Dynamic Epistemic Logics of Introspection

Raul Fervari[1] and Fernando R. Velázquez-Quesada[2(✉)]

[1] FaMAF, Universidad Nacional de Córdoba, and CONICET, Córdoba, Argentina
rfervari@conicet.gov.ar
[2] ILLC, Universiteit van Amsterdam, Amsterdam, The Netherlands
FRVelazquezQuesada@uva.nl

Abstract. This work studies positive and negative introspection not as properties, but rather as actions that change the agent's knowledge. The actions are introduced as model update operations, with matching modalities expressing their effects. Sound and complete axiom systems are provided, and some properties 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]) is that it deals not only with an agent's knowledge about propositional facts, but also with her knowledge about her own (and eventually other agents') knowledge (*high-order* knowledge). This has been the starting point for the study of more complex multi-agent epistemic notions (e.g., common knowledge) that are crucial in multi-agent interaction, thus allowing *EL* to extend its range of applications, including not only philosophy (epistemology [2]), but also computer science (artificial intelligence [3]) and economics (game theory [4]).

In the study of agents with high-order knowledge, two of the most important concepts have been *positive introspection* (if the agent knows 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: to work with full positively introspection, it is enough to consider a transitive indistinguishability relation, and to deal with full negative introspection, it is enough to ask for such relation to be Euclidean. When these properties are not enforced, the agent might lack introspection, thus making her more 'real'. But, as in real life, not being introspective should not imply one will never be.

Recent works have studied properties of an *EL* agent's knowledge from a *dynamic* point of view, thinking about them in terms of the actions the agent can perform to achieve them. For example, closure under logical consequence

A. Madeira and M. Benevides (Eds.): DALI 2017, LNCS 10669, pp. 82–97, 2018.
https://doi.org/10.1007/978-3-319-73579-5_6

can be seen not as a 'static' property, but rather as the eventual result of awareness raising and 'syntactic' inference steps within awareness relational models [5,6], and also as the result of dynamics of evidence or deductive inference within neighbourhood models [7–9]. Following this idea, the present work studies introspection properties by defining epistemic actions that allow a non-introspective agent to reach them. These actions are represented in a *dynamic epistemic logic* (*DEL*; [10,11]) style: as accessibility-changing model operations. There are several examples of such operations in the literature, as the actions for belief revision and/or preference change studied in [12–15] and the logics for reasoning about dynamic policies investigated in [16,17]. There are also the more 'abstract' edge-deleting sabotage operation of [18], the edge-adding and swapping proposals in [19–22] and the general arrow update approach of [23].

The article is organised as follows. Section 2 introduces basic definitions about epistemic logic and propositional dynamic logic. Section 3 defines model operations to achieve positive introspection for general knowledge and also with respect to a formula. Section 4 focuses on similar operations for negative introspection. In all cases we study some properties of the operations, providing also sound and complete axiomatizations for their respective modalities. Finally, Sect. 5 draws conclusions.

2 Basic Definitions

This section recalls not only the basic definitions of basic epistemic logic, but also extensions that will be useful when providing axiom systems for modalities representing the introspection operations. 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 (which is not required to satisfy any property). A* relational model *is a tuple* $M = \langle F, V\rangle$ *with* F *a relational frame and* $V : \mathtt{P} \to \wp(W)$ *an 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.

Next we introduce the *basic epistemic language* $\mathcal{L}_\Diamond$.

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 $p \in \mathtt{P}$. *Other Boolean connectives and constants as well as the modality* $\Box$ *are defined as usual (*$\Box\varphi := \neg\Diamond\neg\varphi$ *for the latter), and formulas of the form* $\Box\varphi$ *are read as* "the agent knows φ"*. For the semantic interpretation, given a relational state* (M, w) *with* $M = \langle W, R, V\rangle$*, formulas in* $\mathcal{L}_\Diamond$ *are interpreted as usual, with the cases of atomic propositions and the 'diamond' modality being*

$(M, w) \Vdash p$ *iff* $w \in V(p)$
$(M, w) \Vdash \Diamond \varphi$ *iff* *there is* $u \in W$ *such that* Rwu *and* $(M, u) \Vdash \varphi$.

A formula φ *is* true *at* w *in* M *when* $(M, w) \Vdash \varphi$. *A formula* φ *is* valid *(notation:* $\Vdash \varphi$*) when it is true in every world* w *of every model* M.

Theorem 2.1 (Axiom System for $\mathcal{L}_{\Diamond}$). *As it is well-known (e.g., [24,25]), axiom schemes and rules on the first block of Table 1 form a sound and strongly complete axiom system (*$\mathsf{L}_{\Diamond}$*) for formulas of* $\mathcal{L}_{\Diamond}$ *w.r.t. relational models.*

Table 1. Axiom systems for $\mathcal{L}_{\Diamond}$ and some of its extensions.

