Epistemic Self-Analysis and Epistemic Bounded Rationality*

Konrad Grabiszewski[†]

2015

Abstract

I consider an agent who conducts epistemic self-analysis. That is, she analyzes her model of knowledge and determines whether she satisfies the three knowledge axioms (the Truth Axiom, the Positive Introspection Axiom, and the Negative Introspection Axiom). I show that the consequence of epistemic self-analysis is that the agent must know that she satisfies all three knowledge axioms. I rely on using the "knowing whether" operator.

Keywords: Knowledge; Epistemic Self-analysis; Knowledge Axioms

JEL classification: D80; D82

---

*I thank Adam Brandenburger, Adam Dominiak, Amanda Friedenberg, Joseph Halpern, Jean-Philippe Lefort, Erkut Ozbay, David Pearce, David Ross, Vasiliki Skreta, two referees, and audiences at New York University, Instituto Tecnológico Autónomo de México, University of Alicante, and National University of Singapore for their comments. Support from the Asociación Mexicana de Cultura A.C. is gratefully acknowledged.

[†]Address: Department of Business Administration, ITAM, Av. Camino a Sta. Teresa No. 930, Col. Magdalena Contreras, C.P. 10700, Mexico D.F., Mexico and Department of Economics, University of Miami, Coral Gables, FL 33124, USA; konrad.grabiszewski@gmail.com

1 Introduction

There are three properties of knowledge which are of particular interest for philosophers, logicians, and economists: the Truth Axiom, the Positive Introspection Axiom, and the Negative Introspection Axiom. I will refer to these three properties as *the knowledge axioms*. If the agent's knowledge satisfies all three knowledge axioms, then the agent is epistemically rational; otherwise, the agent exhibits *epistemic bounded rationality* (see Geanakoplos (1989), Grabiszewski (2015), and Rubinstein (1998)). I ask whether it is possible for an agent to know that she violates these knowledge axioms. I find that the answer is negative: if the agent analyzes her knowledge (epistemic self-analysis), then she must know that she is epistemically rational.

My analysis is based on the standard model of knowledge, introduced in Aumann (1976), which is denoted by (Ω, P), where Ω is a state space and P is a possibility correspondence. I consider an agent, Ann, who conducts epistemic self-analysis. That is, she determines whether or not she satisfies the knowledge axioms. I assume that the analysis is conclusive: after the analysis of a given axiom, Ann knows either that it holds or that it is violated, and she has no doubts about whether the axiom is true. Her certainty about whether the claim is true is the hallmark of a conclusive analysis.

I argue that the consequences of epistemic self-analysis can be expressed using the "knowing whether" operator (see Aloni et al. (2013), Hart et al. (1996), Heifetz and Samet (1993), Heifetz and Samet (1999), Fan et al. (2013)). I apply this operator to the knowledge axioms. I show that if Ann knows *whether* the knowledge axioms hold, then she knows *that* these axioms do in fact hold.

My paper is related to Gossner and Tsakas (2012), which shows that if the agent satisfies the knowledge axioms on non epistemic propositions, then the agent is epistemically rational (that is, satisfies the knowledge axioms on all propositions). This result is important for a researcher who tests whether or not the agent is characterized by epistemic bounded rationality: Gossner and Tsakas (2012) proves that it is enough for the researcher to test the agent's knowledge only on primitive propositions. I ask a new question: What if the agent acts like a researcher? Epistemic self-analysis can be interpreted as the agent acting like a researcher. On the one hand, there is Ann-the-agent, whose knowledge is depicted by the model of knowledge. On the other hand, there is Ann-the-researcher who analyzes that model. I show that if Ann acts like a researcher, then Ann-the-agent cannot know that she violates the knowledge axioms.

In Section 2, I briefly review the standard model of knowledge, its basic properties, and two knowledge operators: "knowing that" and "knowing whether."[1] Section 3 contains the main results. In Section 4, I present two implications of my results.

2 Standard Model of Knowledge and Knowledge Operators

The standard model of knowledge (Aumann (1976)) consists of a state space Ω and a possibility correspondence $P : \Omega \to 2^{\Omega} \setminus \emptyset$. An element of Ω is called a state and is denoted by ω. An event E is a subset of Ω. The complement of E, $\Omega \setminus E$, is denoted by $\neg E$. Ann's knowledge is represented by a possibility correspondence. If ω is a true state, then $P(\omega)$ is a subset of Ω that contains all states Ann considers to be possible.

