MULTI-PERSON UNAWARENESS

Aviad Heifetz* Martin Meier† Burkhard C. Schipper‡

May 2003

Abstract

Standard state-space models, which are widely used in economics, preclude non-trivial forms of unawareness as shown by Dekel, Lipman and Rustichini (1998). We define a generalized state-space model that allows for unawareness. In order to facilitate applications we make no explicit use of modal syntax within the sematic model. Our model satisfies all "S4" properties as well as all desiderata on unawareness proposed by Modica and Rustichini (1999), Halpern (2001) as well as Dekel, Lipman and Rustichini (1998).

JEL-Classifications: C70, D80.

Keywords: unawareness, awareness, knowledge, interactive epistemology, bounded perception.

*The School of Economics, Tel Aviv University and CORE, Université Catholique de Louvain. Email: heifetz@post.tau.ac.il

†CORE, Université Catholique de Louvain. Email: meier@core.ucl.ac.be

‡Dept. of Economics, University of Bonn. Email: burkhard.schipper@wiwi.uni-bonn.de, Web: www.bgse.uni-bonn.de/~burkhard

1 Introduction

It is hard to argue that decision makers are aware of all facts effecting the outcome of their decisions. Thus unawareness is a rather natural state of mind and its role especially in interactive decision making should be investigated. Yet modeling unawareness proves to be a tricky task. Geanakoplos (1989) suggested using non-partitional information structures (i.e. Kripke structures) to this effect. In such a model one can have states in which an individual doesn't know an event and is ignorant of her ignorance. However, as Dekel, Lipman and Rustichini (1998) show, in such states she knows that she is ignorant of her ignorance, and therefore it is not appropriate to ascribe unawareness of the event to the individual in such states. More generally, Dekel, Lipman and Rustichini (1998) showed that no standard information structures (i.e. Kripke structures) can capture adequately the notion of unawareness.[1]

Modica and Rustichini (1999) suggest an enhanced structure in order to model unawareness of an individual. They consist of an "objective" space, describing the world with the full vocabulary, and a "subjective" space for each sub-vocabulary. When an individual is unaware of an event, the states

[1]Ewerhart (2001) suggests a way to model unawareness in a standard information structure. However, in his modeling if an individual is unaware of an event then she is aware of its negation. While this property may be suitable for some aspect or view of unawareness, it is incompatible with all the other formal approaches cited here, as well as with the approach of the current contribution.

she considers as possible belong to a subjective space in which this event cannot be described. Halpern (2001) offers an alternative formulation with one space but two different knowledge operators – implicit knowledge and explicit knowledge. Halpern (2001) proves that a particular kind of his awareness structures are equivalent to the Modica-Rustichini structure as a semantics for a modal syntax that includes both a knowledge and an awareness modality.

Both these approaches suffer from the following limitations. First, they involve an explicit use of the modal syntax *within* the semantic structures. This limits the audience that is capable of applying this machinery to specific problems. Just as the short paper by Aumann (1976) introduced to economists the partitional Kripke structures as a logic-free tool to model knowledge, and was thus seminal to a large body of consecutive work in Economics, the analogue of such a presentation is still lacking for unawareness.

Second, only one-person unawareness is treated explicitly both by Modica and Rustichini (1999) and Halpern (2001). While in the case of knowledge, the passage from the single-person case to multi-person case involves no substantial complications, the modeling of multi-person unawareness is more involved. An individual may be unaware of some fact, but believe that another person is unaware of another. To model this appropriately, one needs an explicit *ordered structure of spaces,* where the possibility set of an individual in a state of one space may reside in another space, while the possibility set of a different individual in one of these possible states may reside in yet another space.