Prop	$\vdash \varphi$ for φ a propositional tautology	*MP*	If $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$
K	$\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	*N*	If $\vdash \varphi$, then $\vdash \Box\varphi$
Dual	$\vdash \Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$		
$K_{\boxplus}$	$\vdash \boxplus(\varphi \to \psi) \to (\boxplus\varphi \to \boxplus\psi)$	$Nec_{\boxplus}$	If $\vdash \varphi$, then $\vdash \boxplus\varphi$
$Dual_{\boxplus}$	$\vdash \boxplus\varphi \leftrightarrow \neg\oplus\neg\varphi$		
$FP_{\boxplus}$	$\vdash \oplus\varphi \leftrightarrow \Diamond(\varphi \vee \oplus\varphi)$	$Ind_{\boxplus}$	$\vdash \boxplus(\varphi \to \Box\varphi) \to (\Box\varphi \to \boxplus\varphi)$
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 $\vdash [\alpha]\varphi$
$Dual_{\alpha}$	$\vdash [\alpha]\varphi \leftrightarrow \neg\langle\alpha\rangle\neg\varphi$	?	$\vdash \langle ?\varphi\rangle\psi \leftrightarrow (\varphi \wedge \psi)$
$\triangleleft_1$	$\vdash \varphi \to [\triangleright]\langle\triangleleft\rangle\varphi$	$\triangleleft_2$	$\vdash \varphi \to [\triangleleft]\langle\triangleright\rangle\varphi$
$\cup$	$\vdash \langle\alpha \cup \beta\rangle\varphi \leftrightarrow (\langle\alpha\rangle\varphi \vee \langle\beta\rangle\varphi)$	;	$\vdash \langle\alpha ; \beta\rangle\varphi \leftrightarrow \langle\alpha\rangle\langle\beta\rangle\varphi$
FP_{*}	$\vdash \langle\alpha^{*}\rangle\varphi \leftrightarrow (\varphi \vee \langle\alpha\rangle\langle\alpha^{*}\rangle\varphi)$	Ind_{*}	$\vdash [\alpha^{*}](\varphi \to [\alpha]\varphi) \to (\varphi \to [\alpha^{*}]\varphi)$

The following sections study languages with modalities for actions of introspection. To introduce their corresponding axiom systems, some extensions of the basic epistemic language will be useful. First, a transitive closure modality.

Definition 2.3 (Language $\mathcal{L}_{\Diamond,\oplus}$). *The language* $\mathcal{L}_{\Diamond,\oplus}$ *adds* $\oplus$ *to* $\mathcal{L}_{\Diamond}$. *Given a relational state* (M, w) *with* $M = \langle W, R, V\rangle$ *and* R^{+} *the transitive closure of* R,

$(M, w) \Vdash \oplus\varphi$ *iff* *there is* $u \in W$ *such that* $R^{+}wu$ *and* $(M, u) \Vdash \varphi$.

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

Theorem 2.2 (Axiom System for $\mathcal{L}_{\Diamond,\oplus}$). *The axioms and rules on the first and second block of Table 1 form sound and weakly complete axiom system (*$\mathsf{L}_{\Diamond,\oplus}$*) for formulas of* $\mathcal{L}_{\Diamond,\oplus}$ *w.r.t. relational models [3].*

Second, the propositional dynamic logic (*PDL*; [26]) framework with a converse modality, with operations for building more complex relations (cf. [27]).

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*

$$\varphi, \psi ::= p \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\alpha\rangle\,\varphi \qquad \alpha, \beta ::= \rhd \mid \lhd \mid \alpha \cup \beta \mid \alpha\,;\beta \mid \alpha^* \mid ?\varphi,$$

with $p \in \mathtt{P}$. The fragment of $\mathcal{L}_{PDL^{\triangleleft,?}}$ without ? is called $\mathcal{L}_{PDL^{\triangleleft}}$. Given (M, w) with $M = \langle W, R, V\rangle$, the semantics of the new modality is defined as

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

with the relation R_α defined inductively as

$$R_\rhd := R,\ R_\lhd := \Re,\ R_{\alpha\cup\beta} := R_\alpha \cup R_\beta,\ R_{\alpha;\beta} := R_\alpha \circ R_\beta,\ R_{\alpha^*} := (R_\alpha)^*,\ R_{?\varphi} := \mathrm{Id}^M_\varphi,$$

where $\Re := \{(v, u) \mid Ruv\}$, $\mathrm{Id}^M_\varphi := \{(u, u) \mid (M, u) \Vdash \varphi)\}$ and $R^* := R^+ \cup \mathrm{Id}^M_\top$.

Theorem 2.3 (Axiom System for $\mathcal{L}_{PDL^{\triangleleft,?}}$). *The axioms and rules on the third block of Table 1 form sound and weakly complete axiom system ($\mathsf{L}_{PDL^{\triangleleft,?}}$) for formulas of $\mathcal{L}_{PDL^{\triangleleft,?}}$ w.r.t. relational models [26, 28, 29]. $\mathsf{L}_{PDL^{\triangleleft}}$ denotes the axiom system for the fragment $\mathcal{L}_{PDL^{\triangleleft}}$, given by $\mathsf{L}_{PDL^{\triangleleft,?}}$ minus axiom ?.*

3 Positive Introspection

3.1 General Positive Introspection

When looking for a model operation for representing an action of positive introspection, the first idea is simple: if transitivity makes the positive introspection axiom $\Box\,\varphi \to \Box\,\Box\,\varphi$ valid, then make the accessibility relation transitive.

Definition 3.1 (General Positive Introspection Operation). *Take a relational model $M = \langle W, R, V\rangle$. The* general positive introspection *operation yields the model $M^+ = \langle W, R^+, V\rangle$.*

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

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

As the model operation is deterministic and its associated modality lacks a precondition, the dual modality $[+]\,\varphi := \neg\,\langle+\rangle\,\neg\varphi$ is equivalent to $\langle+\rangle$ (self-duality).

Some Properties. The operation makes the accessibility relation transitive; then, after applying it, the agent has full positive introspection about any φ.

Proposition 3.1. *Let φ an $\mathcal{L}_{\Diamond,+}$-formula. Then $\Vdash [+]\,(\Box\,\varphi \to \Box\,\Box\,\varphi)$.*

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 3.1. *The formula* $\Box\varphi \to [+](\Box\varphi \wedge \Box\Box\varphi)$ *is not valid, not even for* φ *propositional.*

Proof. Take φ *as* p*. In the relational state below on the left (reflexivity assumed),* $(M, w_1) \Vdash \Box p \wedge \neg\Box\Box p$*. Nevertheless, after the operation (relational state on the right), she does not know* p *anymore:* $(M^+, w_1) \Vdash \neg\Box p$*, i.e.,* $(M, w_1) \Vdash \langle+\rangle\neg\Box p$*. Thus,* $(M, w_1) \Vdash \Box p \wedge \langle+\rangle\neg\Box p$*.*

Making the accessibility relation transitive might increase the worlds reachable in one step. Thus, although the operation makes the agent's knowledge positively introspective, it does not do it by increasing her knowledge; rather, it discards the knowledge that was non-introspective.