The agent's knowledge is defined formally in the following way: Ann knows event E at state ω if $P(\omega)$ is a subset of E. This set-theoretical definition of knowledge leads to the construction of a knowledge operator K, called "knowing that." K is defined on subsets of Ω as $K(E) = \{\omega : P(\omega) \subset E\}$. Whenever it is not confusing, I will write $K(E)$ as KE. Note that KE is an event: a collection of all states at which Ann knows E. Every operator K defined from P satisfies the MC property, which says that Ann knows both events E and F if and only if she knows E and she knows F. Formally, $K(E \cap F) = KE \cap KF$.

There are three properties of K which I am calling the knowledge axioms. The Truth Axiom holds if $KE \subset E$ for each event E, the Positive Introspection Axiom holds if $KE \subset KKE$ for each event E, and the Negative Introspection Axiom holds if $\neg KE \subset K\neg KE$ for each event E. Note that the Truth Axiom and the Negative Introspection Axiom imply the Positive Introspection Axiom (see, for instance, Chapter 3 in Rubinstein (1998)).

[1] More extensive and advanced analysis of this model is provided by, among others, Dekel and Gul (1997), Fagin et al. (1995), Geanakoplos (1989), and Rubinstein (1998). In addition, knowledge modeling and its applications in economics are surveyed in Brandenburger (1992), Geanakoplos (1992), Reny (1992), and Samuelson (2004).

If the Truth Axiom does not hold, then it might be more appropriate to say that K stands for "believing that" rather than "knowing that." As noted in Fagin et al. (1995), "the Truth Axiom has been taken by philosophers to be the major one distinguishing knowledge from belief. Although you may have false beliefs, you cannot know something that is false." However, in order to keep terminology simple, I will consider K to stand for "knowing that," irrespective of the Truth Axiom holding.

Beside the "knowing that" operator, there also is the "knowing whether" operator, denoted by J and defined as $JE = KE \cup K\neg E$. "Ann knows whether E" means that she either knows that E is true or she knows that $\neg E$ is true. She is not allowed to not know — she must know which one is true. Note that Ann knows whether E if and only if she knows whether $\neg E$.

3 Main Results

In this note, I am concerned with epistemic self-analysis in the sense that Ann verifies whether or not she satisfies the knowledge axioms. Since the relevant events that Ann analyzes are "Ann satisfies the Truth Axiom," "Ann satisfies the Positive Introspection Axiom," and "Ann satisfies the Negative Introspection Axiom," first, it is necessary to describe these three events as subsets of Ω.

Fix an event E and a state ω. I say that the Truth Axiom holds for event E at state ω if $\omega \in \neg KE \cup E$. If the axiom is true at ω for every event, then I say that the axiom holds at ω. If, in addition, the axiom is true for every state and every event, then I simply say that the axiom holds. Note that this new description of the Truth Axiom ($\omega \in \neg KE \cup E$ for each ω and each E) is equivalent to the standard description presented in the previous section ($KE \subset E$ for each E). This is because if A and B are two subsets of Ω, then $A \subset B$ if and only if $\omega \in \neg A \cup B$ for each ω.

Regarding the Positive Introspection Axiom, I say that it holds if $\omega \in \neg KE \cup KKE$ for each ω and each E. And the Negative Introspection Axiom holds if $\omega \in KE \cup K\neg KE$ for each ω and each E.

Next, I argue that the "knowing whether" operator captures Ann analyzing these events.

I explain in detail the consequence of Ann's verifying the Truth Axiom. Fix an event E and a state ω. Recall that the Truth Axiom holds for E at ω if $\omega \in \neg KE \cup E$, and is violated if $\omega \notin \neg KE \cup E$ (i.e., $\omega \in KE \cap \neg E$). The consequence of Ann's epistemic self-analysis is that she knows (at ω) that either the Truth Axiom holds for E at ω or the Truth Axiom does not hold for E at ω. If the former is true, then $\omega \in K(\neg KE \cup E)$, and if the latter is true, then $\omega \in K(KE \cap \neg E)$. Consequently, it must be true that $\omega \in K(\neg KE \cup E) \cup K(KE \cap \neg E)$; that is, at ω, Ann knows whether the Truth Axiom holds for E at ω. However, at ω, Ann does not limit her self-analysis to one event E; rather, she analyzes all events. Consequently, I say that, at a given state ω, Ann's epistemic self-analysis implies that $\omega \in K(\neg KE \cup E) \cup K(KE \cap \neg E)$ for each event E; that is, at ω, Ann knows whether the Truth Axiom holds. Finally, Ann conducts the epistemic self-analysis at every state. Hence, it is true that $\omega \in K(\neg KE \cup E) \cup K(KE \cap \neg E)$ for each event E and each state ω; that is, Ann knows whether the Truth Axiom holds.

