

## **Martin Aranguren (CNRS). Misrecognitive discrimination: inferring standards of adequate regard from individual emotions**

Is it possible to infer social norms from the emotions of individuals? More specifically, is it possible to infer standards of adequate regard or recognition from the emotional responses of participants to episodes of social interaction?

The aim of the talk is to prepare the ground for a positive answer to these questions. To this purpose, I will first outline an original research programme concerned with the exploration of a phenomenon that I propose to call misrecognitive discrimination. In the second part of the talk, I will build on Arto Laitinen's proposal to assume "minimal objectivism" about recognition, and to base that objectivism on the recourse to social norms. But, drawing on a particular reading of Durkheim and Garfinkel, I will also suggest that knowledge of those norms need not derive from public and explicit sources such as law, but may also come from indirect sources such as the emotions of a population.

### A research program on misrecognitive discrimination

Misrecognitive discrimination can be defined as differential treatment on the grounds of membership in a socially salient group (e.g. being a Muslim) that functions to degrade (disregard, depreciate) the target of the treatment. This programme combines the theory of recognition, the social psychology of prejudice, and the sociology of discrimination.

Recognition. You recognize me when you take me as a person (*lkäheimo*). If the recognition I get from you is inadequate, then you have misrecognized me. Love, respect, and esteem, are specific forms of recognition, and so neglect, disrespect, and disesteem are specific forms of misrecognition (*Honneth*). Arguably, receiving love, respect, and esteem from others are directly and indirectly relevant for basic human needs, and hence severe harm can be caused by misrecognizing people (*Laitinen*). Now, discrimination research has proceeded as though the only important consequences of differential treatment were to create disparities in access to material resources such as jobs, healthcare, or housing. But discrimination is also the source of inequities in the allocation of recognition, especially in the forms of respect and esteem, which as said are relevant to basic human needs.

Discrimination. One way of defining actual discrimination is to say that X discriminates against Y if and only if i) X treats Y differently from Z; ii) the differential treatment is disadvantageous to Y; and iii) the differential treatment is suitably explained by Y's and Z's being members of different, socially salient groups (*Lippert-Rasmussen*). Thus actual discrimination is an essentially comparative notion. The epistemological implication is that there is no way to demonstrate actual discrimination against Y unless there is some Z to provide a "control". This is, in effect, the rationale on which "discrimination testing" research relies. For example, a number of companies offering job vacancies receive two CVs identical in content but presenting one applicant as the member of a majority and the other as the member of a minority. The investigators then record some measure of success (e.g. the number of job interview proposals received), and finally compare the success rate of the applicants. If there is a statistically significant difference in favor of the majority member (i.e. the control), the conclusion is that the minority member was objectively discriminated against. The present programme uses the rationale of "discrimination testing" research.

Misrecognitive discrimination against Muslims and gays/lesbians. The program focuses on two European minorities that can be suspected to be the target of misrecognitive discrimination on a day-to-day basis: Muslims and gays/lesbians. Extant research indicates that members of negatively viewed minorities experience misrecognition most frequently in public places. Further, in a historical context in which blatant intergroup antipathy is mostly regarded as socially undesirable, the project will focus on subtle forms of bias against Muslims and gays/lesbians, and more specifically on the nonverbal dimension. Thus the specific objectives of the programme are: 1) to identify forms of differential treatment (especially nonverbal) toward Muslims in the context of face-to-face encounters with strangers in public places; and 2) to assess whether the forms of discrimination unveiled as a result of objective 1 are also misrecognitive.

The programme will pursue objective 1 by means of a field experiment known as the helping behavior paradigm, and objective 2 through an original invention that could be named the "survey paradigm." In both field experiments randomly chosen passengers interact with a confederate actor on a metro platform. Discrimination is "produced" by having passengers interact with an actor who is recognizable as a Muslim or gay/lesbian in the "treatment" condition but not in the "control" condition. Pilot studies already conducted in the Paris metro advocate for the feasibility and adequacy of this two-fold methodology.

