# Axioms of Distinction in Social Software

Vincent F. Hendricks
Department of Philosophy and Science Studies
Roskilde University, Denmark
vincent@ruc.dk

June 7, 2008

> 'Over a ten year period starting in the mid 90's I became convinced that all these topics – game theory, economic design, voting theory – belonged to a common area which I called Social Software.' —Rohit Parikh, [Parikh 05]: p. 252

Around the turn of the millenium Rohit Parikh launched a new important program on the borderline of computer science and epistemology. 'Social software' has aquired two meanings recently: One in which social software denotes various web-based software programs of a 'social nature' from *Facebook* to *Wikipedia* and a meaning, now canonical, according to an entry exactly in *Wikipedia*, in which social software

> ... studies the procedures of society whether elections, conferences etc. as analogous to computer programs to be analyzed by similar logical and mathematical tools.

The formal tools of the trade are described in Parikh's seminal papers [Parikh 01], [Parikh 02] and subsequenly spelled out as "(1) logic of knowledge, (2) logic of games, and (3) game theory and economic design." [Pacuit & Parikh 07]: p. 442.

These tools are also common to a related new trade called *formal epistemology* which also by a recent entry in *Wikipedia* is characterized as

> ... a subdiscipline of epistemology that utilizes formal methods from logic, probability theory and computability theory to elucidate traditional epistemic problems.

Social software and formal epistemology are both interdisciplinary approaches to the study of agency and agent interaction but are not quite to be confused with yet a new trend in philosophy recently *social epistemology* [Goldman 99]. While social software and formal epistemology share the same toolbox, social epistemology pays stronger homage to the standard philosophical methodology of conceptual analysis and intuition pumps.

Although sounding derogatory, conceptual analysis is not necessarily a bane as long as it is regimented by formal structure. In social software, the formal structure is provided by the aforementiond tools. While modelling agency and interaction certain methodological distinctions may have to be explicitly observed exactly in order to make conceptual sense of the use of logic of knowledge, logic of games, game theory and economic design in social software.

No better occassion to review but a few of these methdological distinctions and their conceptual impact in light of the 70th birthday of the founding architect of social software in the important sense—Rohit Parikh.

## 1 Perspectives, Agents and Axioms

Contemporary epistemology often draws a distinction between descriptive and normative theories of knowledge. There is a similar distinction in moral philosophy between descriptive and normative ethics. The former attempts to describe actual moral behavior while the latter sets the standards for correct moral conduct.

Similarly, descriptive epistemologies account for actual epistemic practice while normative ones are to prescribe rules of inquiry in terms of mechanisms for avoiding error and gaining truth, truth-conducive justification criteria, learning and winning strategies, procedures for revising beliefs etc. The distinction is sometimes blurred by the fact that while describing actual epistemic practice one may have to define various notions like, say, knowledge itself, justification and reliability inviting normative aspects.

Both descriptive and normative epistemologies usually subscribe to the common premiss that epistemic practice of agents by and large is 'rational'. What separates the two stances is whether epistemology is simply to describe this very practice or try to optimize it. Irrational epistemic behavior would be to follow some practice which *a priori* may be demonstrated to be en route to error (when this very practice is an available course of conduct to the agent in the environment) [Kelly 96], [Hendricks 01]. It is not necessarily irrational on the other hand not to follow some prescription if the natural epistemic milieu sets the standards for what the agent is able to do and this prescription is not among them. The local epistemic circumstances may for one reason or the other bar the agent in question from choosing the best means for an end. Constraints could even be such that they reward 'irrational' behavior. Calling such epistemic situations irrational would undermine the common premiss which the two approaches subscribe to. Not only may the environment limit the agent's behavior, other agents may as well. This is for instance illustrated by game theory's distinction between cooperative and non-cooperative games.

