* operation yields the model $M_{-U} = \langle W, R', V\rangle$, with its indistinguishability relation R' given by

$$R' := R \cup (\mathfrak{R} \circ (R \cup \mathfrak{R})^{*} \circ R \circ \mathrm{Id}_U^M),$$

with $\mathrm{Id}_U^M := \{(u,u) \mid u \in U\}$. ◀

By instantiating the parameter of this operation with the set of worlds satisfying $\neg\chi$ in the original model, a modality for particular full negative introspection can be defined. Here is a 'preconditionless' version.

Definition 4.4 The language $\mathcal{L}_{\Diamond,-'\chi}$ extends $\mathcal{L}_{\Diamond}$ with a modality $\langle -'\chi\rangle$ for each formula χ. For the semantic interpretation, let (M, w) be a relational state with $M = \langle W, R, V\rangle$. Then,

$$(M, w) \Vdash \langle -'\chi\rangle\,\varphi \quad \text{iff}_{def} \quad (M_{-[\![\neg\chi]\!]^M}, w) \Vdash \varphi.$$

As before, $[-'\chi]\,\varphi := \neg\,\langle -'\chi\rangle\,\neg\varphi$ so, given that the operation is deterministic and its modality does not have a precondition, $[-'\chi]$ is equivalent to $\langle -'\chi\rangle$. ◀

A modality with an appropriate precondition is defined as expected:

$$\langle -\chi\rangle\,\varphi := \neg\,\Box\,\chi \wedge \langle -'\chi\rangle\,\varphi.$$

Thus, the agent can perform an act of particular negative χ-introspection after which φ is the case, $\langle -\chi\rangle\,\varphi$, if and only if she does not know χ, $\neg\,\Box\,\chi$, and after the particular negative χ-introspection operation, φ is the case, $\langle -'\chi\rangle\,\varphi$.

[15] Still, keep in mind its side-effects.

Axiom system In order to formulate reduction axioms for $\langle -'\chi \rangle$, the basic language will be $\mathcal{L}_{PDL^{\triangleleft,?}}$ (Definition 2.4), as now the 'test' operator ? is required. The language $\mathcal{L}_{PDL^{\triangleleft,?,-'\chi}}$ is the result of extending $\mathcal{L}_{PDL^{\triangleleft,?}}$ with the 'dynamic' negative χ-introspection modality, and the reduction axioms and rule of Table 7 below define the required translation from $\mathcal{L}_{PDL^{\triangleleft,?,-'\chi}}$ to $\mathcal{L}_{PDL^{\triangleleft,?}}$, provided that the program transformer from Definition 3.5 is redefined in the following way.

Definition 4.5 (Program transformer T_χ) Let χ be a formula that can be evaluated in relational models. The χ*-program transformer* T_χ, a function from program expressions to program expressions, is defined inductively, with the cases for non-deterministic choice, sequential composition and transitive reflexive closure as in Definition 3.5. The remaining cases, the basic relation, its converse and the test operator, are defined as follows.

$$\mathrm{T}_\chi(\triangleright) := \triangleright \cup (\triangleleft \,;\, (\triangleright \cup \triangleleft)^* \,;\, \triangleright \,;\, ?\neg\chi) \qquad \mathrm{T}_\chi(?\varphi) := ?\,\langle -'\chi \rangle\, \varphi$$
$$\mathrm{T}_\chi(\triangleleft) := \triangleleft \cup (?\neg\chi \,;\, \triangleright \,;\, (\triangleleft \cup \triangleright)^* \,;\, \triangleleft).$$

This program transformer behaves just as the previous one.

Proposition 4.4 *Let M be a relational model $\langle W, R, V \rangle$, with χ a formula that can be evaluated in such models, and recall that $M_{-[\![\neg\chi]\!]^M} = \langle W, R', V \rangle$. For every program expression α, $(R')_\alpha = R_{\mathrm{T}_\chi(\alpha)}$.* ■

Proof. As in Proposition 3.2, the proof is by structural induction on α. The common cases are similar; for the 'test',

$$R'_{?\varphi} = \{(w,w) \mid w \in [\![\varphi]\!]^{M_{-[\![\neg\chi]\!]^M}}\} = \{(w,w) \mid w \in [\![\langle -'\chi \rangle\, \varphi]\!]^M\} = R_{?\langle -'\chi \rangle \varphi} = R_{\mathrm{T}_\chi(?\varphi)}.$$

$-'\chi_p$	$\vdash \langle -'\chi \rangle\, p \;\leftrightarrow\; p$
$-'\chi_\neg$	$\vdash \langle -'\chi \rangle\, \neg\varphi \;\leftrightarrow\; \neg\, \langle -'\chi \rangle\, \varphi$
$-'\chi_\vee$	$\vdash \langle -'\chi \rangle\, (\varphi \vee \psi) \;\leftrightarrow\; (\langle -'\chi \rangle\, \varphi \vee \langle -'\chi \rangle\, \psi)$
$-'\chi_{\langle\alpha\rangle}$	$\vdash \langle -'\chi \rangle\, \langle \alpha \rangle\, \varphi \;\leftrightarrow\; \langle \mathrm{T}_\chi(\alpha) \rangle\, \langle -'\chi \rangle\, \varphi$
SE	If $\vdash \psi_1 \leftrightarrow \psi_2$ then $\vdash \varphi \leftrightarrow \varphi[\psi_2/\psi_1]$, with $\varphi[\psi_2/\psi_1]$ any formula obtained by replacing one or more *non-modality* occurrences of ψ_1 in φ with ψ_2 (see *SE* on Table 6).

Table 7: Axioms and rule for the modality $\langle -'\chi \rangle$.