Axiom System. When providing an axiom system for a modality representing a model operation, a useful *DEL* 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). 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.

Note how this strategy requires a basic language expressive enough to describe the changes the model operation induces. In this case, $\mathcal{L}_{\Diamond}$ is not enough to deal with the changes the general positive introspection operation brings about: it cannot describe what holds in worlds that can be reached by an *arbitrary* (finite non-zero) number of R-steps (i.e., a single R^+-step). Thus, in order to provide reduction axioms for $\langle+\rangle$, the basic language will be $\mathcal{L}_{\Diamond,\oplus}$.

Table 2. Axioms and rule for the modality $\langle+\rangle$.

$+_p$	$\vdash \langle+\rangle p \leftrightarrow p$	$+_{\Diamond}$	$\vdash \langle+\rangle\Diamond\varphi \leftrightarrow \oplus\langle+\rangle\varphi$
$+_{\neg}$	$\vdash \langle+\rangle\neg\varphi \leftrightarrow \neg\langle+\rangle\varphi$	$+_{\oplus}$	$\vdash \langle+\rangle\oplus\varphi \leftrightarrow \oplus\langle+\rangle\varphi$
$+_{\vee}$	$\vdash \langle+\rangle(\varphi \vee \psi) \leftrightarrow (\langle+\rangle\varphi \vee \langle+\rangle\psi)$	Nec_+	If $\vdash \varphi$, then $\vdash [+]\varphi$
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 occurrences of ψ_1 in φ with ψ_2.		

Theorem 3.2 (Axiom System for $\mathcal{L}_{\Diamond,\oplus,+}$). *The axioms and rules of Table 2, together with* $\mathsf{L}_{\Diamond,\oplus}$ *(first and second blocks in Table 1), form a sound and weakly complete axiom system for formulas of* $\mathcal{L}_{\Diamond,\oplus,+}$ *w.r.t. relational models.*

3.2 Particular Positive Introspection

The operation of Definition 3.1 allows the agent to have positive introspection at the cost of losing knowledge. As such, it does not follow the intuition of what an actual positive introspection reasoning step should do. An operation closer to this intuition would take the agent from knowing χ without knowing she knows it, to knowing χ and knowing she knows it. But then the operation should be radically different. If at (M, w) the agent knows a given χ without having full positive introspection about it, then although every world R-reachable from w *in one step* satisfies χ, there is at least one world R-reachable from w (in two or more steps) where χ fails. In order for the agent to have full positive introspection about χ, such $\neg\chi$-worlds should not be R-reachable anymore. In other words, the operation should not add edges, but rather remove them.

Definition 3.3 (U-disconnecting Operation). *Let $M = \langle W, R, V\rangle$ be a relational model; take $U \subseteq W$. The U-*disconnecting *operation yields the model $M_{+U} = \langle W, R', V\rangle$, with $R' := R \setminus (U \times \overline{U})$ (for $\overline{U} := W \setminus U$). Thus, this operation removes edges from worlds on U to worlds not in U.*

When the parameter U of this model operation is given by the truth-set of a formula χ, then the operation can be understood as a *particular positive χ-introspection operation*: it removes edges from worlds satisfying χ to worlds not satisfying χ. The modality for this operation will be introduced in two stages, the first one being the definition of an auxiliary modality.

Definition 3.4 (Language $\mathcal{L}_{\Diamond,+'\chi}$). *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; use $[\![\chi]\!]^M$ to denote the set of worlds in M in which χ holds.*

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

The operation is deterministic and its modality does not have a precondition, so the modality $[+']$, defined as $[+'\chi]\,\varphi := \neg\,\langle +'\chi\rangle\,\neg\varphi$, is equivalent to $\langle +\rangle$.

This auxiliary modality allows the language to describe the effects of the positive χ-introspection operation. Still, it 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, 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:

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

The reader familiar with *DEL* might notice here a departure from the standard approach: why the auxiliary 'preconditionless' modality $\langle +'\chi\rangle$ instead of defining $\langle +\chi\rangle$ directly with the appropriate precondition? The reason is that the former simplifies the formulation of reduction axioms.

Some Properties. First, here it is a validity characterizing the knowledge of the agent after the operation.

Proposition 3.2. *Let χ and φ be formulas in $\mathcal{L}_{\Diamond,+'\chi}$. The agent can perform a particular positive introspection step for χ after which she will know φ iff she knows both χ and that, after the 'preconditionless' operation, φ will be the case. More precisely,* $\Vdash \langle +\chi\rangle \Box\varphi \leftrightarrow \Box(\chi \wedge [+'\chi]\varphi)$.

Proof. Take any (M,w) with $M = \langle W,R,V\rangle$. From left to right, $(M,w) \Vdash \langle +\chi\rangle \Box\varphi$ yields, 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, $(M_{+\chi},w) \Vdash \Box\varphi$, i.e. $R'wu$ implies $(M_{+\chi},u) \Vdash \varphi$. Take now any $u \in W$ with Rwu: then $(M,u) \Vdash \chi$ and hence, from the definition of R' in $M_{+\chi}$, $R'wu$, so $(M_{+\chi},u) \Vdash \varphi$ and then $(M,u) \Vdash [+'\chi]\varphi$. Thus, Rwu implies $(M,u) \Vdash \chi \wedge [+'\chi]\varphi$; hence, $(M,w) \Vdash \Box(\chi \wedge [+'\chi]\varphi)$. 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 $u \in W$ with $R'wu$: since $R' \subseteq R$, then Rwu and hence $(M,u) \Vdash [+'\chi]\varphi$, i.e., $(M_{+\chi},u) \Vdash \varphi$. Thus, $(M_{+\chi},w) \Vdash \Box\varphi$ and so $(M,w) \Vdash \langle +'\chi\rangle \Box\varphi$. 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 this 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. One might expect for the second conjunct inside the scope of $\Box$ in the right-side, $[+'\chi]\Box\chi$, to collapse to $\top$, so the necessary and sufficient condition for the agent to reach one-level positive χ-introspection is for her to know χ. This is not the case.

