Proposition de symposium

(Version française)

Attitudes de niveaux supérieurs: connaissance, croyance et interaction sociale

Description de l'atelier :

Les *attitudes de niveaux supérieurs* occupent une place de plus en plus importante en philosophie analytique contemporaine. Par attitudes de niveau supérieur nous entendons ici des attitudes, tant cognitives que conatives, portant sur d'autres attitudes. En théorie de la connaissance, le principe d'introspection fait l'objet de débats animés, et ce depuis plusieurs décennies déjà. Avons-nous, ou devrions-nous avoir des croyances correctes à propos de nos propres croyances? Avons-nous connaissance de notre ignorance ? De telles questions ont montré leur pertinence en épistémologie des attitudes individuelles de croyance et de connaissance, mais également pour des notions plus spécifiques comme *la connaissance approximative* et l'*irreprésentation* (*unawareness*). En épistémologie sociale, la notion de *désaccord* est sans doute l'une des plus discutées. Ici les attitudes de niveaux supérieurs sont aussi d'une importance capitale, en particulier les attitudes portant sur les attitudes d'autrui. Est-ce qu'être informé à propos de ce que sait une autre personne constitue pour moi aussi une évidence en faveur de ce qu'elle sait? Devons-nous toujours prendre en compte les croyances d'autrui lorsque nous formons nos propres croyances? Enfin les attitudes de niveau supérieur occupent aussi une place importante en méta-éthique. Pensons seulement aux discussions sur *l'identité* ou sur les *raisons* et la *normativité*, et la place qu'y occupent les notions de désirs à propos des désirs.

Cet atelier a pour but de présenter, de comparer et de relier ces débats et questions contemporaines touchant les attitudes de niveaux supérieurs. L'atelier sera composé de quatre exposés, chacun faisant écho à un champ de recherche particulièrement actif en philosophie analytique contemporaine. Les deux premiers exposés porteront sur les débats contemporains autours de la notion d'introspection en épistémologie individuelle. Le premier exposé vise à comparer deux formes d'ignorance, et corrélativement, deux formes d'introspection négative, mettant en jeu la connaissance et la représentation (*awareness*), respectivement, en puisant à la fois dans les recherches en science cognitive, en philosophie et en logique. Le second se penchera quant à lui sur l'introspection positive, cette fois en relation avec la connaissance approximative, et le principe de marge d'erreur proposé par Williamson pour rendre compte de la fiabilité de la connaissance. Les troisième et quatrième exposés se tourneront vers les enjeux sociaux des attitudes de niveaux supérieurs. Le premier d'entre eux étudiera la relation entre deux modèles mathématiques de formation de consensus. Le second portera sur les notions de raisons et de rationalité en situations d'interaction sociale et montrera, en utilisant des modèles de théorie des jeux, que les attitudes de niveaux supérieures y sont d'une importance capitale. Ensemble, ces quatre exposés esquisseront quatre champs de la philosophie analytique contemporaine ou les attitudes de niveaux supérieurs jouent un rôle-clef.

***

Exposé 1 : L'irreprésentation, l'incertitude et la connaissance de notre ignorance

Que requiert-il de savoir que l'on ne sait pas une chose ? La réponse dépend du type d'ignorance en question. Il existe au moins deux variétés distinctes d'ignorance, à savoir l'ignorance due à un défaut de certitude (l'incertitude) et l'ignorance due à un défaut de représentation (l'irreprésentation, néologisme par lequel nous traduisons le terme d'*unawareness* étudié en épistémologie formelle : voir Franke et de Jager 2010 pour un panorama récent, et Bromberger 1987 pour un article précurseur de cette distinction).

