The Romanian Journal of Analytic Philosophy
Vol. VII, 1°, 2013, pp. 42-51

A LEWISIAN ARGUMENT AGAINST USING EQUIVALENCE ACCESSIBILITY RELATIONS IN THE LEARNING BY ERASING FRAMEWORK

Alexandru DRAGOMIR*

Abstract. The first part of this paper will present (1) the concept of knowledge as it is used in epistemic logic, (2) Public Announcement Logic's concept of learning as elimination of epistemic alternatives, and (3) a game-theoretical perspective on scientific discovery (research as a game between Nature and Scientist) together with Nina Gierasimczuk's established connection between (2) and (3). The second part will be concerned with arguing that the accessibility relations between hypotheses within the framework of „learning by erasing" should not be equivalence relations.

Keywords: epistemic logic, public announcement logic, learning by erasing.

I. INTRODUCTION

Epistemic logic (hereafter EL) is widely used as a tool for reasoning about knowledge (see Fagin, R. *et al.* 1995 for an introduction). Its semantics is based on epistemic models composed of a set of possible worlds (epistemic alternatives), an equivalence accessibility relation that links possible worlds (also called „the epistemic indistinguishability relation"). However, EL lacks the technical means to describe the evolution of an agent's knowledge after receiving truthful information. However, this task has been accomplished by dynamic epistemic logics (see van Ditmarsch, H.P., B. Kooi., W. van der Hoek 2006). A particular dynamic epistemic logic, Public Announcement Logic (hereafter PAL), discovered by Plaza (1989) can describe an agent's epistemic state (what the agent knows) after learning different truths. The

* Alexandru DRAGOMIR, Faculty of Philosophy, University of Bucharest.
Email: alexdragomirs@gmail.com.

effect of learning a truth, say ϕ, is computed by eliminating all the alternatives to ϕ from the model, meaning all possible worlds that do not satisfy ϕ. PAL has also been used by Gierasimczuk (2009a, 2009b, 2010) to offer a model for scientific inquiry in the „learning by erasing" framework. All the logics and frameworks introduced above will be presented in separate sections. We will argue that the accessibility relations used to link epistemic alternatives should not be equivalence relations (meaning reflexive, symmetric and transitive relations). In order to arrive at this conclusion, we will argue that the accessibility relation is not always symmetrical: there are cases in which a world u is accessible from w, but w is not accessible from u. These cases appear if we follow David Lewis' (see Lewis 1979) intuition that conceivability of other worlds from world w depends on the scientific background in w.

II. THE NOTION OF KNOWLEDGE IN EPISTEMIC LOGIC AND LEARNING AS ELIMINATION OF ALTERNATIVES

We will begin by defining the language of epistemic logic, hereafter L_{EL}. For p in P, a countable set of propositional variables, and a in A, a finite set of agents, L_{EL} is given by the following rules:

$$\phi ::= p \mid \neg\phi \mid \phi \,\&\, \phi \mid K_a\phi$$

Formula $K_a\,\phi$ reads agent a knows that ϕ. Consequently, the formula $K_a(\phi\ \&\ \neg\psi)\ \&\ \neg K_a\chi$ will be part of L_{EL} and will be read „agent a knows that ϕ and not ψ, and agent a does not know that χ. A typical model for propositional epistemic logic comprises a domain of possible worlds, an equivalence[1] accessibility relation that links worlds and a valuation function that tells us what propositions are true at each world[2]: $M=(W, R, V)$, $R_a \subseteq W \times W$, $V: P \to 2^W$, for P a countable set of propositional variables.

The fact that the accessibility relation is an equivalence relation ensures certain intuitive properties of knowledge:

(1) Facticity: $K_a\phi \to \phi$ (you know only truths),

(2) Positive introspection: $K_a\phi \to K_aK_a\phi$ (if you know something, you know that you know it),

(3) Negative introspection: $\neg K_a\phi \to K_a\neg K_a\phi$ (if you do not know something, you know that you don't know it).

[1] An equivalence relation is any relation that is reflexive, symmetric and transitive.

[2] We will denote the powerset of W by 2^W.

These properties along with the technically motivated distribution of knowledge over implication: $K_a(\phi \rightarrow \psi) \rightarrow (K_a\phi \rightarrow K_a\psi)$ compose the axioms of the *S5* system of epistemic logic. *S5* is the logical tool of choice of many philosophers, economists and computer scientists, despite notable exceptions: Jaakko Hintikka prefers *S4*, an epistemic logic without the axiom of negative introspection, as the most suitable logic of knowledge.