Sometimes agents would be able to have more knowledge than they actually have if they were not tied up in their local epistemic theater. Then they could freely pursue the optimal means for obtaining some desirable result whether truth, epistemic strength or winning in some other sense. One may then rightfully ask why epistemologists sometimes are in the business of means-ends prescriptions that no local agent is able to meet. There are two with each other related answers to this question:

- As epistemologists we are not only in the business of ascribing ourselves knowledge, but equally much in the business of ascribing knowledge to other agents. Lewis has pointed out that there is a significant difference between one agent ascribing himself knowledge in his local epistemic situation, and us ascribing him knowledge given the situation we are in. The two situations do not always coincide. First and third persons do not share the same real world in many contexts. There are rules to follow under knowledge attribution to oneself and others to know what we think we know.

- Rather principled information about what it would take to solve the epistemic task at hand, than no information at all. Epistemology is about whether knowledge is possible and about what agents can and cannot know insofar knowledge is possible. The problem is that it is not always clear from within whether something is knowable or not. One recurs to a perspective from without for a principled answer which may then spill over into the local circumstances.

According to Lewis [Lewis 96], the agent may actually know more than we are able to ascribe to him. On his account this is due to the fact that the attribution of knowledge is highly sensitive to which world is considered actual and for whom in a given context. An agent in his environment is more likely to be aware of what the relevant possibilities are given the world considered actual by him than the knowledge ascriber standing by him or even outside. Lewis refers to these two stances as a *first* versus a *third* person perspective on inquiry.

Observe that an agent is free to be prescribe recommendations for himself to follow as long as the means suggested are available to him where he is. Outsiders may also freely prescribe recommendations for the agent inside as long as they are available to the agent in question. If the outsider decides to prescribe a course of conduct to solve an epistemic problem for the agent in the environment unavailable to the agent then the situation changes. Emphasized is then what it would take to solve the epistemic problem regardless of whether the agent is capable of actually performing the action(s) it would take .

Normative/descriptive, first person versus third person perspectives are not mutually exclusive distinctions. The distinction between descriptive and normative theories of knowledge together with a modified version of Lewis' first and third person perspective dichotomy are subsumed in the following two formulations:

> First person perspective—A perspective on scientific inquiry is **first person** if it considers what an agent can solve, can do or defend considering the available means for an end given the epistemic environment he is sunk into

Third person perspective—A perspective on scientific inquiry is **third person** if it considers what an agent could solve, could do or defend considering the best means for an end independently of the epistemic environment he is sunk into

In criticizing some epistemological position, whether mainstream or formal, without noticing that the criticism is based on a third person perspective and the position advocated is first person may again turn out to be criticizing an apple for not being an orange [Hendricks 06]. The dichotomy has a significant bearing on the general epistemological and conceptual plausibility given the formal tools utilized by social software in modelling agent and agent interaction.

## 2 Setting up the Matrix

What is being studied from the first and the third person perspective respectively using the logics of knowledge (and games) in social software are

- an agent or multiple agents in concert, and
- various epistemic axioms and systems characterizing the knowledge of one or more agents.

The task now is to tell a plausible epistemological story about the axioms valid while modelling, say, one agent, from a first person perspective and so forth for the remaining possibilities in the matrix

| | First person | Third person |
|---|---|---|
| One agent | $x$ | $x$ |
| Multiple agents | $x$ | $x$ |

for $x \in \{\mathsf{T},\mathsf{K},\mathsf{D},\mathsf{4},\mathsf{5}\}$[1]

Here is an example of an epistemological story told about modelling one agent from a first person perspective all the way to **S4**: Hintikka stipulated that the axioms or principles of epistemic logic are conditions descriptive of a special kind of general (strong) *rationality* for a single agent and on a first person perspective [Hintikka 62]. The statements which may be proved false by application of the epistemic axioms are not inconsistent meaning that their truth is logically impossible. They are rather rationally ‘indefensible’. Indefensibility is fleshed out as the agent’s epistemic laziness, sloppiness or perhaps cognitive incapacity whenever to realize the implications of what he in fact knows. Defensibility then means not falling victim of ‘epistemic neglience’ as Chisholm calls it. The notion of indefensibility gives away the status of the epistemic axioms and logics. Some epistemic statement for which its negation is indefensible is