Fact 3.3 *The formula $\Box\chi \to [+'\chi]\Box\chi$ is not valid, and so neither is $[+'\chi]\Box\chi$.*

Proof. Take $\chi := p \wedge \Diamond\Diamond\neg p$ and the relational state below on the left (reflexivity assumed); χ holds at w_1 and w_2 (so $(M,w_1) \Vdash \Box\chi$), but fails at w_3. Thus, the operation yields the relational state on the right, with χ false at w_2; then, $(M_{+\chi},w_1) \Vdash \neg\Box\chi$ and hence $(M,w_1) \Vdash \Box\chi \wedge \langle +'\chi\rangle \neg\Box\chi$: the agent knows χ, but she will not know it anymore after a positive χ-introspection action.

Note how $(M,w_1) \Vdash \neg\Box\Box\chi$, so the introspection action is not redundant. Even more, $(M,w_1) \Vdash \Box\chi$, so the state satisfies $\langle +\chi\rangle \neg\Box\chi$ and hence $\neg[+\chi]\Box\chi$.

Fact 3.3 is just one more instance of Moorean phenomena, commonly known as formulas which, after being truthfully announced, become false [30].[1] Here it appears as formulas that are known but, after a particular positive introspection action, are not known anymore. This is because, though the operation does not

[1] The paradigmatic example is $p \wedge \neg\Box p$.

change the atomic valuation, it changes the accessibility relation, thus affecting the agent's knowledge. Nevertheless, the operation behaves as expected when the truth-value of the involved formula χ is preserved by the operation.

Proposition 3.3. *If* $\Vdash \chi \to [+'\chi]\,\chi$, *then after the operation the agent will have positive introspection about* χ, $\Vdash \langle +\chi\rangle \Box\Box\chi \leftrightarrow \Box\chi$.

Proof. The '$\to$*' direction follows by replacing* φ *with* $\Box\chi$ *in the validity of Proposition 3.2. For '*$\leftarrow$*', take any* (M,w) *with* $M = \langle W,R,V\rangle$, *and suppose* $(M,w) \Vdash \Box\chi$; *then* Rwu *implies* $(M,u) \Vdash \chi$. *Now take any* $u \in W$ *with* $R'wu$ *and any* $v \in W$ *with* $R'uv$. *Since* $R' \subseteq R$, *then* Rwu *and hence* $(M,u) \Vdash \chi$. *But* $R'uv$ *so the definition of* R' *yields* $(M,v) \Vdash \chi$. *Then, by the assumption,* $(M,v) \Vdash [+'\chi]\,\chi$, *that is,* $(M_{+\chi},v) \Vdash \chi$. *Since* v *is an arbitrary* R'*-successor of* u, $(M_{+\chi},u) \Vdash \Box\chi$; *since* u *is an arbitrary* R'*-successor of* w, $(M_{+\chi},w) \Vdash \Box\Box\chi$. *Hence,* $(M,w) \Vdash \langle +'\chi\rangle \Box\Box\chi$ *and, as the precondition holds,* $(M,w) \Vdash \langle +\chi\rangle \Box\Box\chi$.

The right-to-left direction of this validity, $\Box\chi \to \langle +\chi\rangle \Box\Box\chi$, is a *dynamic* version of the positive introspection axiom $\Box\chi \to \Box\Box\chi$: the agent might lack positive introspection for χ, but she can achieve it. Even more: under the same requirement for χ, after the action the agent will have *full* positive χ-introspection.

Proposition 3.4. *If* $\Vdash \chi \to [+'\chi]\,\chi$, *then after the operation the agent will have* full *positive introspection about* χ, *that is,* $\Vdash \Box\chi \to \langle +\chi\rangle \Box^n \Box\chi$ *for any* $n \geq 0$, *with* $\Box^0\varphi := \varphi$ *and* $\Box^{k+1}\varphi := \Box^k\Box\varphi$.

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

For $(M,w) \Vdash \langle +\chi\rangle \Box^n \Box\chi$, *take* $n \geq 0$ *and any* $u \in W$ *with* $(R')^{n+1}wu$. *Then* $(M,u) \Vdash \chi$ *and hence, by the assumption,* $(M,u) \Vdash [+'\chi]\,\chi$, *i.e.,* $(M_{+\chi},u) \Vdash \chi$. *Thus,* $(R')^{n+1}wu$ *implies* $(M_{+\chi},u) \Vdash \chi$, *that is,* $(M_{+\chi},w) \Vdash \Box^n\Box\chi$ *so* $(M,w) \Vdash \langle +'\chi\rangle \Box^n\Box\chi$. *But* $\langle +\chi\rangle$ *'s precondition holds; thus,* $(M,w) \Vdash \langle +\chi\rangle \Box^n\Box\chi$.

Thus, if the operation does not affect χ's truth-value, the action's precondition (to know χ) guarantees that the agent will have (knowledge and) full positive introspection about χ. This operation is closer to what comes to mind when one thinks about 'real life': the agent knows χ without noticing it, and thus she only needs to make a further 'introspective' effort to realise it. The operation does not yield positive introspection for all formulas, but it does the work for the particular χ (modulo the extra assumption).