To wit, we consider a complete lattice of state spaces accompanied by suitable projections among them. The partial order of spaces indicates the strength of their expressive power. The possibility set of an individual in a state of one space may reside in a less-expressive space. A crucial feature of the model is that it limits the subsets (of the union of all spaces) which are considered as events – those that can be "known" or be the object of awareness. The special structure of events is natural, in the sense that it is the same as that of subsets of states in which a particular formula obtains – if states were to consist of maximally-consistent sets of formulas in an appropriate logical formulation.[2] In particular, in our setting the negation of an event is different from its set-theoretic complement. As a result, there are states that belong neither to an event nor to its negation. When the possibility set of an individual consists of such states, the individual is unaware of the event.

2 Model

$\mathcal{S} = \{S_\alpha\}_{\alpha \in A}$ is a complete lattice of disjoint spaces, with $\preceq$ a partial order on $\mathcal{S}$. Denote $\Sigma = \bigcup_{\alpha \in A} S_\alpha$ the union of these spaces.

If $S' \succeq S$ ("S' is more expressive than S – states of S' describe situations with a richer vocabulary than states in S"), then $r^{S'}_S : S' \to S$ is a surjective

[2]We show this formally in a companion work.

projection. ("$r_S^{S'}(\omega)$ is the restriction of the description ω to the more limited vocabulary of S.") Projections commute: If $S'' \succeq S' \succeq S$ then $r_S^{S''} = r_{S'}^{S''} \circ r_S^{S'}$. If $\omega \in S'$, denote $\omega_S = r_S^{S'}(\omega)$. If $B \subseteq S'$, denote $B_S = \{\omega_S : \omega \in B\}$.

Denote by $g(S) = \{S' : S' \succeq S\}$ the set of spaces that are at least as expressive as S. For $B \subseteq S$, denote by $B^{\uparrow} = \bigcup_{S' \in g(S)} \left(r_S^{S'}\right)^{-1}(B)$ all the "extensions of descriptions in B to more expressive vocabularies."

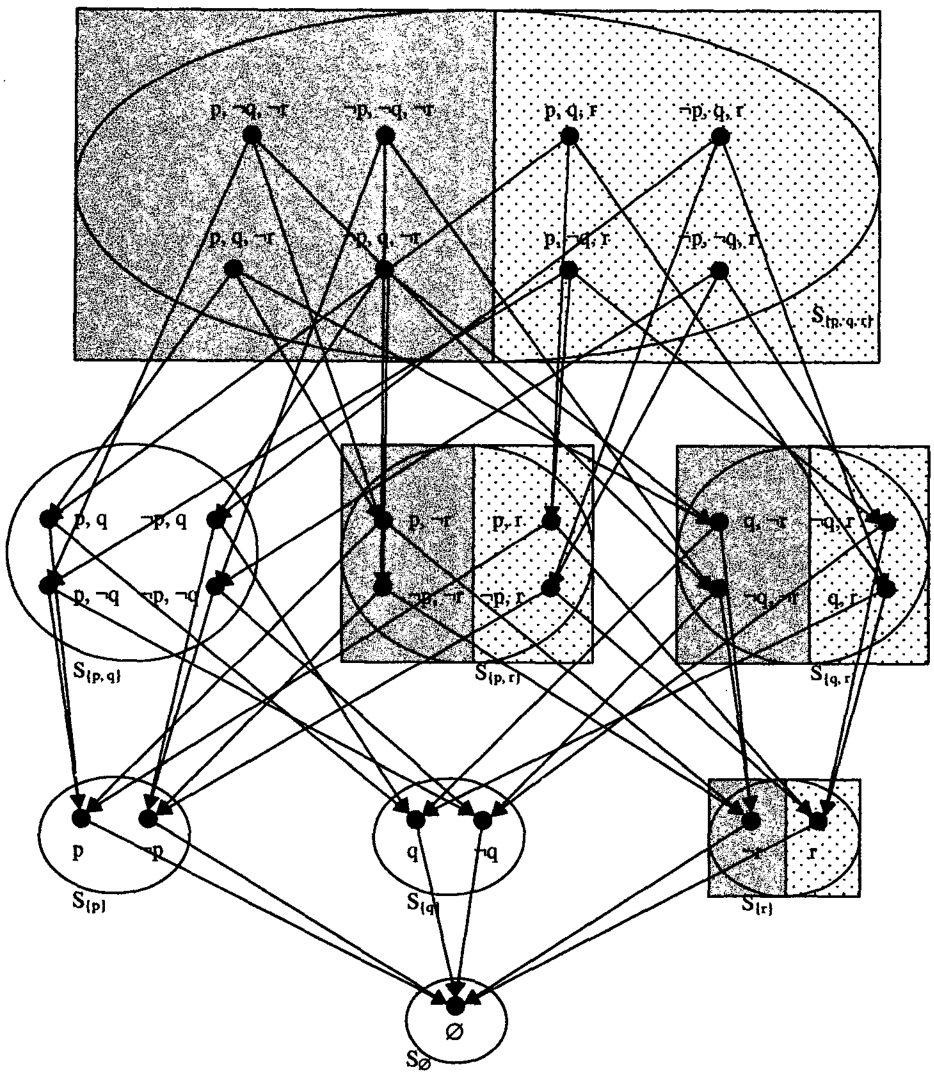
An event E is a subset of Σ of the form $B^{\uparrow}$ for some $B \subseteq S$, where $S \in \mathcal{S}$. In such a case we call S the base-space of the event E, and denote it by $S(E)$. Notice that not every subset of Σ is an event.