The semantics of the knowledge operator is based on the intuition that someone knows that ϕ if and only if she cannot conceive an alternative to it. In formal terms, ϕ is known by agent *a* iff in each accessible possible world (or epistemic alternative) it is true that ϕ:

$$M, w \vDash K_a\phi \text{ iff } \forall u: wRu \Longrightarrow M, u \vDash \phi$$

To illustrate the philosophical intuition and the formal definition of epistemic logic's concept of knowledge, consider the following situation. Suppose we have two possible states of affairs: *w*, at which it is true that it is raining in Bucharest and *u* at which it is sunny in Bucharest. Suppose the actual state of affairs is *w*: actually it is raining in Bucharest, and let ϕ denote „it is raining in Bucharest". The epistemic model *M* that captures this situation contains two possible worlds, *w* and *u*, and the valuation function *V* assigns *w* to ϕ: $V(\phi)=\{w\}$. Now, John lives in London and has no information about the weather in Bucharest. That is, even if he inhabits a world in which it rains in Bucharest, *w*, he considers both *w* and *u* as candidates for describing the actual world. The fact that he cannot distinguish between these two worlds is modeled by linking *w* and *u* by an equivalence accessibility relation: $\{(w, w), (w, u), (u, w), (u, u)\} \in R$. In this model, because at *u* it is false that ϕ and John cannot distinguish between *w* and *u* (*u* is accessible from *w*), at *w*, meaning in the actual world, John does not know that ϕ: $M, w \vDash \neg K_a \phi$.

How can one agent come to know that ϕ? How can we represent learning in this semantic framework? An agent can achieve knowledge (in the sense of EL) of ϕ if the model is transformed in such way that:

(1) In each of her accessible worlds ϕ becomes true by changing the valuation of the model, or:

(2) All possible worlds in which ϕ is false are eliminated from the domain.

We will take the latter route.[3] For John to acquire knowledge of ϕ at *w*, he would have to eliminate all doubt of it, meaning all worlds connected to *w* in which ϕ is false. This would happen if he would receive the information

[3] Logics in which the assignment function *V* is dynamic can be found in van Ditmarsch, van der Hoek and Kooi 2005, and van Ditmarsch and Kooi 2008.

that it is raining in Bucharest from a trusted source, like a friend who lives in Bucharest and knows what the weather is like. Then, he will exclude *u* from the set of possible worlds. Note that if a model contains a single world, for example *w*, then, for all formulas ϕ true at w, it holds that $K_a\phi$ at *w*, for each agent *a*. In other words, everything would be known, since no epistemic alternative would be possible.

Although (static) epistemic logic does not have the technical means to modify the domain, dynamic epistemic logic (in particular PAL) has the formal instrument that allows eliminating states that do not satisfy given formulas: the public announcements of formulas.

II.1. PUBLIC ANNOUNCEMENT LOGIC

PAL, discovered by Jan Plaza (1989), is used to describe and predict the evolution of different notions of knowledge as a result of the epistemic interaction between agents in a group. After a public announcement of ϕ, the model changes so as to contain only ϕ–satisfying worlds. PAL contains a double-argument operator that reads: after every public announcement of ϕ it is true that ψ. Its semantics is as follows:

$$M,w \vDash [!\phi]\psi \text{ iff } M,w \vDash \phi \Longrightarrow M!\phi,w \vDash \psi$$

Where $M!\phi = (W', R', V')$, the model updated with ϕ, is the following model:

(1) $W'=\{w \mid M,w \vDash \phi\}$

(2) $R'=R \cap (W' \times W')$

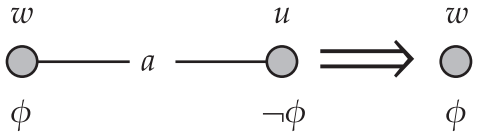
(3) $V'=V \mid W'$ (the restriction of function V to the set W')

The model is transformed so as to contain only worlds satisfying ϕ, and the accessibility relation and valuation are modified so as to range over the new domain. The dual of $[!\,\phi]\,\psi$ is $<!\,\phi>\psi$, and is read: „after a public announcement of ϕ it becomes true that ψ“:

$$M, w \vDash <!\,\phi>\psi \text{ iff } M, w \vDash \phi \text{ and } M!\phi, w \vDash \psi$$

Recall the example given above; the picture below is a graphical representation of John's epistemic state before and after the announcement of ϕ (the elimination of the world at which it is false that ϕ)[4]:

[4] We have omitted the reflexive arrows, but keep in mind that for each possible world *w*, $(w, w) \in R$.