Theorem 7 (Axiom system for $\mathcal{L}_{PDL^{\triangleleft},-}$) *The axioms and rules of Table 7, together with the axiom system $\mathsf{L}_{PDL^{\triangleleft,?}}$ (Table 3) form a sound and weakly complete axiom system for formulas of $\mathcal{L}_{PDL^{\triangleleft,?,-'\chi}}$ w.r.t. relational models.* ■

Properties of the operation First, as only edges to former $\neg\chi$-worlds are added, it is clear not only that the accessibility relation in $M_{-[\![\neg\chi]\!]^M}$ does not need to be Euclidean, but also that in such model the agent does not need to have full negative introspection about every formula. However, under the proper circumstances, the added edges are enough to obtain full negative introspection about the given χ, as the following analogous of Proposition 4.3 shows.

Proposition 4.5 *Let χ be a formula in $\mathcal{L}_{\Diamond,-'\chi}$ such that $\Vdash \neg\chi \to [-'\chi]\,\neg\chi$. If the agent does not know χ, then after the operation the agent will have* full *negative introspection about χ. More precisely, for any $n \geq 0$ we have*

$$\Vdash \neg\,\Box\,\chi \to \langle -\chi\rangle\,\Box^n\,\neg\,\Box\,\chi. \qquad (3) \qquad \blacksquare$$

Proof. Let (M, w) be a relational state, $M = \langle W, R, V\rangle$, and suppose $(M, w) \Vdash \neg\,\Box\,\chi$; then there is $v \in W$ such that Rwv and $(M, v) \Vdash \neg\chi$, with the latter implying $\mathrm{Id}^M_{\neg\chi}\,vv$ (by definition) and $(M_{-[\![\neg\chi]\!]^M}, v) \Vdash \neg\chi$ (by the assumption).

The first step is to show (by induction on $n \geq 0$) how, in $M_{-[\![\neg\chi]\!]^M}$, any world that can be reached from w in zero or more steps can also reach v, that is, how $(R')^n wu$ implies $R'uv$.

- Suppose $(R')^0 wu$; then, no R-steps are needed to reach u, i.e., $u = w$. But Rwv so $R'wv$, that is, $R'uv$.
- Suppose $(R')^1 wu$; then $R'wu$. Now, from the definition of R' there are two possibilities.
  - If Rwu then $\mathfrak{R}uw$, which together with Rwv and $\mathrm{Id}^M_{\neg\chi}\,vv$ implies $(\mathfrak{R} \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$, so $R'uv$.
  - If $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})wu$, then there are $u_1, u_2 \in W$ such that $\mathfrak{R}wu_1$, $(R \cup \mathfrak{R})^* u_1u_2$ and Ru_2u; thus, $\mathfrak{R}uu_2$, $(R \cup \mathfrak{R})^* u_2u_1$ and Ru_1w, which together with Rwv and $\mathrm{Id}^M_{\neg\chi}\,vv$ imply $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$, so $R'uv$.
- Suppose $(R')^{n+2} wu$; then there is $u' \in W$ such that $(R')^{n+1} wu'$ and $R'u'u$, and hence (inductive hypothesis) $R'u'v$ and $R'u'u$. From the definition of R', this yields four cases.
  - If $Ru'u$ and $Ru'v$, then $(\mathfrak{R} \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$, so $R'uv$.
  - If $Ru'u$ and $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})u'v$, then $(\mathfrak{R} \circ \mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$ so $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$, that is, $R'uv$.
  - If $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})u'u$ and $Ru'v$, then $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R)u'u$ and $Ru'v$, so $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$ and hence $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$, that is, $R'uv$.
  - If $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})u'u$ and $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})u'v$, then $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R)u'u$ and $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})u'v$, so $(\mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathfrak{R} \circ (R \cup \mathfrak{R})^* \circ R \circ \mathrm{Id}^M_{\neg\chi})uv$ and hence $R'uv$.

Now, for $(M,w) \Vdash \langle -\chi\rangle \Box^n \neg \Box \chi$, take $n \geq 0$ and $u \in W$ such that $(R')^n wu$. Then $R'uv$ and, from $(M_{-[\![\neg\chi]\!]^M}, v) \Vdash \neg\chi$, it follows that $(M_{-[\![\neg\chi]\!]^M}, u) \Vdash \Diamond \neg\chi$, i.e., $(M_{-[\![\neg\chi]\!]^M}, w) \Vdash \Box^n \neg \Box \chi$, and hence $(M,w) \Vdash \langle -'\chi\rangle \Box^n \neg \Box \chi$. But $\langle -\chi\rangle$'s precondition holds; thus, $(M,w) \Vdash \langle -\chi\rangle \Box^n \neg \Box \chi$, as required. ∎

This validity is, again, a *dynamic* version of the negative introspection axiom, now focussing on a specific χ. The axiom system also allow us to obtain further validities as the following one, characterising the knowledge of the agent after the operation (without unfolding $\mathrm{T}_\chi(\triangleright)$, to simplify its presentation).

Proposition 4.6 *Let χ and φ be formulas in $\mathcal{L}_{PDL^{\triangleleft,?,-'\chi}}$. Then,*

$$\Vdash \langle -\chi\rangle [\triangleright] \varphi \leftrightarrow (\neg [\triangleright] \chi \wedge [\mathrm{T}_\chi(\triangleright)][-'\chi] \varphi).$$ ∎

As before, it is possible to use the previous validity to obtain one characterising negative introspection about a given χ after the operation.

$$\Vdash \langle -\chi\rangle [\triangleright] \neg [\triangleright] \chi \;\leftrightarrow\; \big(\neg [\triangleright] \chi \wedge [\mathrm{T}_\chi(\triangleright)][-'\chi] \neg [\triangleright] \chi\big).$$

Still, there is s caveat. Just as in the general case, the particular negative introspection operation might cause the agent to drop propositional knowledge.

Fact 4.2 *The formula $\Box q \to [-p] \Box q$ is not valid.*

Proof. In the relational state below on the left, $\Box q$ holds; however, this is not the case in the relational state on the right, the one resulting from a negative introspection operation for the agent's lack of knowledge *about* p.

