Note on Unawareness: Negative Introspection vs. AU Introspection (and KU Introspection)*

Yi-Chun Chen[a], Jeffrey C. Ely[a], Xiao Luo[b,c]

[a] *Department of Economics, Northwestern University, Evanston, IL 60208*

[b] *Department of Economics, National University of Singapore, Singapore 117570, Singapore*

[c] *Institute of Economics, Academia Sinica, Taipei 115, Taiwan, ROC*

This Version: March 2009

Abstract

We show that Negative Introspection equals KU Introspection and AU Introspection for plausible unawareness in a standard state-space model *à la* Dekel et al. [Econometrica 66(1998) 159-173]. If, moreover, the standard state-space model is derived from a possibility correspondence and satisfies Non-Delusion, then AU Introspection alone is equivalent to Negative Introspection. This study provides additional insight into DLR's result that standard state-space models preclude sensible unawareness, in terms of how AU Introspection and KU Introspection clash with Plausibility. We also present a simple model to capture non-trivial unawareness. *JEL Classification: D80.*

Keywords: unawareness, negative introspection, KU introspection, AU introspection, plausibility, standard state-space model

*This paper is based upon part of our previous manuscripts entitled "A Note on Unawareness" and "Another Note on Unawareness." Acknowledgements to be added.

1 Introduction

Unawareness is a real-life phenomenon associated with an unconscious mental state directed toward, or lacking of positive knowledge about, a definite event. Unawareness can play an important role in economic implications. For example, unforeseen contingencies could prevent contracting parties from writing a complete contract with contingencies of which they were unaware at the contractual date.[1] The early work of Modica and Rustichini [19] (henceforth MR) provides a formal model of unawareness and study the logic of awareness; see also Fagin and Halpern [7]. MR define awareness in terms of knowledge or belief: an agent is unaware of an event if he/she does not know the event and does not know that he/she does not know the event.[2] In a propositional model MR show that no nontrivial unawareness is compatible with partitional information structures.[3]

Dekel et al. [4] (henceforth DLR) introduce three axioms for the concept of unawareness – namely Plausibility, KU Introspection, and AU Introspection – and show that standard state-space models preclude sensible unawareness. More specifically, DLR [4, Theorem 1] show that (i) under Necessitation, unawareness must be trivial, and (ii) under Monotonicity, if an agent is unaware of anything, he/she is unaware of everything and knows nothing. For MR's definition of unawareness DLR demonstrate, through parts B and C of Example 1, that unawareness may violate both KU Introspection and AU Introspection.

In this short note, under MR's definition of unawareness, we show that Negative Introspection indeed equals KU Introspection and AU Introspection (Proposition 1). This characterization result provides additional insight into MR's criticism that non-partitional possibility correspondence models are incapable of cap-

[1] Another example of non-trivial unawareness operator can be associated with the concept of "aware of unawareness" which is studied by Chung and Fortnow [3]; see also Geanakoplos [11] for more examples.

[2] MR [19, Proposition 2.1] show that this definition of unawareness implies not knowing the event, not knowing of not knowing the event and so on *ad infinitum*.

[3] Modica and Rustichini [20] take a different route of a syntactic approach and suggest an enhanced structure that allows for non-trivial unawareness in partitional information structures; see also Halpern [12].

turing an interesting form of unawareness; in particular, it illustrates how AU Introspection and KU Introspection clash with Plausibility: for non-trivial unawareness the latter must rule out Negative Introspection which by our result is equivalent to the former with virtually no assumptions on knowledge.

We next extend this result for non-trivial plausible unawareness in a standard state-space model (Proposition 2). One important implication of Proposition 2 is that: if we are willing to drop Negative Introspection (a must if we are to allow for unawareness) this entails relaxing either AU Introspection and KU Introspection in standard state spaces. Moreover, the result can be strengthened for the state-space model in which the knowledge operator satisfies Non-Delusion and Monotonicity (Proposition 3). Consequently, this paper provides some additional insight into DLR's result that standard state-space models preclude sensible unawareness.

We would like to point out that DLR [4, Footnote 10] observe that the primary conflict is between AU Introspection and possibility correspondences. However, DLR [4] do not investigate the logical relationship between the properties of the knowledge operator and the unawareness operator while we will do in this paper. For example, Corollary 1 shows that the conflict between AU Introspection and possibility correspondences originates from the fact that AU Introspection is equivalent to Negative Introspection in this case. Finally, we also present a simple model to capture non-trivial unawareness.