Particular Introspection vs Public Announcement. The reader familiar with *public announcement logic* (*PAL*; [31]) will have noted the similarities between the operation of Definition 3.3 and the public announcement operation: in the new model, former χ-worlds can only reach former χ-worlds. Thus, when the evaluation point is a χ-world, the resulting models are bisimilar. There is, however, an important difference in the precondition of their associated modalities: the one for a χ-announcement requires χ, but the one for a χ-introspection requires $\Box\chi$. This is why, while the public announcement modality has 'straightforward' reduction axioms (there is a match between the precondition and the requirement for a world to be reachable after the operation), the introspection modality requires an auxiliary 'preconditionless' version.

Despite the technical similarities, the two operations represent actions of a very different nature: a public announcement is about external communication, but introspection is about self-reflection. It is then remarkable how, in this setting, their representations are so similar. It could be argued that the presented introspection action is too drastic: it removes any 'eventual' (i.e., possibility of having a possibility) uncertainty the agent might have about the given formula. This is indeed the case, but it is the interpretation of edges in relational models what gives no other choice in order to represent this specific epistemic action.

Axiom System. For an axiom system for the modality $\langle{+}\chi\rangle$, the first step is to provide reduction axioms for its 'preconditionless' counterpart.

Theorem 3.4 (Axiom System for $\mathcal{L}_{\Diamond,+'\chi}$). *The axioms and rules of Table 3, together with the axiom system* $\mathsf{L}_{\Diamond}$ *(see Table 1), form a sound and strongly complete axiom system for formulas of* $\mathcal{L}_{\Diamond,+'\chi}$ *w.r.t. relational models.*

Table 3. Axioms and rule for the modality $+'\chi$.

$+'\chi_p$ $\vdash \langle{+}p\rangle \leftrightarrow p$	$+'\chi_\vee$ $\vdash \langle{+'}\chi\rangle(\varphi \vee \psi) \leftrightarrow \langle{+'}\chi\rangle\varphi \vee \langle{+'}\chi\rangle\psi$
$+'\chi_\neg$ $\vdash \langle{+'}\chi\rangle\neg\varphi \leftrightarrow \neg\langle{+'}\chi\rangle\varphi$	$+'\chi_\Diamond$ $\vdash \langle{+'}\chi\rangle\Diamond\varphi \leftrightarrow (\neg\chi \wedge \Diamond\langle{+'}\chi\rangle\varphi) \vee \Diamond(\chi \wedge \langle{+'}\chi\rangle\varphi)$
$N_{+'\chi}$ If $\vdash \varphi$, then $\vdash [{+'}\chi]\varphi$	
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 (occurrences of ψ_1 which are *not* inside any 'dynamic' modality $\langle{+'}\chi\rangle$.) with ψ_2.	

The previous theorem provides a sound and strongly complete axiom system for $\langle{+'}\chi\rangle$. As $\langle{+}\chi\rangle$ is just an abbreviation, it requires no axioms; still, its definition makes $\langle{+}\chi\rangle\varphi \leftrightarrow (\Box\chi \wedge \langle{+'}\chi\rangle\varphi)$ valid.

4 Negative Introspection

4.1 General Negative Introspection

Analogous to its positive introspection counterpart, the operation for achieving full negative introspection is simply an Euclidean closure operation.

Definition 4.1 (General Negative Introspection Operation). *Take a relational model $M = \langle W, R, V\rangle$. The* general negative introspection *operation yields the model $M^- = \langle W, R^E, V\rangle$ in which R^E is the Euclidean closure of R, that is,*

$$R^E := R \cup (\Re \circ (R \cup \Re)^* \circ R).$$

Definition 4.2 (Language $\mathcal{L}_{\Diamond,-}$). *The language $\mathcal{L}_{\Diamond,-}$ extends $\mathcal{L}_\Diamond$ with $\langle - \rangle$ ($[-]$ defined as usual). For its semantic interpretation, let (M, w) be a relational state.*

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

Clearly, R^E can be equivalently defined in *PDL* plus the converse operator. This suggests that $\mathcal{L}_{PDL^\triangleleft}$ from Definition 2.4 will be useful to provide reduction axioms for this operation. But first, here are some of its properties.

Some Properties. Since R^E is indeed R's Euclidean closure, after the operation the agent has negative introspection.

Lemma 4.1. *For any $R \subseteq (W \times W)$, the relation $R^E := R \cup (\Re \circ (R \cup \Re)^* \circ R)$ is R's Euclidean closure, i.e., the smallest Euclidean relation containing R.*[2]

Proposition 4.1. *Let φ an $\mathcal{L}_{\Diamond,-}$-formula. Then, $\Vdash [-](\neg \Box \varphi \to \Box \neg \Box \varphi)$.*

Even more. Different from the positive introspection case, in the propositional case the operation makes the agent's knowledge negatively introspective in the sense of taking her from $\neg \Box \varphi \wedge \neg \Box \neg \Box \varphi$ to $\neg \Box \varphi \wedge \Box \neg \Box \varphi$.

Proposition 4.2. *If φ is propositional, then $\Vdash \neg \Box \varphi \to [-](\neg \Box \varphi \wedge \Box \neg \Box \varphi)$.*

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

This validity, a *dynamic* version of the negative introspection axiom $\neg \Box \varphi \to \Box \neg \Box \varphi$, shows how the operation behaves properly in the propositional case. Still, as expected, it also has Moorean behaviour for arbitrary formulas.

Fact 4.1 *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) below on the left (reflexivity assumed). Note how $(M, w_1) \Vdash \Diamond \Box p$, i.e., $(M, w_1) \Vdash \neg \Box (\neg \Box p)$. The operation produces the relational state on the right, where $(M^-, w_1) \Vdash \Diamond \Box \Diamond \neg p$, i.e., $(M^-, w_1) \Vdash \neg \Box \neg \Box (\neg \Box p)$ so $(M, w_1) \Vdash \langle - \rangle \neg \Box \neg \Box (\neg \Box p)$.