### Inferring standards of recognition from emotions

It can be expected that at least some standards of adequate recognition exist only à l'état pratique, in the same way that rules of grammar are correctly followed by speakers of a language without them necessarily holding any second-order thought about those rules. To that extent, socially shared standards of recognition are not directly available to the researcher, which means that they have to be inferred from some indirect source. But what source?

In this respect, I propose to follow a principle classically formulated by Durkheim and famously exploited by Garfinkel in his "breaching

experiments.” This is the principle that negative emotions such as anger, shame, or disgust, when they respond to some form of social behavior, and when the emotional response is sufficiently similar within a population, reveal that the behavior is experienced by this population to deviate from social norms. More specifically, recognition theory and research on the self-conscious emotions predict that targets of misrecognition should experience shame or “humiliated fury.”

With emotions, as with anything else in biology or psychology, we should expect some amount of interindividual variation within the relevant population. Thus discovering a deviation from a practical norm of recognition requires to expose samples of the same population to different treatments and to compare the emotional responses elicited by those diverging treatments. I will indicate how this can be achieved with a field experiment.

## Eivind Balsvik (University of Oslo, Faculty of Social Sciences). When interpretation becomes challenging

In favorable circumstances, interpreting social behavior is relatively straightforward. If an individual is sincere, and if his verbal behavior matches his non-verbal behavior, an interpreter can use the verbally expressed beliefs and desires to give meaning to the non-verbal behavior, and the non-verbal behavior to confirm the interpretations based upon the verbal-behavior (Cf. Hopkins' (1999) method of testing for normative accord). However, in cases where there is a mismatch between an individual's verbal and non-verbal behavior, interpretation becomes quite challenging. This presentation investigates difficulties to interpretation posed by three kinds of cases where verbal behavior and non-verbal behavior do not mesh: self-deception, akrasia and implicit racism or sexism. These phenomena occur frequently, and will often be of interest to social scientists. I shall try not to get bogged down by the literature on competing conceptualizations of these phenomena, and will initially base my discussion on vignettes similar to the one's developed and discussed by Gendler (2008), Mele (2009) and Schwitzgebel (2010). When suggesting refined interpretations, I shall however, appeal to theories from the behavioral sciences (social psychology, behavioral economics).

My presentation has four goals:

- I wish to assess an interpretative methodology inspired by Davidson's (1980, 1984, 2001), Henderson's (1993) and Hopkins' (1999) works on interpretation, developed in Author (2017, forthcoming)
- I hope to illuminate the role of the principle of charity and the presumption of first-person authority in interpretation
- I will discuss the corrigibility of self-ascriptions of psychological predicates and of the rationality assumptions interpretation depends upon
- I wish to illuminate the role of psychological theory in interpretation, and will consider whether dual-systems theory can serve as an overarching psychological theory in refined interpretation

My interpretative methodology utilizes Henderson's (1993) distinction between developing a first-approximation scheme and refined interpretation. Whereas the primary goal of first-approximation schemes is to attribute meaning to verbal and non-verbal behaviors, the goal of refined interpretation is to offer explanations by uncovering the mental processes that are the causes of actions. When developing a first-approximation scheme, Davidson's principle of charity and the presumption of first-person authority serve as rather strong constraints on adequate interpretation. In refined interpretation, the interpreter might use psychological theory, and explain some beliefs or actions as seriously mistaken or irrational.

Let me briefly state my working hypotheses regarding the kinds of cases I wish to investigate:

There seem to be many varieties of self-deception. I am particularly interested in cases where an individual's verbally expressed beliefs and desires do not reveal the attitudes that determine (some) of the subject's behavior related to those beliefs. I contend that a presumption of first-person authority is necessary in order to provide evidence for self-deception. I wish to consider whether an adequate explanation of self-deception must explain how the self can be both deceiver and deceived, or whether there are plausible explanations which do not make this assumption. That is, does dual-systems theory provide the best framework for explaining self-deception? Although a presumption of first-person authority is necessary in order to provide evidence for self-deception, self-deception is a kind of case where an individual's claim to self-knowledge is either seriously incomplete or otherwise deficient.