[1] Only these 5 canonical axioms are considered here, others may be discussed as well – see [Hendricks 06].

called ‘self-sustaining’. The notion of self-sustenance actually corresponds to the concept of validity. Corresponding to a self-sustaining statement is a logically valid statement. But this will again be a statement which is rationally indefensible to deny. So in conclusion, epistemic axioms are descriptions of rationality.

There is evidence to the effect that Hintikka early on was influenced by the autoepistemology of Malcolm [Malcolm 52] and took, at least in part, their autoepistemology to provide a philosophical motivation for epistemic logic. There is an interesting twist to this motivation which is not readily read out of autoepistemology. Epistemic axioms may be interpreted as principles describing a certain strong rationality. The agent does not have to be aware of this rationality, let alone able to immediately compute it from the first person perspective as Hintikka argues when it comes to axiom $\mathsf{K}$:

> In order to see this, suppose that a man says to you, ‘I know that $p$ but I don’t know whether $q$’ and suppose that $p$ can be shown to entail logically $q$ by means of some argument which he would be willing to accept. Then you can point out to him that what he says he does not know is already implicit in what he claims he knows. If your argument is valid, it is irrational for our man to persist in saying that he does not know whether $q$ is the case. [Hintikka 62], p. 31.

The autoepistemological inspiration is vindicated while Hintikka argues for the plausibility of $\mathsf{4}$ as a governing axiom of his logic of knowledge as he refers to Malcolm:

> This is especially interesting in view of the fact that Malcolm himself uses his strong sense of knowing to explain in what sense it might be true that whenever one knows, one knows that one knows. In this respect, too, Malcolm’s strong sense behaves like mine. [Hintikka 70], p. 154.

Besides the requirement of closure and the validity of the $\mathsf{4}$, axiom $\mathsf{T}$ is also valid to which Malcolm would object. A logic of autoepistemology is philosophically congruent with Hintikka’s suggestion for a **S4** epistemic logic describing strong rationality from a first person point of view for a singular agent.

The key debate of whether epistemic axioms are plausibly describing agenthood or not seems much to depend on whether one subscribes to a first or a third person perspective on inquiry. Given an autoepistemological inspiration epistemic axioms describe a first person knowledge operator as Hintikka suggested. If epistemic axioms are describing *implicit knowledge* as Fagin et al. suggest [Fagin et al. 95], then what is being modelled is what follows from actual knowledge independently of agent computations. Agents can on this third person perspective not be held actually responsible for failing to exercise some reflective disposition. Closure principles may be problematic from the point of view of the agent, not necessarily from point of view of the ones studying the agent third person. Logical omniscience as a consequence of the epistemic axioms is a problem from a first person perspective but not necessarily from a third person perspective.

We are not going to fill in all the cells of the matrix with epistemological stories, just discuss one additional axiom which is tricky for knowledge, games and economic design, agents and any point of view.

## 3 Negative Introspection of Knowledge

One of the most celebrated motivations for the plausibility of axiom 5, or the axiom of negative introspection, is a closed-world assumption in data-base applications [Fagin et al. 95]: An agent examining his own knowledge base will be led to conclude that whatever is not in the knowledge base he does not know and hence he will know that he does not. This is a first person motivation, but in the same breath, the argument for dodging logical omniscience is based on a third-person operative as seen above. So there is some meandering back and forth while arguing for the axiomatic plausibility of agency.