[2] Proof: http://homepages.cwi.nl/~jve/courses/lai0506/Solutions2.pdf.

Axiom System. $\mathcal{L}_{\Diamond}$ is not expressive enough to describe the effects of this operation, but the clearly more expressive $\mathcal{L}_{PDL^{\triangleleft}}$ is. The Boolean cases are as those in Tables 2 and 3; the modal case is different: in $\langle\alpha\rangle\,\varphi$, the expression α is an arbitrary program expression, and thus an appropriate translation in each case must be presented. The *program transformer* defined below, a simplification of that of [32] for providing reduction axioms for *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 (Proposition 4.3).

Definition 4.3 (Program Transformer). *A* program transformer T *is a function from program expressions to program expressions defined inductively as*

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

Proposition 4.3. *Let $M = \langle W, R, V\rangle$ be any relational model, and recall that $M^{-} = \langle W, R^{E}, V\rangle$. Then, for every program expression α, $R^{E}_{\alpha} = R_{\mathrm{T}(\alpha)}$.*

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

- $R^{E}_{\triangleright} = R^{E} = R \cup (\Re \circ (R \cup \Re)^{*} \circ R) = R_{\triangleright} \cup (R_{\triangleleft} \circ (R_{\triangleright} \cup R_{\triangleleft})^{*} \circ R_{\triangleright})$
  $= R_{\triangleright} \cup (R_{\triangleleft} \circ (R_{\triangleright\cup\triangleleft})^{*} \circ R_{\triangleright}) = R_{\triangleright} \cup (R_{\triangleleft} \circ R_{(\triangleright\cup\triangleleft)^{*}} \circ R_{\triangleright})$
  $= R_{\triangleright} \cup R_{\triangleleft;(\triangleright\cup\triangleleft)^{*};\triangleright} = R_{\triangleright\cup(\triangleleft;(\triangleright\cup\triangleleft)^{*};\triangleright)} = R_{\mathrm{T}(\triangleright)}$

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

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

Theorem 4.2 (Axiom System for $\mathcal{L}_{PDL^{\triangleleft},-}$). *The axioms and rules of Table 4, together with the axiom system $\mathsf{L}_{PDL^{\triangleleft}}$ (see Table 1), form a sound and weakly complete axiom system for formulas of $\mathcal{L}_{PDL^{\triangleleft},-}$ w.r.t. relational models.*

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

$-_{p}$	$\vdash \langle -\rangle\, p \leftrightarrow p$	$-_{\langle\alpha\rangle}$	$\vdash \langle -\rangle\,\langle\alpha\rangle\,\varphi \leftrightarrow \langle \mathrm{T}(\alpha)\rangle\,\langle -\rangle\,\varphi$
$-_{\neg}$	$\vdash \langle -\rangle\,\neg\varphi \leftrightarrow \neg\langle -\rangle\,\varphi$	Nec_{-}	If $\vdash \varphi$, then $\vdash [-]\,\varphi$
$-_{\vee}$	$\vdash \langle -\rangle\,(\varphi \vee \psi) \leftrightarrow (\langle -\rangle\,\varphi \vee \langle -\rangle\,\psi)$	*SE*	As in Table 2

4.2 Particular Negative Introspection

Different from the positive introspection counterpart, the operation of Definition 4.1 already behaves as expected: it preserves the agent's (propositional) lack of knowledge while giving her negative introspection (Proposition 4.2). Still, for uniformity, this section explores a negative introspection action over a given χ.

A model operation for achieving full negative introspection about χ should then 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 4.1 in which the new edges point only to $\neg\chi$-worlds.

Definition 4.4 (U-connecting Operation). *Let $M = \langle W, R, V\rangle$ be a relational model; let $U \subseteq W$. The U-*connecting *operation gives the model $M_{-U} = \langle W, R', V\rangle$, with its indistinguishability relation R' given (with $\mathrm{Id}_U^M := \{(u,u) \mid u \in U\}$) by*

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

A modality for a particular full negative introspection can be defined by instantiating U with the set of worlds satisfying $\neg\chi$ in the original model. Here is a 'preconditionless' version.

Definition 4.5 (Language $\mathcal{L}_{\Diamond,-'\chi}$). *The language $\mathcal{L}_{\Diamond,-'\chi}$ extends $\mathcal{L}_\Diamond$ with a modality $\langle -'\chi\rangle$ for each formula χ. For the semantic interpretation,*

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

A modality with an appropriate precondition is defined in the obvious way:

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

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

Some Properties. As expected, the analogous of Proposition 3.4 holds.

Proposition 4.4. *If $\Vdash \chi \to [-'\chi]\, \chi$, then after the operation the agent will have* full *negative introspection about χ, $\Vdash \neg \Box \chi \to \langle -\chi\rangle\, \Box^n \neg \Box \chi$ for any $n \geq 0$.*

Proof. Take a relational state (M,w) with $M = \langle W, R, V\rangle$, 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 $(M_{-\chi}, v) \Vdash \neg\chi$ (by the assumption). The first step is to show (by induction on $n \geq 0$) how, in $M_{-\chi}$, any world that can be reached from w in zero or more steps can also reach v, that is, how $(R')^n wu$ implies $R'uv$. The base case ($n = 0$, i.e., $u = w$) is immediate, as the supposition states Rwv, and thus $R'wv$. For the inductive case, 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 $R'u'u$. It is not hard to see that, in each of the four cases the definition of R' yields, $R'uv$.

Now, in order to prove $(M, w) \Vdash \langle -\chi \rangle \Box^n \neg \Box \chi$, take any $n \geq 0$ and any $u \in W$ such that $(R')^n wu$. Then $R'uv$ and, from $(M_{-\chi}, v) \Vdash \neg\chi$, it follows that $(M_{-\chi}, u) \Vdash \Diamond \neg\chi$, that is, $(M_{-\chi}, w) \Vdash \Box^n \neg \Box \chi$ so $(M, w) \Vdash \langle -'\chi \rangle \Box^n \neg \Box \chi$. But $\langle -\chi \rangle$'s precondition holds; thus, $(M, w) \Vdash \langle -\chi \rangle \Box^n \neg \Box \chi$, as required.