M $\qquad\qquad$ M^-

Moreover: the operation can be performed (the initial relational state satisfies the action's precondition, $\neg \Box p$) and it is actually needed (the agent is not negatively introspective about her lack of knowledge of p, that is, $\neg \Box \neg \Box p$). Thus, $(M,w) \Vdash \Box q \wedge \langle -p\rangle \neg \Box q$. ∎

5. One-step introspection actions

When successful, the operations discussed in the previous sections make the agent's knowledge fully introspective, either about all her knowledge (Section 3), or else about the given formula (Section 4). Thus, they achieve the goal set in the introduction: although a 'more real' agent might not have full positive/negative introspection, she can achieve it in the appropriate cases by performing the appropriate epistemic actions.

However, these actions can be also seen as too strong, as they give the agent, in a single shoot, *full* introspection. If the goal is indeed to model 'more real' agents, it makes sense to explore also introspection operations that give her not full introspection, but rather increase her introspection level by only 'one degree'. From a technical point of view, operations representing these epistemic actions should be different from the ones studied before: in order to increase the agent's introspection by 'one degree', the operation should take not a *global* perspective, as the previously studied do, but rather a *local* one, so the introspection the agent has *at a given evaluation point* will be increased.[16]

5.1. One-step positive χ-introspection

Take a relational state (M, w) in which the agent knows a given χ, and yet she does not have full positive introspection about it. This means that there is a $n \geq 2$ such that, although every world R-reachable from w in n steps satisfies χ, there is at least one world R-reachable from w in $n+1$ steps in which $\neg\chi$ is the case. In order for the agent to gain one 'degree' of positive introspection about χ, such $\neg\chi$-worlds should not be R-reachable from w anymore; in other words, only χ-worlds should be R-reachable from w in $n+1$ steps. Unlike the full positive introspection of before (Definition 4.1), this is a *local* operation, and thus it intends to go from $\Box^n \chi$ to $\Box^{n+1} \chi$ without necessarily reaching $\Box^{n+2} \chi$.

Use $R[w]$ to denote the set of worlds reachable from w via R in one step (that is, $R[w] := \{v \mid Rwv\}$), and $R^n[w]$ to denote the set of worlds reachable from w via R in n steps (that is, $R^1[w] := R[w]$ and $R^{n+1}[w] := \bigcup_{v \in R[w]} R^n[v]$).

Definition 5.1 (Single-step positive *U*-introspection) Let (M, w) be a relational state with $M = \langle W, R, V\rangle$. Let $U \subseteq W$ a subset of its domain, and let m be the smallest natural number such that $R^m[w] \cap \overline{U} \neq \varnothing$ (i.e., m is the number of steps required to reach the worlds closest to w that are not in U). The *single-step positive U-introspection* operation yields the model $M_{w+U} = \langle W, R_w^{+U}, V\rangle$, with R_w^{+U} given by

$$R_w^{+U} := R \setminus \left\{(u, v) \mid u \in R^{m-1}[w] \text{ and } v \in (R[u] \cap \overline{U})\right\}. \qquad \blacktriangleleft$$

Some words about the just defined operation. First, if no world in $\overline{U}$ is reachable from w (i.e., if only worlds in U can be reached from w), then the mentioned m does not exist. In such cases, the set of pairs that are being removed from R is empty, and thus the operation returns exactly the same model. Second, when such m exists, the operation removes the last edge in

[16] Note how, to increase the introspection degree by one, one has to work locally (with respect to some fixed evaluation point). However, one can work locally and also reach full introspection, simply by performing the operations of Section 5 only on the sub-model generated by the given evaluation point.

every path from w that leads to its closest $\overline{U}$-worlds. Third, when the parameter of this operation, the set U, is given by the truth-set of a formula, say χ, the operation can be understood as single-step positive χ-introspection action, as it takes a relational state in which the worlds closest to w that are not in $[\![\chi]\!]^M$ (i.e., those in $[\![\neg\chi]\!]^M$) are m steps away, and returns a relational state (the new model plus the same evaluation point) in which the worlds closest to w that are in $[\![\neg\chi]\!]^M$ will be at *at least* $m+1$ steps away.

Definition 5.2 The language $\mathcal{L}_{\Diamond,\langle\!\langle+\chi\rangle\!\rangle}$ extends $\mathcal{L}_{\Diamond}$ with the modality $\langle\!\langle+\chi\rangle\!\rangle$. For its semantic interpretation, let (M,w) be a relational state. Then,

$$(M,w) \Vdash \langle\!\langle+\chi\rangle\!\rangle\,\varphi \quad \text{iff}_{def} \quad (M,w) \Vdash \Box\,\chi \text{ and } (M_{w+[\![\chi]\!]^M}, w) \Vdash \varphi.$$

In words, the agent can perform a single-step positive introspection action for χ after which φ is the case, $(M,w) \Vdash \langle\!\langle+\chi\rangle\!\rangle\,\varphi$, if and only if she knows χ, $(M,w) \Vdash \Box\,\chi$, and, after the action, φ is the case, $(M_{w+[\![\chi]\!]^M}, w) \Vdash \varphi$. The universal modality $[[+\chi]]\,\varphi$ is defined as the modal dual of $\langle\!\langle+\chi\rangle\!\rangle\,\varphi$, as usual.◀

Notice that we require the agent to know χ; thus, the m referred in the operation's definition is such that $m \geq 2$. Note also how, when the agent has full positive introspection about χ, no world in $[\![\neg\chi]\!]^M$ is reachable from w, and hence the m mentioned in the operation's definition does not exist. Thus, as remarked above, the operation returns exactly the same model.

Unlike the previous sections, we have not introduced an axiom system for the operators defined here. Due to the locality of the operations introduced in this section, we believe that more powerful tools needs to be used, as it will be discussed in the conclusions.

Properties of the operation Let us analyse the effects of having this new modality. Remember that we only are interested about what are the effects on taking one introspection step from a non-introspective level reachable from the evaluation point, without taking into account what happens beyond such step.

The operator behaves as expected, not only for propositional formulas, but also for those whose truth is preserved by the operation.