2 Results

Following DLR [4], we consider a standard state-space model

$$(\Omega, K, U),$$

where Ω is a state space, the knowledge operator $K : 2^{\Omega} \to 2^{\Omega}$ assigns each event $E \subseteq \Omega$ to a set KE of states in which it is known that E occurs, the unawareness operator $U : 2^{\Omega} \to 2^{\Omega}$ assigns each event $E \subseteq \Omega$ to a set UE of states in which it is unaware of the possibility that E occurs. Throughout this paper, we require that the knowledge operator K satisfies $K\Omega = \Omega$ and $K\varnothing = \varnothing$.[4]

The standard state-space model (Ω, K, U) can be applied to very general situations. For example, a standard propositional model used by MR [19] can be recast as a standard state-space model (see DLR [4, Section 3]). Furthermore, the knowledge operator K can be derived either from an underlying possibility correspondence in a possibility correspondence model (see, e.g., Aumann [1]) or from the Savage-nullity notion in a decision model under uncertainty (see, e.g., Epstein and Wang [6] and Morris [21]); the unawareness operator U may take the specific form of MR's [19] definition:

$$U^{MR}E \equiv \neg KE \cap \neg K\neg KE.$$

Based on the idea that the concept of unawareness corresponds to a complete lack of positive knowledge, DLR [4] suggest three axioms that an unawareness operator should satisfy:

Plausibility: $UE \subseteq \neg KE \cap \neg K\neg KE$.
***KU* Introspection:** $KUE = \varnothing$.
***AU* Introspection:** $UE \subseteq UUE$.

That is, Plausibility says that an agent is unaware of an event only if the agent does not know it is true and the agent does not know that he/she does not know it is true, KU Introspection requires that the agent does not know he\she is unaware, and AU Introspection requires that the agent be unaware of being unaware. We say U is *non-trivial* if $U(E) \neq \varnothing$ for some event E. Observe that no non-trivial plausible U can coexist with the following property for knowledge:

Negative Introspection: $\neg KE \subseteq K\neg KE$.

[4]The propert $K\Omega = \Omega$ is known as Necessitation; see DLR [4, p. 164].

That is, Negative Introspection requires to know that he\she does not know. In the state-space model generated by a possibility correspondence, Negative Introspection is a necessary condition for the possibility correspondence to be partitional.

We start by stating a proposition that Negative Introspection indeed equals AU Introspection and KU Introspection for MR's unawareness operator U^{MR}, as demonstrated partially by DLR's [4] Example 1: Parts B and C. This characterization result provides some additional insight into MR's criticism that possibility correspondence models are incapable of capturing an interesting form of unawareness. Let $\left(\Omega, K, U^{MR}\right)$ denote the state-space model arises from the standard propositional model used by MR [19]. Formally, we have

Proposition 1. *In the state-space model* $\left(\Omega, K, U^{MR}\right)$*,* U^{MR} *satisfies* AU *Introspection and* KU *Introspection if, and only if,* K *satisfies Negative Introspection.*
Proof. $\Rightarrow$: Since $K\Omega = \Omega$ and U^{MR} satisfies Plausibility, by DLR [4, Theorem 1(i)], $U^{MR}E = \varnothing$ for all events E. Therefore, $\neg KE \subseteq K\neg KE$ for all events E.

$\Leftarrow$: By Negative Introspection, $\neg KE \cap \neg K\neg KE = \varnothing$ for all events E. Therefore, $U^{MR}E = \varnothing$ for all events E. So AU Introspection trivially holds. Since $K\varnothing = \varnothing$, $KU^{MR}E = \varnothing$, i.e., KU Introspection holds.■

Next, we extend the result of Proposition 1 to standard state-space models. In doing so, we will keep the most basic axiom of Plausibility for the rest of the paper.[5] Observe that a "sensible" plausible unawareness operator U should satisfy the condition that K violates Negative Introspection only if U is non-trivial; otherwise, if the condition fails to hold, then any plausible unawareness operator U must be trivial. Subsequently, we will only consider the state-space model (Ω, K, U) in which the plausible unawareness operator U satisfies this condition, e.g., $\left(\Omega, K, U^{MR}\right)$ is an outstanding example of such a state-space model. We are

[5] This is in accordance with the spirit of DLR. As DLR [4, Footnote 10] note, "We view plausibility as the most basic of the three axioms, so it seems natural to only consider relaxations of our assumptions which keep this."

in position to state our main result that Negative Introspection is equivalent to AU Introspection and KU Introspection in standard state-space models.