In the updated model, the one on the right-hand side of the graphic, denoted by $M!\phi$, there is only one world, w, and the accessibility relation links w to w as a consequence of its reflexivity. Note that $M, w \vDash <!\phi> K_a\phi$ (after an announcement of ϕ, a knows that ϕ) because $M, w \vDash \phi$ and $M!\phi, w \vDash K_a\phi$, because in $M!\phi=(W', R', V')$: $W'=\{w\}$, $R'ww$, and $M!\phi, w \vDash \phi$.

PAL is reducible to EL through a translation schema (a set of reduction axioms) that transforms each PAL-formula into an EL-formula. The use of reduction axioms takes care of the problems of PAL's completeness and expressivity; for complete proofs, see Kooi 2007, and Solecki *et al.* 1999.

II.2. RESEARCH AS A GAME BETWEEN NATURE AND SCIENTIST

In this section we will present „learning by erasing", a game-theoretical perspective on scientific inquiry (see Martin and Osherson 2002), and Nina Gierasimczuk's account of „learning by erasing" in dynamic epistemic logic (see 2009a, 2009b, 2010). Scientific inquiry can be seen as a guessing game between two agents (see Martin and Osherson 2002; Gierasimczuk 2009a; 2009b; 2010): Nature and Scientist. Nature makes the first move by selecting a possible world from the domain, and in effect making it the actual state of affairs. Scientist tries to deliver the one hypothesis that describes the world she lives in – the actual one. In other words, in terms of this game, she tries to find out what possible world Nature actualized. The fact that she does not know what the actual state of affairs is, is modeled by her having epistemic alternatives to the actual world, or in other words, by her having a set of concurrent hypotheses. But how will Scientist establish which world is actual, or which hypothesis is correct? The game continues in turns: Nature makes the first step by sending a formula (or set of formulas) true in the actual world and, in turn, Scientist tries to figure out what the actual world is by restricting her domain of alternatives to the set of possible worlds that are consistent with the received formula (or set of formulas). This way of viewing the process of scientific inquiry seems intuitive: as a result of her research (conducting experiments, for example), Scientist will receive new information that may be consistent with her hypotheses or may refute them. Also, to simplify the game, we may consider Nature sending a sequence of formulas $!\Phi := !\phi_1!\phi_2!\phi_3 \ldots !\phi_n$ all at once. Scientist will use each formula of the sequence to test her hypotheses. If the formula is not verified by a hypothesis, then the hypothesis must be wrong and be eliminated from the domain and the process will continue by selecting another formula from

How is this testing or verifying procedure accounted for in the formal apparatus presented above? Nina Gierasimczuk argues for using public announcements of the formulas in !Φ (2009a; 2009b; 2010). We have seen that announcing formula ϕ restricts the model to only ϕ-satisfying states. Testing a hypothesis with a formula ϕ results in announcing formula ϕ in the model and failing the test is represented by the elimination of that world (as a result of the announcement made). Note that no sequence of announcements will eliminate the actual world, since the announced formulas are true in it.

Here is an example. Consider three possible worlds:

(1) w, such that $w \vDash \phi_1$, $w \vDash \phi_2$

(2) u, such that $u \vDash \phi_1$, $u \vDash \phi_3$

(3) v, such that $v \vDash \phi_1$, $v \vDash \phi_2$, $v \vDash \phi_4$

Suppose Nature selects v and Scientist receives the sequence $!\Phi=!\phi_1!\phi_2!\phi_4$. After successively announcing each formula, only v remains, and Scientist now knows what the actual world is. After announcing ϕ_1, none of the three are eliminated since all of them satisfy ϕ_1. After the announcement of ϕ_2 only w and v remain in the domain, and $!\phi_4$ will keep only v in the model. So, $!\Phi=!\phi_1!\phi_2!\phi_4$ is a sequence that eliminates all uncertainty towards v, as $!\Phi'=!\phi_1!\phi_4$ and $!\Phi''=!\phi_2!\phi_4$ would.

III. A PHILOSOPHICAL PROBLEM WITHIN THE FRAMEWORK

In this section we will raise a philosophically driven problem within the „learning by erasing" framework. This problem will stem from David Lewis' intuition that conceivability of other possible worlds depends on the scientific background of the actual world.