Note how both negative introspection operations add edges. This differs from the positive introspection case: the general operation adds edges, but the particular one needs to remove them.

Axiom System. The basic language will be now $\mathcal{L}_{PDL^{\triangleleft,?}}$ (Definition 2.4), as the 'test' operator ? is required. Thus, $\mathcal{L}_{PDL^{\triangleleft,?,-'\chi}}$ extends $\mathcal{L}_{PDL^{\triangleleft,?}}$ with the 'dynamic' negative χ-introspection modality; for reduction axioms, the program transformer of Definition 4.3 is redefined in the following way.

Definition 4.6 (Program Transformer). *A χ-program transformer T_χ is a function from program expressions to program expressions defined as follows*

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

The remaining cases (for $\cup$, ; and *) are as in Definition 4.3.

Proposition 4.5. *Let $M = \langle W, R, V\rangle$ be any relational model, and recall that $M_{-\chi} = \langle W, R', V\rangle$. Then, for every program expression α, $R'_\alpha = R_{\mathrm{T}_\chi(\alpha)}$.*

Proof. As in Proposition 4.3, the proof is by structural induction on α. The common cases are similar; for the 'test',

- $R'_{?\varphi} = \{(w, w) \mid (M_{-\chi}, w) \Vdash \varphi\} = \{(w, w) \mid (M, w) \Vdash \langle -'\chi\rangle\,\varphi\} = R_{?\langle -'\chi\rangle\,\varphi} = R_{\mathrm{T}_\chi(?\varphi)}$.

Theorem 4.3 (Axiom System for $\mathcal{L}_{PDL^{\triangleleft},-}$). *The axioms and rules of Table 5, together with the axiom system $\mathsf{L}_{PDL^{\triangleleft,?}}$ (see Table 1) form a sound and weakly complete axiom system for formulas of $\mathcal{L}_{PDL^{\triangleleft,?,-'\chi}}$ w.r.t. relational models.*

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

$-'\chi_p$	$\vdash \langle -'\chi\rangle\, p \leftrightarrow p$	$-'\chi_{\langle\alpha\rangle}$	$\vdash \langle -'\chi\rangle\,\langle\alpha\rangle\,\varphi \leftrightarrow \langle \mathrm{T}_\chi(\alpha)\rangle\,\langle -'\chi\rangle\,\varphi$
$-'\chi_\neg$	$\vdash \langle -'\chi\rangle\,\neg\varphi \leftrightarrow \neg\langle -'\chi\rangle\,\varphi$	$Nec_{-'\chi}$	If $\vdash \varphi$, then $\vdash [-'\chi]\,\varphi$
$-'\chi_\vee$	$\vdash \langle -'\chi\rangle\,(\varphi \vee \psi) \leftrightarrow (\langle -'\chi\rangle\,\varphi \vee \langle -'\chi\rangle\,\psi)$	SE"	Analogous to SE' in Table 3

With the language extended and the axiom system introduced, it is possible to provide further validities describing the behaviour of the operation. First, here is how the operation affects the agent's knowledge (now described by $[\triangleright]$).

Proposition 4.6. *Suppose* χ *and* φ *are both formulas in* $\mathcal{L}_{PDL^{\lhd,?,-'\chi}}$*; then,* $\Vdash \langle -\chi \rangle [\rhd] \varphi \leftrightarrow (\neg [\rhd] \chi \wedge [\mathrm{T}_\chi(\rhd)] [-'\chi] \varphi)$.

From this and the axiom system, one can obtain a validity characterising the requirements for the agent to have negative introspection about a given χ after the operation: $\Vdash \langle -\chi \rangle [\rhd] \neg [\rhd] \chi \ \leftrightarrow \ (\neg [\rhd] \chi \wedge [\mathrm{T}_\chi(\rhd)] [-'\chi] \neg [\rhd] \chi)$.

5 Conclusion and Further Work

This paper studies positive and negative introspection as epistemic actions that modify the agent's knowledge. In both cases two possibilities are considered: a general operation, and a particular one working relative to a given formula. In all cases, the basic epistemic language is extended with modalities representing the effects of the model operations, presenting their sound and complete axiom systems, and exploring some properties of the new languages.

In the case of positive introspection, the general operation follows a straightforward idea: make the accessibility relation transitive. Yet, this approach boils down to assume that introspection fails not because of what the agent knows about her knowledge, but rather because of what she knows; thus, as a result, non-introspective knowledge is lost, and only the introspective one is preserved. The particular operation has the opposite perspective: to get positive introspection about a given χ, it eliminates edges from χ-worlds to $\neg\chi$-worlds, thus forcing positive introspection on χ while keeping the rest of her knowledge 'as before'. For the negative introspection case, the general operation makes the accessibility relation Euclidean, and thus reaches negative introspection by ensuring the agent knows what she does not know. The particular operation follows the same idea while adding only edges pointing to $\neg\chi$-worlds. Both cases about edge-addition; thus, they have a similar behaviour.

For future work, one direction is to explore operations that raise the agent's introspection in just one level (e.g., from $\Box p \wedge \neg \Box \Box p$ to $\Box p \wedge \Box \Box p \wedge \neg \Box \Box \Box p$). A more interesting project is to investigate similar operations in a multi-agent setting (e.g., public, private versions of these operations), focusing also on operations for reaching common knowledge.

Acknowledgements. This work was partially supported by grant ANPCyT-PICT-2013-2011, STIC-AmSud "Foundations of Graph Structured Data (FoG)", SeCyT-UNC, the Laboratoire International Associé "INFINIS", and the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 690974 for the project MIREL: MIning and REasoning with Legal texts.