Proposition 5.1 *Let χ be a formula such that $\Vdash \chi \rightarrow [[+\chi]]\,\chi$. Then,*

$$\Vdash \Big(\bigwedge_{i=1}^{n} \Box^i\,\chi \;\wedge\; \neg\,\Box^{n+1}\,\chi\Big) \rightarrow \langle\!\langle+\chi\rangle\!\rangle \bigwedge\nolimits_{i=0}^{n+1} \Box^i\,\chi.$$

■

Proof. Take any relational model (M,w) with $M = \langle W, R, V\rangle$. The antecedent of the formula whose validity should be proved indicates that the first 'problematic' (i.e., $\neg\chi$-) world is $n+1$ steps away from the evaluation point w; thus, in Definition 5.1, $m = n+1$. Here it needs to be shown that $i \in \{1,\ldots,n+1 = m\}$ implies $(R_w^{+U})^i[w] \subseteq [\![\chi]\!]^{M_{w+U}}$.

Note that the operation does not add edges, so being reachable in the new model implies being reachable in the original one: $(R_w^{+U})^i[w] \subseteq R^i[w]$.

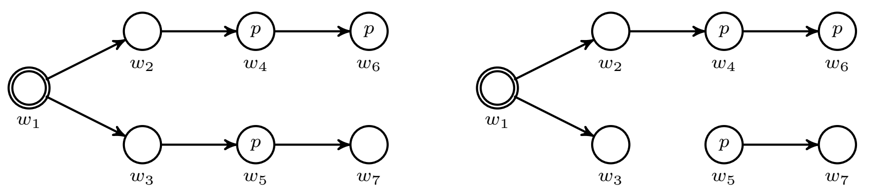
Take first any $i \in \{1, \ldots, n = m-1\}$. The antecedent of the formula gives us $R^i[w] \subseteq [\![\chi]\!]^M$, and the extra requirement states that $[\![\chi]\!]^M \subseteq [\![\chi]\!]^{M_{w+U}}$. Hence, $(R_w^{+U})^i[w] \subseteq [\![\chi]\!]^{M_{w+U}}$. Now take $n+1 = m$. After the operation, links from worlds in $R^{m-1=n}[w]$ to worlds in $[\![\neg\chi]\!]^M$ are cut, so no world reachable from w in $m = n+1$ steps satisfies $\neg\chi$ in M: $(R_w^{+U})^{n+1}[w] \subseteq [\![\chi]\!]^M$. Then, again from the extra requirement, $[\![\chi]\!]^M \subseteq [\![\chi]\!]^{M_{w+U}}$, so $(R_w^{+U})^{n+1}[w] \subseteq [\![\chi]\!]^{M_{w+U}}$. Thus, the operation produces a model satisfying the required formula.

Finally, in the formula, n's lower bound is 1. Thus, the precondition of the modality is satisfied, i.e., its 'diamond' version indeed holds. ■

However, the result cannot be generalized for arbitrary modal formulas:

Fact 5.1 $\not\Vdash (\bigwedge_{i=1}^{n} \Box^i \chi \wedge \neg \Box^{n+1} \chi) \to \langle\!\langle +\chi \rangle\!\rangle \bigwedge_{i=0}^{n+1} \Box^i \chi$. *In particular,* $(\Box\chi \wedge \neg\Box\Box\chi) \wedge \neg\langle\!\langle +\chi \rangle\!\rangle \Box\chi$ *is satisfiable.*

Proof. Let $\chi := \Diamond p$, and consider the relational state (M, w_1) below on the left. Clearly we have $(M, w_1) \Vdash (\Box\Diamond p \wedge \neg\Box\Box\Diamond p)$, since all one-steps successors from w_1 have at least one p-successor, but some two-steps successors (w_5) do not have any p-successor.



M $\qquad$ $M_{w_1+\Diamond p}$

After updating the accessibility relation with w_1 the reference point, we obtain $M_{w_1+\Diamond p}$. Notice that we have now $(M_{w_1+\Diamond p}, w) \not\Vdash \Box\Diamond p$, since the edge between w_3 and w_5 has been removed, so the agent lost some knowledge. ■

Fact 5.1 says that the new modality describes one step of positive introspection, but at the cost of losing knowledge about formulas whose truth-sets might shrink after the operation (i.e., formulas not satisfying the additional requirement on Proposition 5.1). Notice that the defined model update operation removes edges *at the end* of paths leading to the closest problematic worlds (those not satisfying χ). One could then propose an alternative operation that removes edges *at the beginning* of such paths:

$$R_w^{+U} := R \setminus \left\{(w, u) \mid u \in R[w] \text{ and } v \in (R^{m-1}[u] \cap \overline{U})\right\}.$$

Still, although this alternative works properly for formulas whose truth-set does not shrink (the argument is similar to the one given for Proposition 5.1), it still fails in the general case.

Fact 5.2 *The alternative single-step positive introspection operation does not work for arbitrary modal formulas.*

Proof. Let $\chi := \Diamond p$, and consider the relational state (M, w_1) below on the left. Clearly we have $(M, w_1) \Vdash (\Box \Diamond p \wedge \neg \Box \Box \Diamond p)$, since all one-steps successors from w_1 (w_1 and w_2) have at least one p-successor, but some two-steps successors (w_3) do not have any p-successor.

The relational state that results from the above sketched operation appears above on the right. On it, $\Box \Diamond p$ fails: there is a world (w_1) accessible from w_1 that does not have a p-successor. ∎

5.2. One-step negative χ-introspection

Take now a relational state (M, w) in which the agent does not know a given χ, and yet she does not have full negative introspection about it. Then, there is $n \geq 0$ such that, although every world R-reachable from w in n steps has a $\neg\chi$-successor, there is at least one world R-reachable from w in $n+1$ steps that can reach only χ-worlds. In order for the agent to gain one 'degree' of negative introspection about χ, all elements of $R^{n+1}[w]$ should have such successor: for every world in $u \in R^{n+1}[w]$ there should be a world $v \in R[u]$ in which χ fails. Unlike the full negative introspection of before (Definition 4.3), this is a *local* operation, and thus it intends to go from $\Box^n \neg \Box \chi$ to $\Box^{n+1} \neg \Box \chi$ without necessarily reaching $\Box^{n+2} \neg \Box \chi$. Unlike the its single-step positive introspection counterpart (Definition 5.1), edges should be added instead of being removed.