III.1. DAVID LEWIS ON POSSIBILITY

For Lewis, worlds are not mathematical entities (what he calls „ersatz worlds"), but real entities, like our world, the only difference being that their „inhabitants" call their worlds actual, exactly the way we call our and only our world actual and all other worlds, possible. Some propositions may hold at certain possible worlds, but not at others. David Lewis (1979) held that logic and arithmetic are the same in every possible world, that is to say mathematical and logical truths are necessary truths. However, physics is contingent, that is to say some physical laws or principles may not be true in some worlds.

What of the accessibility relation? There are worlds that are not accessible to ours. Some of them because we could not cover the whole set of logically

possible worlds[5] and some of them because our conceivability depends on our own scientific background. As Lewis argues: „If we knew only the physics of 1871, we could fail to cover some of the possibilities that we recognize today. Perhaps we fail today to cover possibilities that will be recognized in 2071." (see Lewis 1979, p. 189). A consequence of Lewis' tenet is that an increase in scientific knowledge leads to conceiving of a larger set of possible worlds. In other words, worlds that were not or could not have been conceived become possible and, implicitly, accessible in the sense of being linked by the accessibility relation of our models for epistemic logic from our actual world.

One could remark that the two views, Lewis' and the „learning by erasing" account, are in some sense opposed[6]: the first views acquiring knowledge as expanding the domain of possible worlds accessible from the actual, the latter views acquiring knowledge as an elimination of possible words. But the feeling that they are opposed will disappear if we note that Lewis' intuition is that the expansion of the domain is a possible effect of knowledge acquisition, whereas the „learning by erasing" paradigm argues that the effect of eliminating epistemic alternatives is acquiring knowledge. We can see the two as complementary if we note that the domain from which we eliminate alternative hypotheses is dynamic, changing in time based on our current scientific knowledge. At different points in time the domains will have different cardinality or different components, i.e. different worlds which we may consider to be actual or, what turns out to be the same thing, different competing hypotheses. The techniques of model transformation come in only after stating the domain of worlds (hypotheses), in order to:

(a) Offer a way to represent scientific discovery as elimination of epistemic alternatives (in the sense of Martin and Osherson 2002 and Gierasimczuk 2009b);

(b) Tell us whether we can reach knowledge after receiving certain different new pieces of information.

To conclude: Lewis' thesis tells us that the possible worlds and their relations with the actual one are dynamic, while epistemic logic comes in only after a domain of possible worlds will have been fixed, as a tool for describing the process of discovery and reasoning about knowledge.

[5] Lewis argues that there are at least $\aleph_2$ worlds (see Lewis 1979, the footnote at p. 188).

[6] The argument was communicated to me by Gheorghe Ştefanov in a private conversation.

III.2. SHOULD EPISTEMIC INDISTINGUISHABILITY BE AN EQUIVALENCE RELATION?

In this subsection we will use David Lewis' remark on conceivability to show that an epistemic indistinguishability relation cannot be symmetric, therefore another kind of models should be employed to represent the acquisition of knowledge about the world.

Suppose Lewis' intuition is correct and consider the following case. At 2.000 BCE we may have considered that the actual world is one in which divine action is responsible for certain phenomena, e.g., combustion is explained by divine action. Now suppose Nature selected a world in which everything is made of atoms and there are no such forces as gods and spirits. Could that world have been conceived by us at 2.000 BCE? It is very likely that because of their very different scientific backgrounds, the „atoms-world" is a good candidate for an unconceivable world from the „2.000 BCE-world". Simply put, it is very likely that we could not have conceived such a thing as a world without divine forces and composed only of material particles, as we do nowadays.[7] If we accept this intuition, then, within an epistemic model, the „atoms-world" should not be linked or accessible from any of the states that we might have considered actual in 2.000 BCE.

But it is consistent with Lewis' intuition to link the „atoms-world" (in only one direction!) to the „2.000 BCE-world", because in our current state of knowledge we can conceive of such a world as that in which divine power is the cause of combustion, and consider this hypothesis as wrong. Note that it is not the distance in time that makes one world inconceivable from the other, it is the fact that the two hypotheses, the one involving explanations in terms of our current scientific ontology and the one involving explanations in terms of divine action, belong to different scientific backgrounds.

As a conclusion, the framework of „learning by erasing" needs an epistemic model whose accessibility relations are not necessarily symmetric, so that even if we may have reasons to link a world w to a world u, this will not imply accessibility to w from u, as in the case of equivalence relations. The plausibility models of Baltag and Smets (2006; 2011) could be a solution, since they include only reflexive and transitive accessibility relations for plausibility between worlds. Gierasimczuk (2009b; 2010) also considered using plausibility models, but for different reasons: they allow for upgrades, model-transforming techniques that do not eliminate possible worlds but only change the plausibility relation between possible worlds and they allow a better logical modeling of learning in the limit.