References

1. Hintikka, J.: Knowledge and Belief. Cornell University Press, Ithaca (1962)
2. Hendricks, V.F.: 8 bridges between formal and mainstream epistemology. Philos. Stud. **128**(1), 1–5 (2006)

3. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning About Knowledge. MIT Press, Cambridge (1995)
4. de Bruin, B.: Explaining Games: The Epistemic Programme in Game Theory. Synthese Library, vol. 346. Springer, Dordrecht (2010). https://doi.org/10.1007/978-1-4020-9906-9
5. van Ditmarsch, H., French, T.: Semantics for knowledge and change of awareness. J. Logic Lang. Inf. **23**(2), 169–195 (2014)
6. Grossi, D., Velázquez-Quesada, F.R.: Syntactic awareness in logical dynamics. Synthese **192**(12), 4071–4105 (2015)
7. van Benthem, J., Pacuit, E.: Dynamic logics of evidence-based beliefs. Stud. Logica **99**(1), 61–92 (2011)
8. Velázquez-Quesada, F.R.: Explicit and implicit knowledge in neighbourhood models. In: Grossi, D., Roy, O., Huang, H. (eds.) LORI 2013. LNCS, vol. 8196, pp. 239–252. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40948-6_19
9. Balbiani, P., Fernández-Duque, D., Lorini, E.: A logical theory of belief dynamics for resource-bounded agents. In: Jonker, C.M., Marsella, S., Thangarajah, J., Tuyls, K. (eds.) Proceedings AAMAS 2016, pp. 644–652. ACM (2016)
10. van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer, Dordrecht (2008). https://doi.org/10.1007/978-1-4020-5839-4
11. van Benthem, J.: Logical Dynamics of Information and Interaction. CUP, New York (2011)
12. van Benthem, J.: Dynamic logic for belief revision. J. Appl. Non-Class. Logics **17**(2), 129–155 (2007)
13. van Benthem, J., Liu, F.: Dynamic logic of preference upgrade. J. Appl. Non-Class. Logics **17**(2), 157–182 (2007)
14. Ghosh, S., Velázquez-Quesada, F.R.: Agreeing to agree: reaching unanimity via preference dynamics based on reliable agents. In: Weiss, G., Yolum, P., Bordini, R.H., Elkind, E. (eds.) Proceedings AAMAS 2015, pp. 1491–1499. ACM (2015)
15. Ghosh, S., Velázquez-Quesada, F.R.: A note on reliability-based preference dynamics. In: van der Hoek, W., Holliday, W.H., Wang, W. (eds.) LORI 2015. LNCS, vol. 9394, pp. 129–142. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48561-3_11
16. Pucella, R., Weissman, V.: Reasoning about dynamic policies. In: Walukiewicz, I. (ed.) FoSSaCS 2004. LNCS, vol. 2987, pp. 453–467. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24727-2_32
17. Göller, S.: On the complexity of reasoning about dynamic policies. In: Duparc, J., Henzinger, T.A. (eds.) CSL 2007. LNCS, vol. 4646, pp. 358–373. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74915-8_28
18. Benthem, J.: An essay on sabotage and obstruction. In: Hutter, D., Stephan, W. (eds.) Mechanizing Mathematical Reasoning. LNCS (LNAI), vol. 2605, pp. 268–276. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-32254-2_16
19. Areces, C., Fervari, R., Hoffmann, G.: Moving arrows and four model checking results. In: Ong, L., de Queiroz, R. (eds.) WoLLIC 2012. LNCS, vol. 7456, pp. 142–153. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32621-9_11
20. Areces, C., Fervari, R., Hoffmann, G.: Swap logic. Logic J. IGPL **22**(2), 309–332 (2014)
21. Fervari, R.: Relation-Changing Modal Logics. Ph.D. thesis, Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Argentina (2014)

22. Areces, C., Fervari, R., Hoffmann, G.: Relation-changing modal operators. Logic J. IGPL **23**(4), 601–627 (2015)
23. Kooi, B., Renne, B.: Arrow update logic. Rev. Symb. Logic **4**, 536–559 (2011)
24. Chellas, B.F.: Modal Logic: An Introduction. Cambridge University Press, Cambridge (1980)
25. Blackburn, P., de Rijke, M., Venema, Y.: Modal logic. CUP, New York (2001)
26. Harel, D., Kozen, D., Tiuryn, J.: Dynamic Logic. MIT Press, Cambridge (2000)
27. van Eijck, J., Wang, Y.: Propositional dynamic logic as a logic of belief revision. In: Hodges, W., de Queiroz, R. (eds.) WoLLIC 2008. LNCS (LNAI), vol. 5110, pp. 136–148. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69937-8_13
28. Prior, A.N.: Time and Modality. Clarendon Press, Oxford (1957)
29. Parikh, R.: The completeness of propositional dynamic logic. In: Winkowski, J. (ed.) MFCS 1978. LNCS, vol. 64, pp. 403–415. Springer, Heidelberg (1978). https://doi.org/10.1007/3-540-08921-7_88
30. Holliday, W., Icard, T.: Moorean phenomena in epistemic logic. In: Beklemishev, L., Goranko, V., Shehtman, V. (eds.) Advances in Modal Logic, College Publications, pp. 178–199 (2010)
31. Plaza, J.A.: Logics of public communications. In: Emrich, M.L., Pfeifer, M.S., Hadzikadic, M., Ras, Z.W. (eds.) Proceedings 4th International Symposium on Methodologies for Intelligent Systems, Oak Ridge National Laboratory, pp. 201–216 (1989)
32. van Benthem, J., van Eijck, J., Kooi, B.: Logics of communication and change. Inf. Comput. **204**(11), 1620–1662 (2006)