The same reasoning applied to the Positive and Negative Introspection Axioms allows us to conclude that a conclusive epistemic self-analysis implies that (i) Ann knows whether the Positive Introspection Axiom holds ($\omega \in K(\neg KE \cup KKE) \cup K(KE \cap \neg KKE)$ for each event E and each state ω), and, (ii) that Ann knows whether the Negative Introspection Axiom holds ($\omega \in K(KE \cup K\neg KE) \cup K(\neg KE \cap \neg K\neg KE)$ for each event E and each state ω).

Proposition 3.1 *Assume that the agent knows whether all three knowledge axioms hold.*

1. *If the Truth Axiom holds, then the agent knows that she satisfies all three knowledge axioms.*
2. *If the Positive Introspection Axiom holds, then the agent knows that she satisfies all three knowledge axioms.*
3. *If the Negative Introspection Axiom holds, then the agent knows that she satisfies all three knowledge axioms.*

Proof: Assuming that a knowledge axiom holds means that Ann knows that the axiom holds. Hence, if a knowledge axiom is assumed, then it is necessary to verify only that Ann knows that the remaining two knowledge axioms also hold.

In order to prove that Ann knows that a knowledge axiom holds, I rely on proof by contradiction. By assumption, Ann knows whether all three knowledge axioms hold. Hence, if she does not know that a knowledge axiom holds, it must be true that she knows that the axiom does not hold. Consequently, it is enough to show that it is impossible for her to know that a knowledge axiom does not hold.

1. Assume the Truth Axiom. (a) Suppose that Ann does not know that the Positive Introspection Axiom holds. Consequently, there must be some ω and E such that $\omega \in K(KE \cap \neg KKE)$. Due to the MC property, $K(KE \cap \neg KKE) = KKE \cap K\neg KKE$. The Truth Axiom implies that $K\neg KKE \subset \neg KKE$. Consequently, $\omega \in KKE \cap \neg KKE$, which is a contradiction. (b) Suppose that Ann does not know that the Negative Introspection Axiom holds. Consequently, there must be some ω and E such that $\omega \in K(\neg KE \cap \neg K\neg KE)$. Due to the MC property, $K(\neg KE \cap \neg K\neg KE) = K\neg KE \cap K\neg K\neg KE$. The Truth Axiom implies that $K\neg K\neg KE \subset \neg K\neg KE$. Consequently, $\omega \in K\neg KE \cap \neg K\neg KE$, which is a contradiction.

2. Assume the Positive Introspection Axiom. (a) Suppose that Ann does not know that the Truth Axiom holds. Consequently, there must be some ω and E such that $\omega \in K(KE \cap \neg E)$. Due to the MC property, $K(KE \cap \neg E) = KKE \cap K\neg E$. The Positive Introspection Axiom implies that $K\neg E \subset KK\neg E$. Consequently, $\omega \in KKE \cap KK\neg E$. But this is a contradiction, since both $\omega \in KKE$ and $\omega \in KK\neg E$ must be true. The former implies that $P(\omega) \subset KE$, while the latter implies that $P(\omega) \subset K\neg E$. Since KE and $K\neg E$ are disjoint sets, both statements cannot be simultaneously true. (b) Suppose that Ann does not know that the Negative Introspection Axiom holds. Consequently, there must be some ω and E such that $\omega \in K(\neg KE \cap \neg K\neg KE)$. Due to the MC property, $K(\neg KE \cap \neg K\neg KE) = K\neg KE \cap K\neg K\neg KE$. The Positive Introspection Axiom implies that $K\neg KE \subset KK\neg KE$. Consequently, $\omega \in KK\neg KE \cap K\neg K\neg KE$. That is, Ann simultaneously knows that event $F = K\neg KE$ holds and that it does not hold, which is a contradiction.

3. Assume the Negative Introspection Axiom. (a) Suppose that Ann does not know that the Truth Axiom holds. Consequently, there must be some $\omega \in K(KE \cap \neg E)$. Due to the MC property, $K(KE \cap \neg E) = KKE \cap K\neg E$. Note that the Negative Introspection Axiom implies that either $\omega \in KE$ or $\omega \in K\neg KE$. The former contradicts $\omega \in K\neg E$, and the latter contradicts $\omega \in KKE$. (b) Since the Truth Axiom and the Negative Introspection

Axiom imply that the Positive Introspection Axiom is a tautology, she therefore knows that she satisfies the Positive Introspection Axiom as well. ■

4 Implications

Economists often (implicitly) assume that the model they construct is commonly known to the agents around whom the model is constructed.[2] However, if the agent knows the model, then she acts as a researcher and, consequently, is able to analyze the model. As already mentioned in the Introduction, epistemic self-analysis can be interpreted as the agent acting like a researcher. Hence, if the model in question is the standard model of knowledge and the agent knows her model of knowledge, then, as proved in Section 3, the agent knows that she satisfies all three knowledge axioms. That is, the usual assumption of the agent's knowing the model is not an innocent assumption. I discuss two implications of this observation.