Proposition 2. *In the state-space model* (Ω, K, U)*, Negative Introspection is equivalent to* AU *Introspection and* KU *Introspection.*

Proof. By Negative Introspection, $\neg KE \cap \neg K \neg KE = \varnothing$ for all events E. By Plausibility, $UE = \varnothing$ for all events E. So AU Introspection trivially holds. By $K\varnothing = \varnothing$, $KUE = \varnothing$, i.e., KU Introspection holds.

Suppose, on the contrary, that AU Introspection and KU Introspection hold. Under Plausibility, by DLR [4, Theorem 1(i)], $UE = \varnothing$ for all events E. Since K violates Negative Introspection only if U is non-trivial, K satisfies Negative Introspection.■

The Proposition 2 provides additional insight into DLR's [4, Theorem 1(i)] result that in standard state-spaces, AU Introspection and KU Introspection clash with Plausibility; in particular, for non-trivial unawareness, Plausibility must rule out Negative Introspection whereas AU Introspection and KU Introspection are equivalent to Negative Introspection with virtually no assumptions on knowledge. As DLR [4, Footnote 10] note, any knowledge operator satisfying Plausibility, Monotonicity, and Non-Delusion must satisfy KU Introspection. Under these additional assumptions of Monotonicity and Non-Delusion, we have the following strengthened Proposition 3.

Proposition 3. *In the state-space model* (Ω, K, U) *which satisfies Non-Delusion (i.e.* $KE \subseteq E$ *for all events* E*) and Monotonicity (i.e.* $KE \subseteq KF$ *for any events* $E \subseteq F$*), Negative Introspection is equivalent to* AU *Introspection.*

Proof. We show first that KU Introspection holds under our assumptions. Suppose $\omega \in KUE$ for some $E \subseteq \Omega$. By Plausibility, $UE \subseteq \neg KE \cap \neg K \neg KE$. By Monotonicity, $KUE \subseteq K(\neg KE \cap \neg K \neg KE) \subseteq K \neg KE$ and, hence, $\omega \in K \neg KE$. By Non-delusion, $K(\neg KE \cap \neg K \neg KE) \subseteq \neg KE \cap \neg K \neg KE$ and, hence, $\omega \notin K \neg KE$, which is a contradiction. Therefore, KU Introspection holds. By Proposition 2, it follows that Negative Introspection is equivalent to AU Introspection.■

Note that the knowledge operator derived from a possibility correspondence satisfies Monotonicity (as well as $K\Omega = \Omega$ and $K\varnothing = \varnothing$). The following corollary is an immediate implication of Proposition 3.

Corollary 1. *In the* state-space *model* (Ω, K, U) *where the knowledge operator* K *is derived from a possibility correspondence and satisfies Non-Delusion, Negative Introspection is equivalent to* AU *Introspection.*

MR [19, Theorem] show that Negative Introspection in equivalent to U^{MR} is symmetric (i.e. $U^{MR}E = U^{MR}(\Omega \backslash E)$) in $\left(\Omega, K, U^{MR}\right)$.[6] Since $\left(\Omega, K, U^{MR}\right)$ also satisfies Non-Delusion and Monotonicity, the following corollary is an immediate implication of Proposition 3.

Corollary 2. *In the state-space model* $\left(\Omega, K, U^{MR}\right)$, U^{MR} *is symmetric if, and only if,* U^{MR} *satisfies* AU *Introspection.*

Remark. Proposition 2 asserts that if we stick to the conventional framework of a standard state space, then we have to drop either AU Introspection and KU Introspection in order to allow for non-trivial unawareness given that Negative Introspection must be dropped at the first place. In the case of more restricted state-space models, Proposition 3 asserts that if we are willing to drop Negative Introspection (a must if we are to allow for unawareness), this entails the willingness to accept awareness of unawareness. Otherwise, we must go beyond standard state spaces to model non-trivial unawareness; a strand of recent literature has emerged to investigate syntactic models with unawareness (see, e.g., Board and Chung [2], Fagin and Halpern [7], Feinberg [8, 9], Halpern [12], Halpern and Rego [13], Heifetz et al. [16], Modica and Rustichini [20], and Sillari [22]), or generalized state-space models, e.g., Galanis [10] models the agent's theorems by the agent's knowledge of impossible states and presents that being more aware leads to more knowledge of the agent; Heifetz et al. [14, 15] propose a model with a