Definition 5.3 (Single-step negative U-introspection) Let (M, w) be a relational state with $M = \langle W, R, V \rangle$. Let $U \subseteq W$ a subset of its domain, and let m be the smallest natural number for which there exists at least one $u \in R^m[w]$ such that $R[u] \subseteq U$ (i.e., m is the number of steps required to reach the worlds closest to w whose successors are all in U). The *single-step negative U-introspection* operation yields the model $M_{w-U} = \langle W, R_w^{-U}, V \rangle$, with R_w^{-U} given by

$$R_w^{-U} := R \cup \left\{ (u, v) \mid u \in R^m[w], R[u] \subseteq U \text{ and } v \in (R[w] \cap \overline{U}) \right\}. \qquad \blacktriangleleft$$

Some words about the just defined operation. First, if every world that can be reached from w has a $\overline{U}$-successor, then the mentioned m does not exist. In such cases, the set of pairs that are being added to R is empty, and thus the operation returns exactly the same model. Second, when such m exists, the operation attempts to add edges from the worlds closest to w whose successors are all in U to all successors of w that are not in U. Of course, if the latter worlds do not exist, the set of pairs that are being added to R is empty too, and thus once again the operation returns exactly the same model. Third, when the parameter of this operation, the set U, is given by the truth-set of a formula, say χ, the operation can be understood as single-step negative χ-introspection action, as it takes a relational state in which the worlds closest to w that have only $[\![\chi]\!]^M$-successors are m steps away, and returns a relational state (the new model plus the same evaluation point) in which the worlds closest to w that have only $[\![\chi]\!]^M$-successors should be at *at least* $m+1$ steps away.

Definition 5.4 The language $\mathcal{L}_{\Diamond,\langle\!\langle -\chi\rangle\!\rangle}$ extends $\mathcal{L}_{\Diamond}$ with the modality $\langle\!\langle -\chi\rangle\!\rangle$. For its semantic interpretation, let (M, w) be a relational state. Then,

$$(M,w) \Vdash \langle\!\langle -\chi\rangle\!\rangle\,\varphi \quad \text{iff}_{def} \quad (M,w) \Vdash \neg\,\Box\,\chi \text{ and } (M_{w-[\![\chi]\!]^M}, w) \Vdash \varphi.$$

In words, the agent can perform a single-step negative introspection action for χ after which φ is the case, $(M,w) \Vdash \langle\!\langle -\chi\rangle\!\rangle\,\varphi$, if and only if she does not know χ, $(M,w) \Vdash \Box\,\chi$, and, after the action, φ is the case, $(M_{w-[\![\chi]\!]^M}, w) \Vdash \varphi$. The universal modality $[[-\chi]]\,\varphi$ is defined as the modal dual of $\langle\!\langle -\chi\rangle\!\rangle\,\varphi$, as usual.◀

Note how the modality's precondition requires for w to have a $\neg\chi$-successor; thus, the m referred in the operation's definition is such that $m \geq 1$. Moreover, this guarantees that, for the 'diamond' existential modality $\langle\!\langle -\chi\rangle\!\rangle$ (the one on which the precondition acts as a conjunct), a $\neg\chi$-world v as the one required by the operation always exists. Note also how when, the agent has full negative introspection about χ, all worlds reachable from w have $\neg\chi$-successors, and hence the m mentioned in the operation's definition does not exist. Thus, as remarked above, the operation returns exactly the same model.

In the same way as for the one step positive introspection operator, we have not introduced an axiom system. Again, the reasons is that further tools are needed in order to deal with this local operation.

Properties of the operation Here is the crucial proposition.

Proposition 5.2 *Let χ be a formula such that $\Vdash \neg\chi \rightarrow [[-\chi]]\,\neg\chi$. Then,*

$$\Vdash \Big(\bigwedge_{i=0}^{n} \Box^i \neg\,\Box\,\chi \;\wedge\; \neg\,\Box^{n+1} \neg\,\Box\,\chi\Big) \rightarrow \langle\!\langle -\chi\rangle\!\rangle \Big(\bigwedge_{i=0}^{n+1} \Box^i \neg\,\Box\,\chi\Big).$$

■

Proof. Take any relational model (M, w) with $M = \langle W, R, V\rangle$. The antecedent of the formula whose validity should be proved indicates that the first 'problematic' world (one without a $\neg\chi$-successor) is $n+1$ steps away from the evaluation point w; thus, in Definition 5.3, $m = n+1$. Here it needs to be shown that $i \in \{1, \ldots, n+1 = m\}$ implies $(R_w^{-U})^i[w] \subseteq [\![\Diamond\,\chi]\!]^{M_{w-U}}$.

Take first any $i \in \{0, \ldots, n = m-1\}$. The antecedent of the formula gives us $R^i[w] \subseteq [\![\Diamond\,\neg\chi]\!]^M$, that is, $u \in R^i[w]$ implies $R[u] \cap [\![\neg\chi]\!]^M \neq \varnothing$. For the proof, take any $u \in (R_w^{-U})^i[w]$. The first edges added by the operation are from worlds that are m steps away from w, so $R^i[w] = (R_w^{-U})^i[w]$. Hence, $u \in R^i[w]$, and therefore there is $v \in (R[u] \cap [\![\neg\chi]\!]^M)$. Now, the operation does not remove edges, so $R[u] \subseteq R_w^{-U}[u]$, and the extra requirement states that $[\![\neg\chi]\!]^M \subseteq$

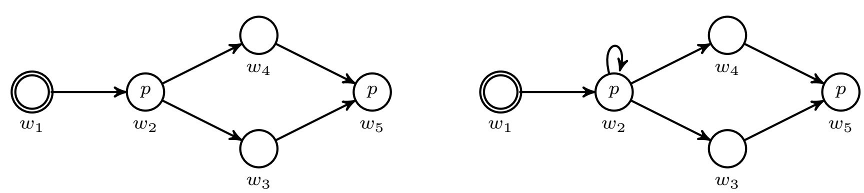
$[\![\neg\chi]\!]^{M_{w\text{-}U}}$; then, $v \in (R_w^{-U}[u] \cap [\![\neg\chi]\!]^{M_{w\text{-}U}})$, that is, $R_w^{-U}[u] \cap [\![\neg\chi]\!]^{M_{w\text{-}U}} \neq \varnothing$. Therefore, so $(R_w^{-U})^i[w] \subseteq [\![\Diamond \neg\chi]\!]^{M_{w\text{-}U}}$, as required.