L'incertitude correspond à toutes les situations dans lesquelles nous hésitons entre des possibilités concurrentes que nous sommes cependant capables d'appréhender et d'articuler. Supposons que l'on me demande si Tolkien et l'auteur de *Bilbo le Hobbit*. Une manière pour moi de manquer de connaitre la réponse est d'hésiter entre la réponse positive et la réponse négative. Il se peut que je conçoive deux possibilités incompatibles, l'une selon laquelle Tolkien est bien l'auteur de *Bilbo*, l'autre selon laquelle c'est C.I. Lewis. L'irreprésentation, en revanche, correspond à toutes les situations pour lesquelles nous manquons de savoir une proposition parce que nous n'avons pas les ressources nécessaires pour articuler et nous représenter la proposition en question. Comparons le cas précédent avec celui dans lequel je n'ai jamais entendu parler de Tolkien, ni de *Bilbo le Hobbit*. Dans ce cas mon défaut de connaissance de ce que Tolkien est l'auteur de *Bilbo le Hobbit* n'est pas le résultat d'une compétition entre diverses possibilités ou sources d'évidence, mais c'est plutôt l'effet d'un défaut de conceptualisation des ingrédients élémentaires de la proposition (voir Heifetz et al. 2006, qui caractérisent ainsi l'irreprésentation).

Un aspect important de la distinction entre incertitude et irreprésentation est que l'irreprésentation, au contraire de l'incertitude, n'est pas introspective (voir Dekel et al. 1998, Franke and de Jager 2008). Cela signifie qu'alors que l'incertitude est accessible à la conscience, de sorte que l'on peut savoir qu'on est incertain que p au moment même où l'on éprouve cette incertitude, on ne peut savoir qu'on manque de se représenter que p de façon contemporaine de cet état de premier niveau, car savoir que l'on manque de se représenter que p impliquerait de se représenter le contenu que p. Il est manifestement possible de réaliser que l'on *était* dans un état d'irreprésentation, mais non de savoir que l'on *est* dans un tel état.

La distinction entre incertitude et irreprésentation est au cœur de nombre de travaux récents en épistémologie formelle. Dans cet exposé, l'objet sera de discuter les implications métacognitives de la distinction à la lumière de plusieurs études sur la psychologie du savoir que l'on ne sait pas. Dans une étude classique, Glucksberg et McCloskey (1982) ont notamment proposé un model dual des décisions touchant l'ignorance. Selon eux, les sujets peuvent rendre un verdit d'ignorance de deux manières distinctes face à une question. Soit parce qu'ils ne trouvent aucune information pertinente dans leur mémoire : dans ce cas, les sujets rendent un verdict rapide d'ignorance. Soit parce que les sujets trouvent des informations pertinentes, mais manquent de trouver des informations concluantes pour répondre la question. Pour ces cas là, les résultats de Glucksberg et McCloskey indiquent que les verdicts d'ignorance prennent plus de temps, en accord avec l'hypothèse d'un coût cognitif plus élevé (voir aussi Hampton et al. 2011).

En lien avec ces travaux, l'aspect que nous souhaitons discuter concerne la fiabilité des déclarations d'ignorance que nous produisons. Les cas dans lesquels nous ne trouvons de

l'information pertinente en mémoire mais sommes incertains devraient par là même être des cas pour lesquels il est difficile de donner une estimation fiable de notre degré d'incertitude, et plus généralement, pour lesquels nous sommes enclins surestimer ou sous-estimer notre ignorance. Par opposition à cela, les cas pour lesquels nous ne trouvons aucune information pertinente en mémoire, tels que les cas fondés sur un arrière-plan d'irreprésentation, devraient apparaître comme des cas pour lesquels nous avons une appréciation plus fiable de l'état réel de notre ignorance.

Bibliographie

Bromberger S. (1987) What we don't know when we don't know why. Repr. in Bromberger S. *What we know we don't know*. University of Chicago Press and CSLI Publications. 1992.

Dekel E., Lipman B.L., Rustichini A. (1998). Standard State-Space Models Preclude Unawareness. *Econometrica* 66 (1): 159-173.

Franke M. & de Jager T. (2010). Now that you mention it: Awareness Dynamics in Discourse and Decisions. In A. Benz et al. (eds), *Language, Games, and Evolution*. Lecture Notes in Computer Science.

Glucksberg, S. & McCloskey, M. (1981). Decisions about Ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory* 7: 311-325.6207, 60-91.

Hampton J. A., Aina B., Mathias Andersson J., Mirza H. Z., Parma S. (2011), The Rumsfeld Effect: the unknown unknown. *Journal of Experimental Psychology: Learning Memory & Cognition*, in press.

Heifetz A. & Meier M. & Schipper B. C. (2006). Interactive unawareness. *Journal of Economic Theory* 130: 7894.