If $B^{\uparrow}$ is an event where $B \subseteq S$, the negation $\neg B^{\uparrow}$ of $B^{\uparrow}$ is $(S \setminus B)^{\uparrow}$. This negation is typically a proper subset of the complement $\Sigma \setminus B^{\uparrow}$. If $B \neq \emptyset$ and $B \neq S$ for some $S \in \mathcal{S}$, then $\neg\neg B^{\uparrow} = B^{\uparrow}$, but otherwise it is not necessarily the case. To circumvent this, for each space $S \in \mathcal{S}$ we devise a distinct vacuous event $\emptyset^S$, and define $\neg S^{\uparrow} = \emptyset^S$ and $\neg \emptyset^S = S^{\uparrow}$. The event $\emptyset^S$ should be interpreted as a logical contradiction phrased with the expressive power available in S.

If $\left\{B_\lambda^{\uparrow}\right\}_{\lambda \in L}$ is a set of events (with $B_\lambda \subseteq S_\lambda$, $\lambda \in L$), their conjunction $\bigwedge_{\lambda \in L} B_\lambda^{\uparrow}$ is just the intersection $\bigcap_{\lambda \in L} B_\lambda^{\uparrow}$ (we will therefore use the conjunction symbol $\wedge$ and the intersection symbol $\cap$ interchangeably). If $S = \sup_{\lambda \in L} S_\lambda$, then this conjunction is $\left(\bigcap_{\lambda \in L}\left(\left(r_{S_\lambda}^{S}\right)^{-1}(B_\lambda)\right)\right)^{\uparrow}$.

As always, the disjunction of $\left\{B_\lambda^{\uparrow}\right\}_{\lambda \in L}$ is defined by the de Morgan law $\bigvee_{\lambda \in L} B_\lambda^{\uparrow} = \neg\left(\bigwedge_{\lambda \in L} \neg\left(B_\lambda^{\uparrow}\right)\right)$. Typically $\bigvee_{\lambda \in L} B_\lambda^{\uparrow} \supsetneqq \bigcup_{\lambda \in L} B_\lambda^{\uparrow}$.