Axiom 5 may seem unrealistically strong for a singular agent in his environment, unless his environment is defined solipsistically, e.g. the closed world assumption. Solipsism is not necessarily a human or a real agent condition but a philosophical thesis; a thesis making idealized sense standing outside looking at the agent in his, admittedly, rather limited epistemic environment. Being a stone-hearted solipsist on a first person basis is hard to maintain coherently as W.H. Thorpe once reported:

> Bertrand Russell was giving a lesson on solipsism to a lay audience, and a woman got up and said she was delighted to hear Bertrand Russell say he was a solipsist; she was one too, and she wished there were more of us.

A reason for adopting the first person perspective and pay homage to axiom 5 for singular agents is that these assumptions provide some nice technical advantages / properties especially with respect to information partition models. There is now also a philosophical basis for doing things in this idealized way—epistemic solipsism and no false beliefs, e.g. infallibilism. Both of these philosophical theses have little to do with logic but plenty to do with the preconditions for studying knowledge from any point of view.

## 4 Negative Introspection in Games

It is more sticky to argue that axiom 5 is reasonable to assume in multi-agent setups. But when game theorists for instance model non-cooperative extensive games of perfect information an **S5** logic of knowledge is used to establish the backward induction equilibrium [Bicchieri 03].

For game theory the untenability of **S5** in multi-agent systems is quite severe. The problem concerns the knowledge of action as Stalnaker has pointed out: It should be possible for a player $\Xi$ to know what a player $\Theta$ is going to do. For instance it should be rendered possible in case $\Theta$ only has one rational choice, and $\Xi$ knows $\Theta$ to be rational, that $\Xi$ can predict what $\Theta$ is going to do. This should not imply however that it is impossible for $\Theta$ to act differently as he has the capacity to act irrationally. In order to make sense of this situation what is needed is a counterfactually possible world such that

1. $\Theta$ acts irrationally, but
2. is incompatible with what $\Xi$ knows.

Now $\Xi$'s prior beliefs in that counterfactual world must be the same as they are in the actual world for $\Theta$ could not influence $\Xi$'s priors beliefs by making a contrary choice (by definition of the game, $\Xi$ and $\Theta$ act independently). Then it has to be the case in the counterfactual world, that $\Xi$ believes he knows something (e.g. that $\Theta$ is irrational) which he in fact does not know. This is incompatible with **S5**.

Additionally, acting in the presence of other agents requires the information to be explicitly available to the agents first person, but it may only be implicitly at the agents' disposal if the over-all model is of implicit knowledge. It is not much help to have the knowledge explicitly available on the third person level if you have to make an informed move on the first person level featuring other agents trying to beat you as you are trying to beat them.

# 5 Negative Introspection in Economics

This discussion of the untenability of **S5** is not in any way linked to a view of inappropriateness of modelling a third person notion of knowledge via the axiom of veridicality $\mathsf{T}$. One may reasonably argue like Stalnaker and Aumann that knowledge requires truth referring to a notion of third-person knowledge. The unintuitive results obtained by Aumann and others indicate that there is something wrong in the information model used by economists, which assumes that agents engaged in economic interactions actually have common knowledge rather than common belief. Thus one can infer that the impossibility of trade can be concluded from assuming that the agents engaged in economic interaction have more powers than they actually have. Once one endows agents with a more realistic epistemic model, it is possible to agree to disagree and trade is made plausible again.

Collins' explanation of what is wrong in Aumann's models is quite plausible. If agents have common belief rather than common knowledge then they cannot share a common prior, a crucial probabilistic assumption in Aumann's seminal paper 'Agreeing to Disagree' [Aumann 76]. An intuitive explanation is provided by Collins:

> Finally, an opponent might challenge my claim that it is belief rather than knowledge that ought to be central to *interactive epistemology*. My response to this is simply to point out that agents, even rational agents,

> can and do get things wrong. This is not a controversial claim, just the commonplace observation that rational agents sometimes have false beliefs. The reason for this is not hard to find. It is because the input on which we update is sometimes misleading and sometimes downright false. To demand that everything an agent fully believes be true is not to state a requirement of rationality but rather to demand that the agent be invariably lucky in the course of her experience. Being completely rational is one thing; always being lucky is another. [Collins 97].