***

Exposé 2: Sur le fondement du principe de marge d'erreur

Selon la théorie de la connaissance inexacte de Timothy Williamson (1994), la connaissance perceptuelle satisfait un principe de marge d'erreur dû aux limites des perceptions individuelles. Ce principe donne lieu au paradoxe du sorite épistémique, que Williamson résout en affirmant que la connaissance ne satisfait pas le principe de l'introspection positive.

Plusieurs auteurs, à commencer par Mott (1998), ont critiqué le raisonnement de Williamson en recourant à une modélisation explicite des perceptions. Ils ont montré que même si les perceptions sont imprécises, la possibilité pour un individu de réaliser des inférences à partir de ses perceptions et de la connaissance de ses propres limites perceptuelles peut conduire à une connaissance qui ne satisfait pas le principe de marge d'erreur. Williamson (2000) répond que cet argument est invalide parce qu'il suppose que les individus possèdent une connaissance parfaite, et non inexacte, de leurs limites perceptuelles.

Cette objection est donc fondée sur des considérations relatives à l'inexactitude de la connaissance d'ordre supérieur, c'est-à-dire aux limites des perceptions sur les limites des perceptions.

Notre contribution vise à reprendre cette controverse à partir d'une modélisation explicite des limites perceptuelles d'ordre supérieur. Nous montrons que pour une certaine classe de signaux, le principe de marge d'erreur pour la connaissance perceptuelle n'est valide que dans

la mesure où il est satisfait à tous les niveaux supérieurs. Mais pour éviter une régression à l'infini, il est nécessaire de supposer l'existence d'un niveau primitif de connaissance, qui ne résulte pas de perceptions. Une modélisation complète des perceptions conduit donc à mettre en cause l'affirmation de Williamson selon laquelle le principe de marge d'erreur peut être justifié par des considérations sur les limites perceptuelles. En outre, cette modélisation peut conduire à rejeter le principe de marge d'erreur pour la connaissance perceptuelle.

Bibliographie

Dutant J. (2007). Inexact Knowledge, Margin-for-Error and Positive Introspection, in D. Samet (ed.), Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK XI), Louvain-la-Neuve: Presses Universitaires de Louvain-la-Neuve, 118-124.

Dokic J. and Egré P. (2009). Margin for error and the transparency of knowledge, Synthese, 166(1), 1-20.

Mott P. (1998). Margins for Error and the Sorites Paradox. The Philosophical Quarterly, 48(193), 494-504.

Williamson T. (1994). Vagueness. London and New York: Routledge.

Williamson T. (2000). Margins for Error: a Reply. The Philosophical Quarterly, 50(198), 76-81.

***

Exposé 3: Consensus et information d'ordre supérieur

L'absence de consensus, en tant que propriété des croyances de premier ordre, et le respect, en tant que propriété des croyances d'ordre supérieur, semblent parfaitement compatibles. Deux personnes peuvent être en désaccord tout en ne considérant pas l'opinion de l'autre comme dépourvue de valeur. Néanmoins, à y regarder de plus près, les choses ne sont pas si simples : accorder de la valeur à une opinion devrait nous conduire à nous en rapprocher, de sorte qu'une forme de consensus potentiel pourrait être cachée dans le respect. Le but de cet exposé sera de parvenir à une meilleure compréhension des scénarios dans lesquels le respect est susceptible de conduire au consensus, en étudiant les relations entre consensus et information d'ordre supérieur.

Lehrer et Wagner (1981) ont proposé un modèle formel des consensus potentiels en caractérisant les conditions sous lesquelles un consensus actuel serait atteint à la limite d'une suite de révisions des croyances individuelles fondées sur le respect mutuel. Dans leur cadre, la représentation des croyances est en partie non-standard : les croyances de premier ordre sont représentées par des probabilités subjectives, mais les croyances d'ordre supérieur sont représentées par des poids que les agents accordent aux croyances de premier ordre des autres agents. A chaque étape, les agents mettent à jour leur croyance sur la base d'une moyenne pondérée par ces poids des opinions dans la communauté. Lehrer et Wagner montrent qu'il suffit que tous les poids soient positifs pour qu'il y ait convergence vers le consensus.