[7] Although atomism has ancient roots, the „atoms" natural philosophers postulated are not the atoms that we talk about in our day and time.

IV. CONCLUSION

To summarize, in the first section we presented the notion of propositional knowledge that is used in EL and we showed how learning can be modeled in a particular kind of dynamic epistemic logic, PAL. We also described Gierasimczuk's „learning by erasing", an account of a game theoretical perspective on scientifc inquiry in PAL. We used Lewis' intuitions on conceivability of other possible worlds to argue against using equivalence epistemic indistinguishability relations in the „learning by erasing" framework and proposed Baltag and Smets' plausibility models as an alternative.

REFERENCES

Baltag, A. and S. Smets. 2006. Conditional doxastic models: A qualitative approach to dynamic belief revision. In G. Mints and R. de Queiroz, eds., Proceedings of WOLLIC 2006, *Electronic Notes in Theoretical Computer Science*, 165:5-21.

Baltag, A. and S. Smets. 2011. Keep changing your beliefs, aiming for the truth. In T. Kuipers and G. Schurz, eds., *Erkenntnis*, 75(2): 255-270.

Solecki, S., A. Baltag, and L.S. Moss. 1999. The logic of public announcements, common knowledge and private suspicions. Technical report, Centrum voor Wiskunde en Informatica, Amsterdam. CWI Report SEN-R9922.

Blackburn, P., M. de Rijke and Y. Venema, Y. 2002. *Modal Logic*. Cambridge University Press.

Demey, L. 2011. Some remarks on the model theory of epistemic plausibility models. *Journal of Applied Non-Classical Logics*, 21 (3-4): 375-395.

van Ditmarsch, H.P., B.P. Kooi and W. van der Hoek. 2006. *Dynamic Epistemic Logic*. Springer Publishing Company.

van Ditmarsch, H.P., W. van der Hoek and B.P. Kooi. 2005. Dynamic epistemic logic with assignment. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS05): 141–148. ACM Inc., New York.

van Ditmarsch, H.P., B.P. Kooi. 2008. Semantic results for ontic and epistemic change. In: G. Bonanno, W. van der Hoek and M. Wooldridge, eds., *Logic and the Foundations of Game and Decision Theory (LOFT 7). Texts in Logic and Games 3*: 87-117, Amsterdam University Press, Amsterdam.Fagin, R., J.Y. Halpern, Y. Moses and M. Vardi. 1995. *Reasoning about Knowledge*. Cambridge, Massachusetts: The MIT Press 1995.

Groeneveld, W. and Gerbrandy, J. 1997. Reasoning about Information Change. *Journal of Logic, Language and Information*, 6(2): 147-169.

Gierasimczuk, N. 2009a. Bridging learning theory and dynamic epistemic logic. *Synthese*, 169(2): 371-384.

Gierasimczuk, N. 2009b. Learning by erasing in dynamic epistemic logic. In *Proceedings of the 3rd International Conference on Language and Automata Theory and Applications*, LATA '09, pp. 362-373, Berlin, Heidelberg, 2009. Springer-Verlag.

Gierasimczuk, N. 2010. *Knowing One's Limits. Logical Analysis of Inductive Inference*. PhD thesis, University of Amsterdam.

Hintikka, J. 1962. *Knowledge and Belief*. Ithaca, N.Y., Cornell University Press.

Kooi, B.P. 2007. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17 (2): 231-253.

Lewis, D. 1979. Possible worlds. In M.J. Loux, editor, *The Possible and the Actual*, pp. 225-235. Cornell University Press.

Martin, E. and D. Osherson. 2002. Scientific discovery from the perspective of hypothesis acceptance. *Proceedings of the Philosophy of Science Association*, 2002 (3): 331-341.

Plaza, J. 1989. „Logics of public communcations". In M. L. Emrich, M. Z. Pfeifer, M. Hadzikadic, and Z. W. Ras, eds., *Proc. 4th Inernational Symposium on Methodologies for Intelligent Systems*, pp. 201–216. Oak Ridge National Laboratory, ORNL/DSRD-24.

Pacuit, E., J. van Benthem and O. Roy. 2011. Toward a theory of play: A logical perspective on games and interaction. *Games*, 2 (1): 52-86.