[6] Xiong [23] defines an unawareness operator which satisfies symmetry and is non-trivial, but violates the requirement of Plausibility. Xiong considers a variant of "plausibility" for the unawareness operator, i.e., the lack of positive knowledge regarding whether an event E or its complement $\neg E$.

lattice of state spaces, ordered by "expressive power," and allow the possibility set at a state to reside in a different state space; Li [17] proposes a product model that overcomes the inconsistency between AU Introspection and Negative Introspection by modeling information as a pair and exploiting a product structure on the state space.[7]

3 A Simple Model of Unawareness

In this section, we will revisit the leading example in DLR [4] to demonstrate the connection between AU Introspection and Negative Introspection in a simple state-space model associated with a possibility correspondence, and present a slightly generalized state-space model to analyze non-trivial unawareness. In particular, we will argue that there is an intuitive connection behind the equivalence between AU Introspection and Negative Introspection. Moreover, we will see that this intuitive connection helps us to motivate a simple model to capture unawareness.

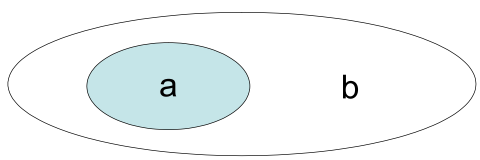

Fig. 1. The DM's knowledge structure

Consider the example where $\Omega = \{a, b\}$, $P(a) = \{a\}$, and $P(b) = \Omega$, and the knowledge operator is generated from P in the usual way (see Figure 1). This is a standard state-space model which is non-partitional. Firstly, observe that $Ka = a$ and $K\neg Ka = Kb = \varnothing$ and hence Negative Introspection is violated.[8] In

[7]For example, for the unawareness structures due to Heifetz et al. [14, 15], it can be shown that our results hold true for all events for which the base-space is the meet (i.e. the space with the lowest expressive power) of the lattice. For events with higher base-spaces, counterexamples can be constructed. That is, our results hold precisely because the "language" of the standard state-space is not rich enough to capture non-trivial unawareness.

[8]In this example, the relevant non-trivial events are singletons, so for notational ease we will avoid brackets.

this case, we take the viewpoint that the DM must *not* fully realize the mechanism which generates his knowledge. To see this, note that if the DM correctly perceive $Ka = a$ at b, then he/she would be subject to *reductio ad absurdum*:

1. I don't know that a is true.

2. If a were true, I would know it.

Hence, he/she would infer that the state is b since he/she does not know a. Indeed, if $Ka = \varnothing$ instead of a, we have $\neg Ka = \Omega = K\Omega = K\neg Ka$ and there is no violation of Negative Introspection.

Secondly, it follows from our Proposition 3 that AU Introspection must also be violated for any plausible unawareness. KKKIndeed, if $Ua = \neg Ka \cap \neg K\neg Ka = b \cap \Omega = b$, then $UUa \subseteq \neg Kb \cap \neg K\neg Kb = \Omega \cap \varnothing = \varnothing$. Hence, $Ua \nsubseteq UUa$. In other words, the MD's unawareness structure displays a failure of AU Introspection: at b the the MD is aware that he/she is unaware of a. Now suppose that $Ka = \varnothing$ at b for the DM. Then, $Ua = \neg Ka \cap \neg K\neg Ka = \Omega \cap \varnothing = \varnothing$ and hence there is no violation of AU Introspection. Note that, only if the DM at b correctly understood the knowledge mechanism and hence stipulated that $Ka = a$, would he be enforced to violate AU Introspection.KKK

The discussion motivates a simple approach to capturing unawareness in an essentially standard state-space model.[9] The theme of the approach is that all logical inconsistencies are detectable only by an analyst who can consult the entire knowledge mechanism. At any given state, the decision-maker's view of the world is invulnerable to *reductio ad absurdum*. However, he/she may maintain different views of the world in different states, and these views may be inconsistent. These inconsistencies survive the decision-maker's introspection because at any given state, he/she is unaware of whatever mechanism would lead him/her to a different model at some different state. This sounds a lot like the approach sketched in DLR; certainly the two approaches follow from the same intuition about the nature of unawareness. The present approach however seems simpler and probably more tractable. In fact, it can be represented by a generalized standard state-space model.

[9] See Li [18] for a general model along this line.