Figure 1: State-Spaces, Projections and Event Structure



Example 1. Let Φ be a set of facts, and $A = 2^{\Phi}$. For $\alpha \in A$, $S_\alpha = \{\omega : \omega = \{true, false\}^\alpha\}$. I.e., a state in S_α is a string indicating which facts in α are true and which are false. $S_\alpha \preceq S_{\alpha'}$ whenever $\alpha \subseteq \alpha'$. Consider for instance a set of three facts $\Phi = \{p, q, r\}$. For example, we write $\omega = (p, \neg q)$ for a state in $S_{\{p,q\}}$ in which the fact p is true and q is false. Clearly, we have for example $g(S_\emptyset) = \mathcal{S}$, $g(S_{\{p,q,r\}}) = \{S_{\{p,q,r\}}\}$ and $g(S_{\{r\}}) = \{S_{\{r\}}, S_{\{p,r\}}, S_{\{q,r\}}, S_{\{p,q,r\}}\}$. Figure 1 illustrates the state-spaces with the states. The projections are indicated by arrows (for clarity we do not consider in this figure any compositions of projections and the identity maps). Consider now the event that fact r is true [r is true]. The base-space is $S_{\{r\}}$, the basis of this event is $\{(r)\} \subset S_{\{r\}}$. Considering all extensions of $\{(r)\}$ we obtain the event $\{(r)\}^\uparrow = \{(r), (p, r), (\neg p, r), (q, r), (\neg q, r), (p, q, r), (p, \neg q, r), (\neg p, q, r), (\neg p, \neg q, r)\}$ $=$ [r is true]. This is the set of states in which fact r obtains. In Figure 1 the event [r is true] is indicated by the union of the dotted rectangles. The event that r is false [r is false] is the negation $\neg$[r is true] $= (S_{\{r\}} \setminus \{(r)\})^\uparrow = \{(\neg r), (p, \neg r), (\neg p, \neg r), (q, \neg r), (\neg q, \neg r), (p, q, \neg r), (p, \neg q, \neg r), (\neg p, q, \neg r), (\neg p, \neg q, \neg r)\}$. In Figure 1 it is indicated by the union of the grey rectangles. It becomes obvious that [r is true] $\cup\, \neg$[r is true] $\subsetneqq \Sigma$. I.e., there are states such as (q) which belong neither to [r is true] nor $\neg$[r is true].

I is the set of individuals. For each individual $i \in I$ there is a nonempty possibility correspondence $\Pi_i : \Sigma \to 2^{\Sigma}$ with the following properties:

0. Confinedness: If $\omega \in S$ then $\Pi_i(\omega) \subseteq S'$ for some $S' \preceq S$.

1. Generalized Reflexivity: $\omega \in \Pi_i^{\uparrow}(\omega)$ for every $\omega \in \Sigma$

2. Stationarity: $\omega' \in \Pi_i(\omega)$ implies $\Pi_i(\omega') = \Pi_i(\omega)$

3. Projections Preserve Awareness: If $\omega \in S'$, $\omega \in \Pi_i(\omega)$ and $S \preceq S'$ then $\omega_S \in \Pi_i(\omega_S)$

4. Projections Preserve Ignorance: If $\omega \in S'$ and $S \preceq S'$ then $\Pi_i^{\uparrow}(\omega) \subseteq \Pi_i^{\uparrow}(\omega_S)$

5. Projections Preserve Knowledge: If $S \preceq S' \preceq S''$, $\omega \in S''$ and $\Pi_i(\omega) \subseteq S'$ then[3] $(\Pi_i(\omega))_S = \Pi_i(\omega_S)$

Remark 1 *Property* 1. *implies that if* $S' \preceq S$, $\omega \in S$ *and* $\Pi_i(\omega) \subseteq S'$ *then* $r^S_{S'}(\omega) \in \Pi_i(\omega)$.

The knowledge operator of individual i on events is defined, as usual, by

$$K_i(E) = \{\omega \in \Sigma : \Pi_i(\omega) \subseteq E\}.$$

Proposition 1 *If* E *is an event, then so is* $K_i(E)$.