Faut-il alors considérer que le modèle LW montre que le respect est une condition suffisante pour le consensus ? Dans un article récent, Bradley (2006) met en doute la pertinence de ce résultat en soutenant que la représentation des informations d'ordre supérieur

par Lehrer et Wagner est inadéquate. L'argument de Bradley repose sur une comparaison avec le modèle bayesien dominant, et la critique principale consiste à dire que le modèle LW ne permet pas de prendre en compte l'impact des indépendances ou des corrélations entre les opinions des agents sur la manière dont ceux-ci doivent mettre à jour leur croyance individuelle sur la base des croyances des autres agents.

Notre but sera d'approfondir l'examen de la compatibilité entre l'approche bayesienne et le modèle LW :

*** nous soutiendrons que la critique de Bradley sur l'articulation entre poids et indépendance est trop rapide, parce que Bradley présuppose à tort que les poids épistémiques dans le modèle LW ne peuvent et ne doivent refléter que l'estimation de la compétence individuelle.

*** nous étudierons les conditions dans lesquelles les modèles LW peuvent être plongés dans les modèles bayesiens. Une condition importante concerne les relations entre les priors de l'agent et les valeurs qu'il considère comme possible pour les croyances des autres agents. Nous montrerons que l'hypothèse de respect contraint de façon très sévère ces relations, de sorte que l'applicabilité du modèle LW à des scénarios concrets peut être mise en doute.

Bibliographie

Bradley, R. (2006) "Taking advantage of difference in opinion", Episteme, Vol. 3 (3), 141-155.

Lehrer, K. & Wagner, C. (1981) Rational consensus in science and society, D. Reidel.

***

Exposé 4: Normativité en interaction : le cas des attitudes de niveaux supérieurs.

Dans plusieurs contextes sociaux, il semble que nous devons, au sens normatif du terme, tenir compte de certains faits à propose de ce que les autres croient, ou s'attendent de nous. Certaines actions peuvent être jugées irrationnelles si elles sont accomplies en l'absence de telles attitudes de niveaux supérieurs. Prenons par exemple la scène célèbre du film Dr. Strangelove de Stanley Kubrick, ou le premier déclare à l'ambassadeur russe : « The whole point of the Doom's Day Machine is lost... *if you keep it a secret; why didn't you tell the world, he*?!?! ». Strangelove remarque ici qu'il est difficile de rationaliser la construction de la machine en question si tous les partis en cause ne sont pas informés de l'existence de cette machine, ou même si ce fait n'est pas connaissance commune.

Dans cette présentation nous nous pencherons sur de telles situations, ou des jugements normatifs sont portés sur ce qui devrait être croyance, connaissance ou attente mutuelle ou commune, et sur la façon dont ces attitudes de niveaux supérieurs sont prises en compte lors de la délibération pratique. Notre cadre de référence sera la théorie épistémique des jeux (Brandenburger, 2007) et la logique épistémique dynamique (van Ditmarsch *et al*. 2007). Après avoir expliqué comment les « règles de choix » (par exemple dominance, maximisation de l'utilité attendue, admissibilité et maximin) peuvent être vues comme sources de jugements normatifs, nous ferons un survol de résultats connus concernant la réactivité de ces règles de choix à des perturbations dans les attitudes de niveaux supérieur (e.g. Rubinstein 1989, Apt, 2007, Trost, Manuscript), et montrerons leur pertinence du point de vue d'une théorie générale et normative de la rationalité en situation d'interaction sociale.

Références:

K.R. Apt. The many faces of rationalizability. *The B.E. Journal of Theoretical Economics*, 7(1), 2007. Article 18.

A. Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35:465– 492, 2007.

H. van Ditmarsch, W. van de Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of Synthese Library Series. Springer, 2007.

A. Rubinstein. The Electronic Mail Game: Strategic Behavior Under ”Almost Common Knowledge”. *The American Economic Review*, 79(3):385– 391, 1989. ISSN 0002-8282.