Now, the $n+1=m$ case. Take any $u \in (R_w^{-U})^{n+1=m}[w]$. Again, the first edges added by the operation are from worlds that are m steps away from w, so $R^{n+1}[w] = (R_w^{-U})^{n+1}[w]$, so $u \in R^{n+1=m}[w]$. Here there are two possibilities, depending on whether $R[u] \subseteq [\![\chi]\!]^M$ holds. On the one hand, if it does not (so no edges are added from u), then $R[u] \cap [\![\neg\chi]\!]^M \neq \varnothing$, that is, there is $v \in (R[u] \cap [\![\neg\chi]\!]^M)$. But then $v \in R_w^{-U}[u]$ (no edges are removed by the operation) and $v \in [\![\neg\chi]\!]^{M_{w\text{-}U}}$ (the extra requirement); thus, $R_w^{-U}[u] \cap [\![\neg\chi]\!]^{M_{w\text{-}U}} \neq \varnothing$. On the other hand, if it does, the operation guarantees that, in the new model, there is a v reachable from u such that $[\![\neg\chi]\!]^M$ (such v in $R[w] \cap [\![\neg\chi]\!]^M$ exists because of the antecedent of the formula) and, from the requirement once more, $v \in (R_w^{-U}[u] \cap [\![\neg\chi]\!]^{M_{w\text{-}U}})$, that is, $R_w^{-U}[u] \cap [\![\neg\chi]\!]^{M_{w\text{-}U}} \neq \varnothing$ again. Thus, in both cases, in the new model u has a successor that satisfies $\neg\chi$; hence, $(R_w^{-U})^{n+1}[w] \subseteq [\![\Diamond \chi]\!]^{M_{w\text{-}U}}$, as required. ■

Again, the property above does not hold for arbitrary formulas:

Fact 5.3 $\not\Vdash (\bigwedge_{i=0}^{n} \Box^i \neg \Box \chi \;\wedge\; \neg \Box^{n+1} \neg \Box \chi) \rightarrow \langle\!\langle -\chi \rangle\!\rangle (\bigwedge_{i=0}^{n+1} \Box^i \neg \Box \chi)$. *In particular,* $\neg \Box \chi \wedge \neg \Box \neg \Box \chi \wedge \langle\!\langle -\chi \rangle\!\rangle (\Box \chi)$ *is satisfiable.*

Proof. Let $\chi := \Diamond p$, and consider the relational state (M, w_1) below on the left. Clearly we have $(M, w_1) \Vdash (\neg \Box \Diamond p \wedge \neg \Box \neg \Box \Diamond p)$, or equivalently $(M, w_1) \Vdash (\Diamond \Box \neg p \wedge \Diamond \Box \Diamond p)$, since there exists a successor of w_1 whose successors satisfy both $\neg p$ and $\Diamond p$.



M $\qquad\qquad$ $M_{w_1 - \Diamond p}$

After updating the accessibility relation with w_1 the reference point, we obtain $M_{w_1 - \Diamond p}$. Notice that we have now $(M_{w_1 + \Diamond p}, w) \Vdash \Box \Diamond p$, because we add a loop in the node w_2, producing a change on the original knowledge. ■

6. Conclusion and further work

This paper studies positive and negative introspection as epistemic actions that modify the agent's knowledge. In both cases, three possibilities are considered: operations changing the general knowledge of the agent, operations updating the agent's knowledge with respect a particular formula, and local operations adding one step to the agent's introspection.

The general operations work by giving the indistinguishability relation the relational properties that guarantee the introspection properties. In the positive introspection case, edges are *added* to make the relation transitive; yet, as discussed, this idea works by considering that introspection fails not because of what the agent knows about her knowledge, but rather because of what she knows. Thus, the operation drops non-introspective knowledge, only preserving the introspective one. In the negative introspection case, edges are *added* to make the relation Euclidean, and the operation works for those formulas whose falsity is preserved by the operation.

The particular operations focus in a specific formula instead. In the positive introspection case, it now *eliminates* edges from χ-worlds to $\neg\chi$-worlds, forcing the agent to know that she knows χ while keeping the rest of her knowledge 'as before'. In the negative introspection case, it *adds* edges once again. In both cases, the operation works as expected for formulas whose truth/falsity is preserved by the operation, respectively.

Finally, the operations for increasing introspection by 'one degree', i.e., for going from $\Box^n \varphi$ to $\Box^{n+1} \varphi$ without reaching necessarily $\Box^{n+2} \varphi$, and for going from $\Box^n \neg \Box \varphi$ to $\Box^{n+1} \neg \Box \varphi$, without necessarily reaching $\Box^{n+2} \neg \Box \varphi$. In order to achieve that, both operations work locally, looking for the first 'problematic' world (the closest to the evaluation point disproving the introspection property), and deleting/adding edges accordingly, respectively.

Modalities for the four global operations (two general, two particular) have been axiomatised by means of the *DEL* reduction axioms strategy, translating formulas with dynamic modalities into formulas without them. However, the local operations require further tools, as they should work only on worlds affecting an introspection property *from the point of view of the given evaluation point*. Therefore, we conjecture that more powerful tools are needed, such as nominals and hybrid logic binders (see e.g. [58] for details) in order to refer to specific points in a path of the model.