Proposition 2 *The Knowledge operator* K_i *has all the "S4" properties* ($K_i(\Sigma) = \Sigma$, $K_i(E \cap F) = K_i(E) \cap K_i(F)$,[4] $K_i(E) \subseteq E$ *and* $K_i(E) \subseteq K_i K_i(E)$).[5] *Instead of the property* (5) $\neg K_i(E) \subseteq K_i \neg K_i(E)$, *the weaker property* $\neg K_i(E) \cap \neg K_i \neg K_i(E) \subseteq \neg K_i \neg K_i \neg K_i(E)$ *obtains.*

[3]We could have assumed $\supseteq$ and deduce $=$ from $\supseteq$, 3. and 2.

[4]In fact also $K_i\left(\bigcap_{\lambda \in L} E_\lambda\right) = \bigcap_{\lambda \in L} K_i(E_\lambda)$ for each set of events $\{E_\lambda\}_{\lambda \in L}$

[5]Monotonicity, i.e., $E \subseteq F$ implies $K_i(E) \subseteq K_i(F)$, also obtains.

The unawareness operator of individual i from events to events is now defined by[6]

$$U_i(E) = \neg K_i(E) \cap \neg K_i \neg K_i(E),$$

and the awareness opertor is then naturally defined by

$$A_i(E) = \neg U_i(E).$$

Proposition 3 *The following properties obtain:*

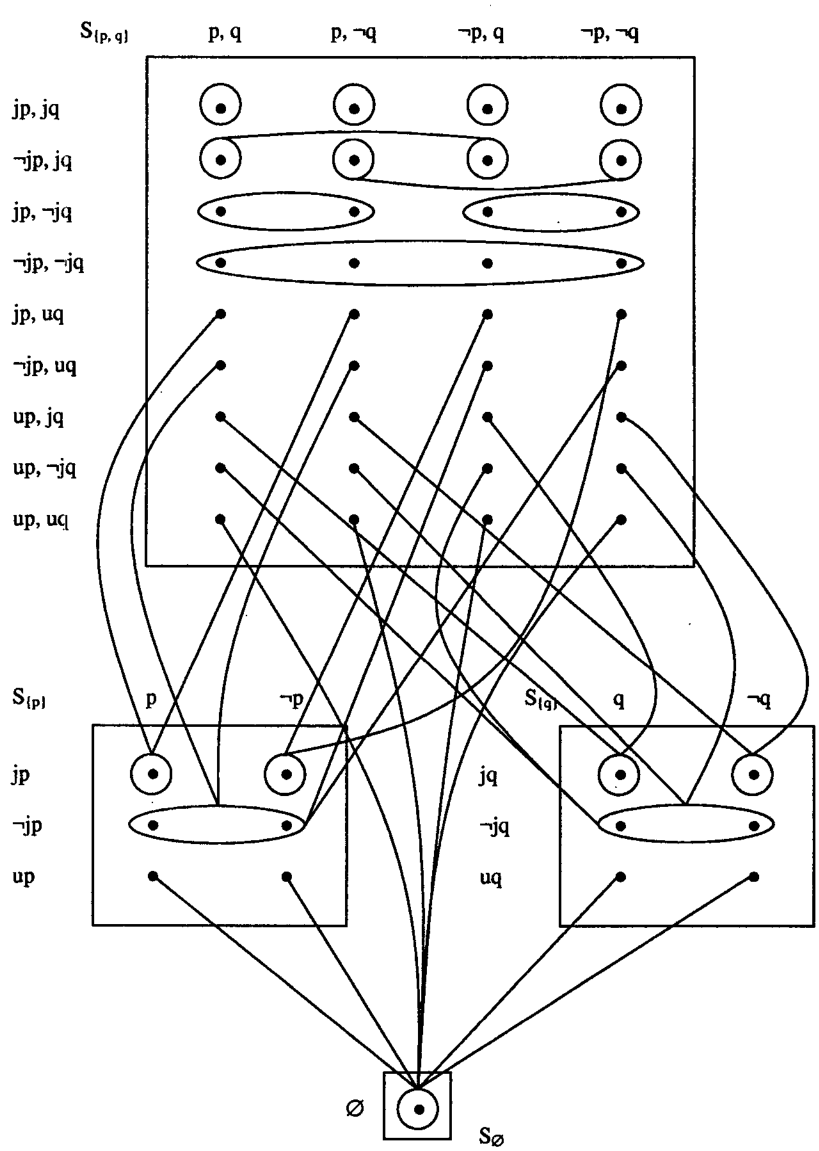
1. *KU Introspection:* $K_i U_i(E) = \emptyset^{S(E)}$
2. *AU Introspection:* $U_i(E) \subseteq U_i U_i(E)$
3. *Weak Necessitation:* $A_i(E) = K_i \left(S(E)^{\uparrow} \right)$
4. *Strong Plausibility:* $U_i(E) = \bigcap_{n=1}^{\infty} (\neg K_i)^n (E)$
5. *Weak Negative Introspection:* $\neg K_i(E) \cap A_i \neg K_i(E) \subseteq K_i \neg K_i(E)$
6. *Event Awareness:* $A_i(\neg E) = A_i(E)$
7. *A-Intersection:* $\bigcap_{\lambda \in L} A_i(E_\lambda) = A_i \left(\bigcap_{\lambda \in L} E_\lambda \right)$
8. *AK-Awareness:* $A_i K_i(E) = A_i(E)$
9. *AA-Self Reflection:* $A_i A_i(E) = A_i(E)$