M. Trost, „On the Equivalence of Iterated Application of a Choice Rule and Common Belief of Applying that Rule”, Manuscript, 2010

Symposium proposal

(English version)

Higher-order attitudes: knowledge, beliefs and social interaction

Description of the workshop:

Higher-order attitudes occupy an increasingly important place in many areas of contemporary analytic philosophy. Higher-order attitudes are attitudes, cognitive or conatives, about attitudes. In epistemology, *introspection* principles have been the object of heated debates for many decades now. Does one always have, or should have, (correct) beliefs about one's beliefs? knowledge of one's ignorance? In recent years, questions of higher-order attitudes have proven to be of importance not only for these two classical pillars of epistemology, knowledge and beliefs, but also for notions such as *awareness* and *higher-ordervagueness*. To take another example, *disagreement* is arguably one of the most discussed notions in contemporary social epistemology. Here higher-order attitudes turn out to be crucial as well, but this time in the form of attitudes about the attitudes of others. Has information about others' information any epistemic significance? Should one always take others' beliefs into account while forming one's own beliefs? Finally, higher-order attitudes have also occupy an important place in meta-ethics, be it in the discussion of the importance of higher-order desires in views about *personal identity*, or about the source of *reasons* and *normativity*.

The aim of this workshop is to present, compare and relate a number of debates and questions involving higher-order attitudes. The workshop will consist of four talks, each of which is representative of a particularly active area of contemporary analytic philosophy. The first two talks will address questions raised in the modern debates about introspection in the epistemology of individual knowledge. The first talk will compare two forms of ignorance and two related forms of *negative* introspection, respectively involving knowledge and awareness, combining insights from cognitive sciences, epistemology and logic. The second talk will turn to positive introspection, but this time in relation with Williamson's margin for error principle for knowledge. The third and fourth talk will explore the social importance of higher-order attitudes, related to questions of disagreement and normativity, respectively. The third talk will explore the relation, both at the formal and at the philosophical level, between two well-known mathematical models of consensus formation. The fourth talk will turn to the notions of reasons and rationality in social interaction and will investigate, from a game-theoretical perspective, their dependence on higher-order attitudes. All in all, these four talks will provide a snapshot of contemporary areas of analytic philosophy where higher-order attitude play a key role and will, we hope, lay the ground for a fruitful interaction between these.

***

Talk 1: Unawareness, uncertainty and the knowledge of one's ignorance

What does it take to know that one does not know proposition? The answer to this question depends on the kind of ignorance under consideration. Ignorance comes in at least two distinct varieties, which are called *uncertainty* and *unawareness* in the recent literature in formal epistemology (see Franke and de Jager 2010 for a recent overview, and Bromberger 1987 for a precursor of the distinction).

Uncertainty corresponds to cases in which we hesitate between competing possibilities that we can grasp and articulate. Suppose I am asked whether Tolkien is the author of *The Hobbit*. One way in which I might fail to know the answer would be for me to hesitate between a yes and a no answer. I might entertain two competing alternatives, one on which Tolkien is the author of *the Hobbit*, the other on which it is C.I. Lewis. Unawareness, on the other hand, corresponds to cases in which we fail to know a proposition because we do not have the wherewithal to even represent or articulate the proposition in question. Compare the previous case with a situation in which I never heard of Tolkien, nor of *The Hobbit*. In that case my failure to know that Tolkien is the author of *the Hobbit* is not the result of a competition between various possibilities or pieces of evidence, but rather, it is the effect of a *lack of conception* of the basic components of the proposition (see Heifetz et al. 2006, who propose this characterization of unawareness).

An important aspect of the distinction between unawareness and uncertainty is that unawareness, unlike uncertainty, is *not introspective* (see Dekel et al. 1998, Franke and de Jager 2008). This means that whereas uncertainty is accessible to consciousness, so that one can know that one is uncertain that *p* at the moment one experiences that uncertainty, one cannot know that one is unaware that *p* simultaneously with that first-order state, for knowing that one is unaware that *p* would require one to represent the content that *p*. It is obviously possible to realize that one *was* unaware of something, but not to know that one *is* unaware proper.