The knowledge structure is given by a set of states Ω, and a *collection* of partition $\Pi(\omega)$, one for each $\omega \in \Omega$. The partition $\Pi(\omega)$ represents the decision-maker's model of the world when the state ω is realized, and plays the role of the subjective factual information in the partition model. Because it is a partition, the decision-maker's model survives Negative Introspection and AU Introspection at every state. However, from an analyst's perspective, he/she may violate both. An analyst who knows the mapping $\omega \to \Pi(\omega)$, can reduce the knowledge model into a standard possibility correspondence $\omega \to P(\omega)$, where $P(\omega)$ is the set of states deemed possible by the decision maker according to his/her model $\Pi(\omega)$ at ω. That is, $P(\omega) = \Pi(\omega)(\omega)$. In general, $P(\omega)$ constructed in this way need not be a partition, in which case the analyst will detect a logical inconsistency of which the decision-maker is unaware.

For example, in the previous example, we can model the DM's knowledge and awareness as follows. At a, the DM has the discrete partition $\Pi(a) = \{\{a\}, \{b\}\}$ and at b, the DM has the trivial partition $\Pi(b) = \{\{a, b\}\}$. We can derive the DM's (now state-dependent) knowledge operator: $K_\omega : \Omega \to 2^\Omega$. At b, according to the model of the world that the DM holds at b, the set of states at which the DM knows that a is true, $K_b(a)$ is empty. This matches the postulate from the informal argument above. For completeness, $K_b(b) = \varnothing$, $K_a(a) = a$, $K_a(b) = b$. The analyst reduces this knowledge structure to the usual non-partitional possibility correspondence in Figure 1. Indeed, dropping the subscripts to denote the analyst's model of the DM's knowledge, we have $Ka = a$, $Kb = \varnothing$ as desired.

References

[1] R. Aumann, Agreeing to disagree, Annals of Statistics 4 (1976), 1236-1239.

[2] O. Board, K.S. Chung, Object-based unawareness, mimeo (2007), University of Minnesota.

[3] K.S. Chung, L. Fortnow, Loopholes, mimeo (2006), University of Minnesota.

[4] E. Dekel, B.L. Lipman, and A. Rustichini, Standard state-space models preclude unawareness, Econometrica 66 (1998), 159-173.

[5] J.C. Ely, A note on unawareness, mimeo (1998), Northwestern University.

[6] L. Epstein, T. Wang, 'Beliefs about beliefs' without probabilities, Econometrica 64 (1996), 1343-1373.

[7] R. Fagin, J. Halpern, Belief, awareness, and limited reasoning, Artificial Intelligence 34 (1988), 39-76.

[8] Y. Feinberg, Subjective reasoning - games with unawareness, mimeo (2004), Stanford University.

[9] Y. Feinberg, Games with incomplete awareness, mimeo (2005), Stanford University.

[10] S. Galanis, Unawareness of theorems, mimeo (2007), University of Southampton.

[11] J. Geanakoplos, Games theory without partitions, and applications to speculation and consensus, Cowles Foundation Discussion Paper No. 914 (1989).

[12] J.Y. Halpern, Alternative semantics for unawareness, Games and Economic Behavior 37 (2001), 321-339.

[13] J.Y. Halpern, L.C. Rego, Reasoning about knowledge of unawareness, in: Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning (2006).

[14] A. Heifetz, M. Meier, and B.C. Schipper, Interactive unawareness, Journal of Economic Theory 130 (2006), 78-94.

[15] A. Heifetz, M. Meier, and B.C. Schipper, Unawareness, beliefs and games, mimeo (2007), UC_Davis.

[16] A. Heifetz, M. Meier, and B.C. Schipper, A canonical model for interactive unawareness, Games and Economic Behavior 62 (2008), 305-324.

[17] J. Li, Information structures with unawareness, Journal of Economic Theory, forthcoming.

[18] J. Li, Modeling unawareness in arbitrary state spaces (2008), University of Pennsylvania.

[19] S. Modica, A. Rustichini, Awareness and partitional information structures, Theory and Decision 37 (1994), 107-124.

[20] S. Modica, A. Rustichini, Unawareness and partitional information structures, Games and Economic Behavior 27 (1999), 265-298.

[21] S. Morris, The logic of beliefs and beliefs change: a decision theoretic approach, Journal of Economic Theory 69 (1996), 1-23.

[22] G. Sillari, Models of unawareness, in: Logic and the Foundations of Game and Decision Theory: Proceedings of the Seventh Conference (2006), ed. by G. Bonanno, W. van der Hoek, and M. Woolridge.

[23] S. Xiong, Unawareness and the standard state space, mimeo (2007), Northwestern University.