[6]This is the Modica-Rustichini (1999) definition. In particular, the Dekel-Lipman-Rustichini (1998) Plausibility requirement $U_i(E) \subseteq \neg K_i(E) \cap \neg K_i \neg K_i(E)$ is satisfied by this definition.

Properties 1. to 4. have been proposed by Dekel, Lipman and Rustichini (1998), properties 6. to 9. by Modica and Rustichini (1999), and properties 5. to 9. by Halpern (2001).

Example 2. Let Φ be the set of atomic propositions, and let $A = 2^{\Phi}$. For a subset $\alpha \subseteq \Phi$ (i.e. $\alpha \in A$), let S_α be the set of maximally consistent sets of formulas in the sub-language $\mathcal{L}_\alpha$ containing only the atomic propositions of α. Again, $S_\alpha \preceq S_{\alpha'}$ whenever $\alpha \subseteq \alpha'$. Let for instance $\Phi = \{p, q\}$. The four spaces $\mathcal{S} = \{S_{\{p,q\}}, S_{\{p\}}, S_{\{q\}}, S_{\emptyset}\}$ are indicated in Figure 2 by rectangles. For convenience in the presentation we use the "knowing whether" operator j defined by $jp := kp \vee k\neg p$, i.e., jp means that an individual knows p or knows not p (see Hart, Heifetz and Samet, 1996). We can recover $kp \Leftrightarrow p \wedge jp$ and $k\neg p \Leftrightarrow \neg p \wedge jp$. For simplicity, each state is described by the atomic propositions that hold in this state as well as the information at that state. Thus we present in Figure 2 each state-space in a matrix-style. That is, the state (jp, jq, p, q) means that p and q obtains, and that the agent knows whether p and knows whether q. This of course implies that the agent knows p and q indicated by the singleton possibility set. The possibility sets are indicated by circles or ovals, some eventually connected by lines. Other lines relate non-reflexive states to their possibility sets. u is defined by $up = \neg kp \wedge \neg k \neg kp$. Using the possibility correspondences, we can build events such as $K[p]$, $\neg K[p]$, $K\neg K[p]$, $\neg K \neg K[p]$ and $U[p]$. It is easy to see that for some states (exactly in all states in which the individual is unaware of an

Figure 2: Information Structure



event) the possibility correspondence is non-reflexive. This causes negative introspection to fail. To see this consider the event $[p]$, i.e., all states in which p obtains. It is easy to see that $(up, jq, p, q) \in \neg K[p]$. Since $(up, jq, p, q) \notin K\neg K[p]$, negative introspection fails. Moreover, also $K\left(S^{\uparrow}_{\{p,q\}}\right) = S^{\uparrow}_{\{p,q\}}$ fails since for instance $(up, jq, p, q) \in \Sigma_{S_{\{p,q\}}}$ but $(up, jq, p, q) \notin K\Sigma_{S_{\{p,q\}}}$. However, in this example one obtains all "S4" properties of Proposition 2 as well as all properties of unawareness of Proposition 3.