The distinction between uncertainty and unawareness has been at the center of much formal work in epistemic logic recently. The perspective of this talk will be to discuss the metacognitive implications of the distinction in the light of work done on the psychology of known unknowns. In a classic study, Glucksberg and McCloskey (1982) have proposed a two-stage model of decisions about ignorance. On their view, subjects can issue a don't know verdict when faced with a question on two grounds: one concerns cases in which subjects find no potentially relevant evidence in their memory. For such cases, subjects are expected to give fast verdicts of ignorance. A second category of don't know answers corresponds to cases in which subjects do find some relevant evidence in memory, but find no conclusive evidence so as to satisfactorily answer the question. For such cases, Glucksberg and McCloskey results indicate that verdicts of ignorance take longer, consistently with the hypothesis of a heavier processing load (see also Hampton et al. 2011). A related aspect we shall focus on concerns the reliability of decisions about one's ignorance. Cases in which we find relevant information in our memory but remain uncertain should be cases for which we find harder to give a reliable estimate of the degree of our uncertainty, and more generally for which we could easily overestimate or underestimate our ignorance. By contrast, cases in which we find no relevant information in our memory, such as cases grounded in antecedent unawareness, should be cases for which we issue more reliable decisions about the true state of our ignorance.

References

Bromberger S. (1987) What we don't know when we don't know why. Repr. in Bromberger S. *What we know we don't know*. University of Chicago Press and CSLI Publications. 1992.

Dekel E., Lipman B.L., Rustichini A. (1998). Standard State-Space Models Preclude Unawareness. *Econometrica* 66 (1): 159-173.

Franke M. & de Jager T. (2010). Now that you mention it: Awareness Dynamics in Discourse and Decisions. In A. Benz et al. (eds), *Language, Games, and Evolution*. Lecture Notes in Computer Science.

Glucksberg, S. & McCloskey, M. (1981). Decisions about Ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory* 7: 311-325.6207, 60-91.

Hampton J. A., Aina B., Mathias Andersson J., Mirza H. Z., Parma S. (2011), The Rumsfeld Effect: the unknown unknown. *Journal of Experimental Psychology: Learning Memory & Cognition*, in press.

Heifetz A. & Meier M. & Schipper B. C. (2006). Interactive unawareness. *Journal of Economic Theory* 130: 7894.

***

Talk 2: On the foundation of the margin for error principle

According to Timothy Williamson's (1994) theory of inexact knowledge, perceptual knowledge satisfies a margin for error principle resulting from the limited accuracy of individual perceptions. This principle in turn gives rise to the epistemic sorites paradox, which Williamson solves by rejecting the view that knowledge satisfies positive introspection.

Several authors (Mott, 1998; Dutant, 2007; Dokic and Egré, 2009) have criticized Williamson's reasoning by resorting to an explicit modeling of perceptions. They showed that even if perceptions are imprecise, individuals can make inferences based on their perceptions and on their knowledge of their perceptual limitations, resulting in knowledge that may violate the margin for error principle. Williamson (2000) retorts that this argument is invalid because it assumes individuals to have perfect rather than inexact knowledge of their perceptual limitations.

Williamson's rebuttal of the rebuttal of the margin for error principle therefore appeals to the inexactness of higher-order knowledge, i.e., to the limits of individual perceptions about the limits of individual perceptions.

Our aim is to assess the merits of the two sides of this debate by explicitly modelling perceptual limitations at various orders. We show that, for a certain class of signal structures, the margin for error principle for perceptual knowledge holds only to the extent that it holds at all higher orders. However, in order to avoid an infinite regress, one must assume that there exists some primitive knowledge not resulting from perceptions. A full modelling of perceptions at all orders thus casts doubt on Williamson's claim that the margin for error principle can be justified by considerations on perceptual limitations. Furthermore, we find

that such a modelling may warrant the rejection of the margin for error principle for perceptual knowledge.

References

Dutant J. (2007). Inexact Knowledge, Margin-for-Error and Positive Introspection, in D. Samet (ed.), *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge* (TARK XI), Louvain-la-Neuve: Presses Universitaires de Louvain-la-Neuve, 118-124.

Dokic J. and Egré P. (2009). Margin for error and the transparency of knowledge, *Synthese*, 166(1), 1-20.

Mott P. (1998). Margins for Error and the Sorites Paradox.*The Philosophical Quarterly*, 48(193), 494-504.

Williamson T. (1994). *Vagueness*. London and New York: Routledge.

Williamson T. (2000). Margins for Error: a Reply. *The Philosophical Quarterly*, 50(198), 76-81.