Nothing is here said about what it would actually mean to have knowledge in economic exchanges. Perhaps to be always lucky aside from rational. This entails that the notion of knowledge does require truth in order for it to be intelligible. Collins points out that agents get things wrong all the time, even while being completely rational. Aumann's theorem demonstrate how alien to our everyday endeavors the notion of knowledge is. The notion of rationality can at most require that the agent only holds beliefs that are full beliefs, i.e., beliefs which the agent takes as true from his first person point of view.

Disagreement alone does not suffice for altering anyone's view. Each agent will therefore have some type of acceptance rule that will indicate to him whether it is rational or not to incorporate information. Sometimes the agent might lend an ear to an incompatible point of view for the sake of the argument and this might end up in implementing a change of view. When a network of agents is modeled from the outside, endowing these agents with third-person knowledge (as is customarily done in economic models) seems inappropriate. Be that as it may, if the agents are rational one should assume that their theories are in autoepistemic equilibrium, and this leads to the assumption that the first person views are each one **S5** [**Arló-Costa 98**]. These two things are perfectly compatible, which does not imply that certain type of inputs (the ones that are marked positive by your preferred theory of acceptance) might require perturbing the current autoepistemic equilibrium via additions or revisions. The philosophical results questioning the use of **S5** in interactive epistemology, question the fact that economic agents can be modeled *by the game theorist* as having third person knowledge. These results do not necessarily have a bearing for what one might or must assume about the first-person point of view of each agent.

# 6 Axioms of Distinction

In a recent interview, Parikh takes stock of the contemporary situation in epistemology:

> Currently there is sort of a divide among two communities interested in knowledge. One is the community of epistemologists for whom the 1963 paper by Gettier has created a fruitful discussion. The other community is the community of formal epistemologists who trace their roots back to modal logic and to Hintikka's work on knowledge and belief. There is relatively little overlap between the two communities, but properties of

> the formal notion of knowledge, e.g., positive and negative introspection, have caused much discussion. I would suggest that knowledge statements are not actually propositions. Thus unlike “There was a break-in at Watergate,” “Nixon knew that there was a break-in at Watergate,” is not a proposition but sits inside some social software. It may thus be a mistake to look for a fact of the matter. The real issue is if we want to hold Nixon responsible. [Parikh 05]: 145.

Parikh is right in emphasizing that too much ink perhaps has been spilled over formal properties of knowledge rather than on how knowledge is acquired. Surely, much epistemic and moral success has to do with procedures and processes for getting it right rather than static epistemic definitions and axioms of charaterization.

The two endavours are not mutually exclusively. Axioms are used to a procedural end in agency and agent interaction because much of the (game theoretical) dynamics is dependent on the epistemic powers of the agents given axiomatically. "The real issue is if we want to hold Nixon responsible", or put differently, the real issue is whether there is a reliable (possibly effective) procedure, or strategy, for converging to the correct epistemic result (axiom or verdict) or winning the game against nature or other agents. Formal learning theory or ‘computational epistemology’ as Kelly recently redubbed the paradigm is extensively concerned with the former, social software with the latter. By the end of the day, the two approaches are very similar in their procedural outlook on inquiry and success.

A successful epistemic story told is dependent on

- what the agent(s) can do in the environment,
- given the
  - the axioms describing their epistemic powers, and
  - the procedures of knowledge acquisition,
- pace the first person / third person point of view.

And so to finish off with a question based on the matrix above:

> **Are there axioms of distinction exclusively separating the first-person perspective from the third-person perspective?**

Such an axiomatic separation would hopefully supplement the great epistemological range and plausibility of the tools employed in Rohit Parikh’s seminal and important program, *social software.*

# References

[Arló-Costa 98] Arló-Costa, H. (1998). ‘Qualitative and Probabilistic Models of Full Belief’, *Proceedings of Logic Colloquim’98, Lecture Notes on Logic* **13**, Buss, S., Hajék, P. and Pudlák, R. (eds.), ASL.