3 Discussion

We build a state-space model with unawareness without an explicit use of the modal syntax within the semantic structure. For a wide audience, this should be helpful for developing various applications. However, there is a canonical model for our structure which we elaborate formally in a companion work. Such canonical model is built starting from sets of maximally consistent sets of formulas in sub-languages containing subsets of atomic propositions. This makes the interpretation of "$\preceq$" as relating "expressive power" natural since a language is "more expressive" than a sub-language containing only a subset of former's atomic propositions. An alternative but very similar interpretation of $\preceq$ is given by sets of facts describing the states in Example 1. Further alternative but non-formal motivations for $\preceq$ can be found for example in cognitive psychology. There it is suggested that perception is guided for instance by mental models or categorization. A mental model is an individual representation of the world (Johnson-Laird, 1983). Mental models may differ in terms of comprehensiveness motivating an order relation of expressive power. Categorization is suggested to guide a human's perception by filtering observations (Goldstone and Kersten, 2002). These filters may differ in their filtration motivating too an order relation of expressive power. In this sense, our structure may be useful to model bounded perception in (interactive) decision making.

The special structure of events emerges from the complete lattice of spaces. The meet of all spaces in this lattice may be a space containing a single state only, the empty set. Note, however that this space is not an empty set. For each space $S \in \mathcal{S}$ there is the universal event S and the vacuous event $\emptyset^S$ corresponding respectively to a tautology and a logical contradiction phrased with the expressive power available in S. There might be, and in general there are, subsets of the union of all spaces, which are not events. Negation of an event is typically a proper subset of the relative complement with respect to the union of all spaces. Thus an event is neither "true" nor "false" exactly in spaces that can not express this event, i.e., states that belong neither to the event nor to its negation.

The possibility correspondence satisfies our generalized reflexivity for all states but not necessarily reflexivity. In analogy to dynamic systems, one may distinguish between "stationary" states and "transient" states. In former, the state is an element of the possibility set at that state. In latter, the possibility set at that state resides in a space with less expressive power than the state's space. At each transient state the individual is unaware of some event although the event can be expressed in the state's space.

We hope that our model will be helpful to develop various applications. Particularly, we think about applications of unawareness to agreement, speculative trade, Dutch books, consumption behavior, emergence of novelty, insurance, inconceivable contingencies in (incomplete) contracting etc. This shall be left to further research.

References

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

[2] Dekel, Eddie, Lipman, Bart and Aldo Rustichini (1998). Standard state-space models preclude unawareness, *Econometrica* **66**, 159-173.

[3] Ewerhart, Christian (2001). Heterogeneous awareness and the possibility of agreement, mimeo., University of Mannheim.

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

[5] Goldstone, Robert and Alan Kersten (2002). Concepts and categorization, forthcoming in: Healy, A.F. and R.W. Proctor (eds.). *Comprehensive handbook of psychology, Vol. IV: Experimental psychology*, New York: Wiley.

[6] Halpern, Joseph Y. (2001). Alternative sematics for unawareness, *Games and Economic Behavior* **37**, 321-339.

[7] Hart, Sergiu, Heifetz, Aviad and Dov Samet (1996). "Knowing whether", "knowing that", and the cardinality of state spaces, *Journal of Economic Theory* **70**, 249-256.

[8] Johnson-Laird, P. N. (1983). *Mental models - Towards a cognitive science of language, inference and consciousness*, Cambridge et al.: Cambridge University Press.

[9] Modica, Salvatore and Aldo Rustichini (1999). Unawareness and partional information structures, *Games and Economic Behavior* **27**, 265-298.