The Aumann Agreement Theorem (Aumann (1976)) shows that if the agents satisfy all three knowledge axioms and that their priors are the same, then their posteriors are also the same if these posteriors are commonly known. Aumann also added "the implicit assumption that the information partitions $\mathscr{P}_1$ and $\mathscr{P}_2$ are themselves common knowledge." But is this assumption necessary? Samet (1990) shows that the agents will not agree to disagree, even if the agents violate the Negative Introspection Axiom. The result of my note indicate that the agents in Samet's paper do not know their own models of knowledge. Consequently, the possibility correspondences need not be known for the agreement theorem to hold.

The Milgrom-Stokey No-trade Theorem (Milgrom and Stokey (1982)) shows that if the risk averse agents share a common prior and satisfy all three knowledge axioms, then they will not trade. Morris (1994) shows that the agents will trade if the common-prior assumption is removed. Geanakoplos (1989) shows that trade occurs when the knowledge axioms are violated. Note that the agents' knowledge is included in the model's setup which, as the results of my note indicate, implies that the agents do not violate the knowledge axioms. This, with common priors, takes us back to the Milgrom-Stokey construction. Consequently, if the agents do trade, then it must be because of their having different priors (Morris (1994))

[2]For example, Myerson (1991) writes that "whatever model of the game we may study, we must assume that the players know this model."

rather than their having violated the knowledge axioms (Geanakoplos (1989)).

References

Aloni, M., P. Égré, and T. de Jager (2013): "Knowing Whether A or B," *Synthese*, 190, 2595–2621.

Aumann, R. J. (1976): "Agreeing to Disagree," *Annals of Statistics*, 4, 1236–1239.

Brandenburger, A. (1992): "Knowledge and Equilibrium in Games," *Journal of Economic Perspectives*, 6, 83–101.

Dekel, E. and F. Gul (1997): "Rationality and Knowledge in Game Theory," in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, ed. by D. M. Kreps and K. Wallis, Cambridge, UK: Cambridge University Press, vol. 1, 87–172.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995): *Reasoning about Knowledge*, Cambridge, MA: The MIT Press.

Fan, J., Y. Wang, and H. van Ditmarsch (2013): "Knowing Whether," *arXiv:1312.0144v3*.

Geanakoplos, J. (1989): "Game Theory Without Partitions, and Applications to Speculation and Consensus," *Cowles Foundation for Research in Economics*, No. 914.

——— (1992): "Common Knowledge," *Journal of Economic Perspectives*, 6, 53–82.

Gossner, O. and E. Tsakas (2012): "Reasoning-based Introspection," *Theory and Decision*, 73, 513–523.

Grabiszewski, K. (2015): "Rationalizing Epistemic Bounded Rationality," *Theory and Decision*, 78, 629–637.

Hart, S., A. Heifetz, and D. Samet (1996): ""Knowing Whether," "Knowing That," and The Cardinality of State Space," *Journal of Economic Theory*, 70, 249–256.

Heifetz, A. and D. Samet (1993): "Universal Partition Structures," *IIBR Working Paper, Tel Aviv University*, 26.

——— (1999): "Hierarchies of Knowledge: An Unbounded Stairway," *Mathematical Social Sciences*, 38, 157–170.

Milgrom, P. and N. Stokey (1982): "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26, 17–27.

Morris, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62, 1327–1347.

Myerson, R. B. (1991): *Game Theory*, Cambridge, MA: Harvard University Press.

Reny, P. J. (1992): "Rationality in Extensive-Form Games," *Journal of Economic Perspectives*, 6, 103–118.

Rubinstein, A. (1998): *Modeling Bounded Rationality*, Cambridge, MA: The MIT Press.

Samet, D. (1990): "Ignoring Ignorance and Agreeing to Disagree," *Journal of Economic Theory*, 52, 190–207.

Samuelson, L. (2004): "Modeling Knowledge in Economic Analysis," *Journal of Economic Literature*, 42, 367–403.