An aspect that has not been discussed is the computational behaviour of the proposed logics. The global ones can be effectively translated into some decidable logic (*PDL*) via the provided reduction axioms; thus,

Corollary 6.1 *The satisfiability problem for all the logics introduced in Sections 3 and 4 is decidable.* ■

For future work, there are some natural lines. First, on the specific proposal, one can look for the additional tools required to provide a sound and complete axiomatisations of the modalities for local operators. Then, one would like to analyse the exact complexity of all the logics we studied, as well as finding concrete applications, for instance for reasoning problems in artificial intelligence. A second direction consists on taking a step back and look at additional interesting introspection operations: an appealing one would provide both positive and negative introspection *simultaneously*.

Finally, an interesting project is to investigate similar operations in a multi-agent setting (e.g., public, private versions of these operations), focusing additionally on operations for reaching common knowledge.

Acknowledgements. This work was partially supported by ANPCyT- PICTs-2017-1130 and 2016-0215, MinCyT Córdoba, SeCyT-UNC and the Laboratoire International Associé INFINIS.

References

[1] J. Hintikka, Knowledge and Belief, Cornell University Press, Ithaca, N.Y., 1962.

[2] M. Y. Vardi, On epistemic logic and logical omniscience, in: J. Y. Halpern (Ed.), TARK, Morgan Kaufmann, 1986, pp. 293–305.

[3] R. Stalnaker, The problem of logical omniscience, I, Synthese 89 (3) (1991) 425–440. doi:10.1007/BF00413506.

[4] R. Stalnaker, On logics of knowledge and belief, Philosophical Studies 128 (1) (2006) 169–199. doi:10.1007/s11098-005-4062-y.

[5] E. J. Lemmon, If I know, do I know that I know?, in: Stroll [59], pp. 54–83.

[6] A. C. Danto, On knowing that we know, in: Stroll [59], pp. 32–53.

[7] R. Hilpinen, Reviewed work: If I know, do I know that I know? by E. J. lemmon, The Journal of Symbolic Logic 38 (4) (1973) 662. doi:10.2307/2272022.

[8] T. Williamson, Knowledge and its Limits, Oxford University Press, 2002. doi:10.1093/019925656X.001.0001.

[9] D. Bonnay, P. Égré, Inexact knowledge with introspection, Journal of Philosophical Logic 38 (2) (2009) 179–227. doi:10.1007/s10992-008-9087-1.

[10] D. Bonnay, P. Égré, Knowing one's limits: An analysis in centered dynamic epistemic logic, in: P. Girard, O. Roy, M. Marion (Eds.), Dynamic Formal Epistemology, Vol. 351 of Synthese Library, Springer Netherlands, 2011, pp. 103–126. doi:10.1007/978-94-007-0074-1_6.

[11] R. Hilpinen, Knowing that one knows and the classical definition of knowledge, Synthese 21 (2) (1970) 109–132. doi:10.1007/BF00413541.

[12] M. Bremer, On the incompatibility of negative introspection and knowledge as true belief (2009).
URL http://www.mbph.de/Logic/NegativeIntrospection.pdf

[13] A. S. Gillies, A new solution to moore's paradox, Philosophical Studies 105 (3) (2001) 237–250. doi:10.1023/A:1010361708803.

[14] J. van Benthem, What one may come to know, Analysis 64 (2) (2004) 95–105. doi:10.1093/analys/64.2.95.

[15] J. Gerbrandy, The surprise examination in dynamic epistemic logic, Synthese 155 (1) (2007) 21–33. doi:10.1007/s11229-005-2211-7.

[16] H. N. Duc, Reasoning about rational, but not logically omniscient, agents, Journal of Logic and Computation 7 (5) (1997) 633–648. doi:10.1093/logcom/7.5.633.

[17] T. Ågotnes, N. Alechina, The dynamics of syntactic knowledge, Journal of Logic and Computation 17 (1) (2007) 83–116. doi:10.1093/logcom/exl019.

[18] M. Jago, Epistemic logic for rule-based agents, Journal of Logic, Language and Information 18 (1) (2009) 131–158. doi:10.1007/s10849-008-9071-8.

[19] A. Solaki, Steps out of logical omniscience, Master's thesis, Institute for Logic, Language and Computation, Amsterdam, ILLC Master of Logic Thesis Series MoL-2017-12 (2017).

[20] J. van Benthem, F. R. Velázquez-Quesada, The dynamics of awareness, Synthese (Knowledge, Rationality and Action) 177 (Supplement 1) (2010) 5–27. doi:10.1007/s11229-010-9764-9.

[21] D. Grossi, F. R. Velázquez-Quesada, Syntactic awareness in logical dynamics, Synthese 192 (12) (2015) 4071–4105. doi:10.1007/s11229-015-0733-1.

[22] J. van Benthem, E. Pacuit, Dynamic logics of evidence-based beliefs, Studia Logica 99 (1) (2011) 61–92. doi:10.1007/s11225-011-9347-x.

[23] F. R. Velázquez-Quesada, Explicit and implicit knowledge in neighbourhood models, in: D. Grossi, O. Roy, H. Huang (Eds.), Proceedings of LORI 2013 Workshop, Vol. 8196 of Lecture Notes in Computer Science, Springer, 2013, pp. 239–252. doi:10.1007/978-3-642-40948-6_19.

[24] J. E. Hopcroft, R. Motwani, J. D. Ullman, Introduction to automata theory, languages, and computation, Addison-Wesley, 2003.

[25] D. Harel, D. Kozen, J. Tiuryn, Dynamic Logic, MIT Press, Cambridge, MA, 2000.

[26] J. Y. Halpern, M. Y. Vardi, The complexity of reasoning about knowledge and time: Extended abstract, in: J. Hartmanis (Ed.), Proceedings of the 18th ACM Symposium on Theory of Computing, Berkeley, California, USA, ACM, 1986, pp. 304–315. doi:10.1145/12130.12161.

[27] R. Fagin, J. Y. Halpern, Y. Moses, M. Y. Vardi, Reasoning about knowledge, The MIT Press, 1995.