[Aumann 76] Aumann, R.J. (1976). ‘Agreeing to Disagree’, *Annals of Statistics* **4**: 1236–1239.

[Bicchieri 03] Bicchieri, C. (1993). *Rationality and Coordination.* New York: Cambridge University Press.

[Collins 97] Collins, J. (1997). ‘How We Can Agree to Disagree,’ Columbia University, version of July 2, 1997, `http://collins.philo.columbia.edu`.

[Fagin et al. 95] Fagin, R., Halpern, J. Y., Moses Y. and Vardi, M. Y. (1995). *Reasoning about Knowledge.* Cambridge: MIT Press.

[Goldman 99] Goldman, A.I. (1999). *Knowledge in a Social World.* New York: Oxford University Press.

[Hendricks 01] Hendricks, V. F. (2001). *The Convergence of Scientific Knowledge: A View from the Limit.* Trends in Logic: Studia Logica Library Series. Dordrecht: Springer.

[Hendricks 02] Hendricks, V. F. (2003). ‘Active Agents’, *Journal of Logic, Language, and Information*, van Benthem, J. and van Rooy, R. (eds.), vol. **12**, Autumn, no. 4: 469–495.

[Hendricks 06] Hendricks, V.F. *Mainstream and Formal Epistemology.* New York: Cambridge University Press.

[Hendricks & Symons 05] Hendricks, V.F. and Symons, J. (eds.) (2005). *Formal Philosophy.* London & New York: Automatic Press / VIP.

[Hendricks & Hansen 07] Hendricks, V.F. and Hansen, P.G. (eds.) (2007). *Game Theory: 5 Questions.* London & New York: Automatic Press / VIP.

[Hintikka 62] Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell: Cornell University Press.

[Hintikka 70] Hintikka, J. (1970). “Knowing that One Knows’ Revisited’, *Synthese* **21**: 141–162.

[Kelly 96] Kelly, K. (1996). *The Logic of Reliable Inquiry.* New York: Oxford University Press.

[Lenzen 78] Lenzen, W. (1978). *Recent Work in Epistemic Logic*, in *Acta Philosophica Fennica* **30**: 1–219.

[Lewis 96] Lewis, D. (1996). 'Elusive Knowledge', *The Australian Journal of Philosophy* **74**: 549-567.

[Malcolm 52] Malcolm, N. (1952). 'Knowledge and Belief', *Mind* LXI, 242. Reprinted in *Contemporary Readings in Epistemology*, Goodman, M. F. and Snyder, R. A. (eds.). Prentice Hall. New Jersey: Englewood Cliffs (1993): 272–279.

[Parikh 01] Parikh, R. (2001). 'Language as Social Software', in *Future Pasts: The Analytic Tradition in the Twentieth Century Philosophy.* Floyd, J. and Shieh, S. (eds.). New York: Oxford University Press: 339–350.

[Parikh 02] Parikh, R. (2002). 'Social Software', *Synthese*, vol. 132: 187–211.

[Parikh 05] Parikh, R. (2005). 'Interview: 5 Questions on Formal Philosophy', in [Hendricks & Symons 05]:

[Parikh 07] Parikh, R. (2007). 'Interview: 5 Questions on Game Theory', in [Hendricks & Hansen 07]:

[Pacuit & Parikh 07] Pacuit, E. and Parikh, R. (2007). 'Social Interaction, Knowledge and Social Software', in *Interactive Computation: The New Paradigm.* Dina Goldin, Sott Smolka, Peter Wegner (eds.). Dordrecht: Springer 2007, 441-461.

[Stalnaker 96a] Stalnaker, R. (1996). 'Knowledge, Belief and Counterfactual Reasoning in Games', *Economics and Philosophy* **12**: 133–163.