***

Talk 3: Consensus and higher-order information

Prima facie,dissensus, as a property of first-order beliefs, and respect, as a property of second-order beliefs, are perfectly compatible. We might disagree, and I could still value your opinion, just like you could value mine. However, this is not clearly so, because respect might lead us to move a bit and get closer and closer to each other's opinions. At the end of the day, a consensus might be reached, suggesting that consensus is hidden in respect. The aim of this talk will be to get a better understanding of this kind of scenarios by studying the relationships between consensus and higher-order information.

Lehrer and Wagner's theory of rational consensus (1981) provides a formal model of this kind of potential consensus by characterizing the conditions under which an actual consensus may be reached at the limit of a sequence of updates based on respect. In their framework, first-order and higher-order beliefs are represented in a non-homogenous way: first-order beliefs are subjective probabilities whereas higher-order beliefs are represented as weights each agent grants to other agents' subjective probabilities. At each stage, agents update their beliefs to weighted averages of the beliefs in the community. In that framework, Lehrer and Wagner show that convergence toward consensus will occur if all weights are positive.

Should we take Lehrer and Wagner's result to show that respect is a sufficient condition for consensus? In a recent paper, Bradley (2006) casts some doubts on the relevance of the result by arguing that the representation of higher-order information in Lehrer and Wagner's model is inadequate. Bradley's argument rests on a comparison with a Bayesian analysis of what happens when an agent modifies his belief by showing some respect to other agents' beliefs. His main criticism is that the LW model is not able to take into account the impact of independence and correlations between agents' opinions on epistemic updates.

Our aim shall be to examine further the compatibility between the Bayesian model and LW models:

*** we shall argue that the criticism by Bradley about weights and independence is too quick, because Bradley presupposes that epistemic weights are to express nothing more than the value granted to the subjective estimate of the other agent's epistemic competence.

*** we shall study the conditions under which LW models can be embedded into Bayesian models. A significant condition concerns the relationship between the agent's prior and her beliefs about the possible values for the other agents' beliefs. The respect assumption can be shown to constrain this relationship in a very substantial manner, so that the applicability of the LW model to actual scenarios might be put at risk.

References

Bradley, R. (2006) "Taking advantage of difference in opinion", *Episteme*, Vol. 3 (3), 141-155.

Lehrer, K. & Wagner, C. (1981) *Rational consensus in science and society*, D. Reidel.

***

Talk 4: Normativity in Interaction: the Case of Higher-Order Attitudes.

In many social situations, in seems that we are under normative pressure to take into account facts about what others believe about, or expect of us. We can be rationally criticizedfor overlooking or ignoringsuch facts. Take for instance the famous scene of Stanley Kubrick's Dr. Strangelove, when the latter tells the Russian ambassador "The whole point of the Doom's Day Machine is lost... *if you keep it a secret; why didn't you tell the world, he*?!?!". Strangelove is pointing out that building the machine in question makes little sense without making sure that everyone knows about its existence or, even, without it being common knowledge.

This talk will be about such situations, where we raise *normative claims* about what should be *mutually* or *commonly*known, believed or expected, and about how these should bear on actions. Our starting point will be contemporary epistemic game theory (Brandenburger, 2007) and dynamic-epistemic logic (van Ditmarsch *et al*. 2007). After explaining how to see "choice rules" (e.g. dominance, maximization of expected utility, admissibility, maximin) as potential sources of normative statements, we will survey known results concerning the sensitivity of these choice rules to perturbations in higher-order attitudes (e.g. Rubinstein 1989, Apt, 2007, Trost, Manuscript), and explain the significance of theses results from the perspective of a general, normative theory of rational decision making in social interaction.

References:

K.R. Apt. The many faces of rationalizability. *The B.E. Journal of Theoretical Economics*, 7(1), 2007. Article 18.

A. Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35:465– 492, 2007.

H. van Ditmarsch, W. van de Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of Synthese Library Series. Springer, 2007.

A. Rubinstein. The Electronic Mail Game: Strategic Behavior Under „Almost Common Knowledge". *The American Economic Review*, 79(3):385– 391, 1989. ISSN 0002-8282.

M. Trost. On the Equivalence of Iterated Application of a Choice Rule and Common Belief of Applying that Rule, Manuscript, 2010