[28] R. Parikh, R. Ramanujam, A knowledge based semantics of messages, Journal of Logic, Language and Information 12 (4) (2003) 453–467. doi:10.1023/A:1025007018583.

[29] H. van Ditmarsch, W. van der Hoek, B. Kooi, Dynamic Epistemic Logic, Vol. 337 of Synthese Library Series, Springer, Dordrecht, 2008.

[30] J. van Benthem, Logical Dynamics of Information and Interaction, Cambridge University Press, 2011.

[31] G. Aucher, P. Balbiani, L. F. del Cerro, A. Herzig, Global and local graph modifiers, Electr. Notes Theor. Comput. Sci. 231 (2009) 293–307. doi:10.1016/j.entcs.2009.02.042.

[32] A. Baltag, L. S. Moss, S. Solecki, The logic of public announcements and common knowledge and private suspicions, in: I. Gilboa (Ed.), TARK, Morgan Kaufmann, San Francisco, CA, USA, 1998, pp. 43–56.

[33] J. van Benthem, J. van Eijck, B. Kooi, Logics of communication and change, Information and Computation 204 (11) (2006) 1620–1662. doi:10.1016/j.ic.2006.04.006.

[34] A. Baltag, S. Smets, A qualitative theory of dynamic interactive belief revision, in: G. Bonanno, W. van der Hoek, M. Wooldridge (Eds.), Logic and the Foundations of Game and Decision Theory, Amsterdam University Press, 2008, pp. 13–60.

[35] J. van Benthem, Dynamic logic for belief revision, Journal of Applied Non-Classical Logics 17 (2) (2007) 129–155. doi:10.3166/jancl.17.129-155.

[36] J. van Benthem, F. Liu, Dynamic logic of preference upgrade, Journal of Applied Non-Classical Logics 17 (2) (2007) 157–182. doi:10.3166/jancl.17.157-182.

[37] S. Ghosh, F. R. Velázquez-Quesada, A note on reliability-based preference dynamics, in: W. van der Hoek, W. H. Holliday, W. Wang (Eds.), Proceedings of LORI 2015 Workshop, Vol. 9394 of Lecture Notes in Computer Science, Springer, 2015, pp. 129–142. doi:10.1007/978-3-662-48561-3_11.

[38] F. R. Velázquez-Quesada, Reliability-based preference dynamics: lexicographic upgrade, Journal of Logic and Computation 27 (8) (2017) 2341–2381. doi:10.1093/logcom/exx019.

[39] J. van Benthem, An Essay on Sabotage and Obstruction, in: Mechanizing Mathematical Reasoning, 2005, pp. 268–276.

[40] C. Areces, R. Fervari, G. Hoffmann, Moving Arrows and Four Model Checking Results, in: Logic, Language, Information and Computation, Vol. 7456 of Lecture Notes in Computer Science, Springer, 2012, pp. 142–153.

[41] C. Areces, R. Fervari, G. Hoffmann, Swap logic, Logic Journal of the IGPL 22 (2) (2014) 309–332.

[42] R. Fervari, Relation-changing modal logics, Ph.D. thesis, Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Argentina (2014).

[43] C. Areces, R. Fervari, G. Hoffmann, Relation-changing modal operators, Logic Journal of the IGPL 23 (4) (2015) 601–627.

[44] B. Kooi, B. Renne, Arrow update logic, The Review of Symbolic Logic 4 (2011) 536–559. doi:10.1017/S1755020311000189.

[45] R. Fervari, F. R. Velázquez-Quesada, Dynamic epistemic logics of introspection, in: A. Madeira, M. Benevides (Eds.), Dynamic Logic. New Trends and Applications - First International Workshop, DALI 2017, Brasilia, Brazil, September 23-24, 2017, Proceedings, Vol. 10669 of Lecture Notes in Computer Science, Springer, 2017, pp. 82–97. doi:10.1007/978-3-319-73579-5_6.

[46] B. F. Chellas, Modal Logic: An Introduction, Cambridge University Press, 1980.

[47] P. Blackburn, M. de Rijke, Y. Venema, Modal logic, Cambridge University Press, 2001.

[48] R. Thomason, Independence of the dual axiom in modal K with primitive $\diamond$, Notre Dame Journal of Formal Logic 59 (3) (2018) 381–385. doi:10.1215/00294527-3817906.

[49] G. De Giacomo, Eliminating "converse" from converse PDL, Journal of Logic, Language and Information 5 (2) (1996) 193–208. doi:10.1007/BF00173700.

[50] A. N. Prior, Time and Modality, Clarendon Press, Oxford, 1957.

[51] R. Parikh, The completeness of propositional dynamic logic, in: Mathematical Foundations of Computer Science 1978, Springer, 1978, pp. 403–415.

[52] J. P. Burgess, Basic tense logic, in: D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic volume II, Reidel, 1984, pp. 89–133.

[53] Y. Wang, Q. Cao, On axiomatizations of public announcement logic, Synthese 190 (1) (2013) 103–134. doi:10.1007/s11229-012-0233-5.

[54] J. A. Plaza, Logics of public communications, in: M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, Z. W. Ras (Eds.), Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems, Oak Ridge National Laboratory, Tennessee, USA, 1989, pp. 201–216.

[55] J. Gerbrandy, W. Groeneveld, Reasoning about information change, Journal of Logic, Language, and Information 6 (2) (1997) 147–196. doi:10.1023/A:1008222603071.

[56] H. van Ditmarsch, B. P. Kooi, The secret of my success, Synthese 151 (2) (2006) 201–232. doi:10.1007/s11229-005-3384-9.

[57] W. Holliday, T. Icard, Moorean phenomena in epistemic logic, in: L. Beklemishev, V. Goranko, V. Shehtman (Eds.), Advances in Modal Logic, College Publications, 2010, pp. 178–199.

[58] C. Areces, B. ten Cate, Hybrid logics, in: P. Blackburn, F. Wolter, J. van Benthem (Eds.), Handbook of Modal Logics, Elsevier, 2006, pp. 821–868.

[59] A. Stroll (Ed.), Epistemology, Harper and Rowe, New York, 1967.