Akrasia: the presumption of first-person authority is also necessary for providing evidence of akrasia. Contrary to Lawlor (2003), I do not think akrasia undermines an individual's claim to authoritative self-knowledge (Cf. Ferrero (2003)). However, in so far as akratic actions are caused by a reversal of preferences, simply due to the passing of time, it is incompatible with an assumption of unbounded rationality. Metcalfe & Mischel's (1999) distinction between a hot and a cool system in the brain or Thaler's (2015) model of the planner and the doer, coupled with the theory of hyperbolic discounting explains akrasia.

Implicit racism: The presumption of first-person authority is also necessary for providing evidence for implicit racism. An implicit racist will disavow racism in his verbal behavior, yet discriminate in his other behavior. According to the prevailing psychological theory, implicit racism or sexism is due to unconscious biases, expectations or tendencies. Such unconscious biases can be measured by a research paradigm known as the Implicit Association Test. However, recent meta-analysis questions the link between unconscious bias, as measured by the Implicit Association Test, and biased behavior. The correlation between implicit bias and discriminatory behavior appears weaker than previously thought. Moreover, changes in implicit bias does not lead to changes in discriminatory behavior (Forscher et al (2016); Oswald, et al (2013); but see also Poehlman and Banaji 2009). Perhaps implicit racism is better explained as a form of self-deception, or that the discrimination is explicit, and the verbal behavior is simply an attempt to appear to be politically correct?

## **Olle Blomberg (Lund University). Team reasoning, joint action, and acting as if part of one large agent**

There is widespread agreement that joint intentional action is of a cooperative character (even though it is compatible with some forms of coercion and manipulation): Participants treat each other as partners rather than as social tools. An influential way of capturing this cooperative character of joint intentional action is to require that participants have various interconnected intentions or plans (see e.g. Bratman 2014; Ludwig 2016). This theoretical strategy, while elegant, has a cost: It makes joint intentional action complex (overly so, one might think) as well as conceptually and cognitively demanding. The aim of this talk is to consider an alternative strategy and argue that it does not work. It seems that joint intentional action is a complex as well as conceptually and cognitively demanding phenomenon.

The alternative strategy is embedded in several accounts. According to team reasoning-based accounts of “shared intention”—where this is what makes a joint action jointly intentional—individuals engage in joint intentional action partly in virtue of the following: they each conceive of themselves and the other participants as parts of a single agent (see Bacharach 2006; Gold and Sugden 2007; Pacherie 2011, 2013). Similarly, on Margaret Gilbert’s (e.g. 2009) account, parties to a shared intention are jointly committed to intending as a single agent or body to perform an action. The idea is thus to make sense of the complex phenomenon of joint intentional action in terms of the idea of something that is conceptually simpler, namely a kind of singular group-level agency. In the case of team reasoning-based account, the participants are induced to conceive of themselves as parts of such singular group-level agency through a process of group-identification, and in the case of Gilbert’s account, they end up conceiving of themselves and the others in this way as a result of making an irreducibly joint commitment to intend to do something as a single agent or body.

While this may seem like promising strategy, I argue that it fails to capture the cooperative character of joint intentional action. The accounts fail to provide sufficient conditions for shared intention since participants who act to bring about the goal of an additional larger agent, of which they take themselves to be parts, are not barred from intentionally coercing or manipulating the other participants in a way that is incompatible with joint intentional action.

Team reasoners can coordinate their actions to reap the greater mutual benefit in some simple coordination problems that have no unique solution for ordinary individualistic game-theoretical agents. For example, they can rationally coordinate on the Hi equilibrium in the Hi-Lo game. Team reasoning may therefore look like it is essentially cooperative. However, in this talk, I argue that nothing rules out that the goal-directed activity of the single group-level agent that each identifies with involves conflict, disagreement and coercion between parts that implement it. What matters is simply that each part is playing the right causal role. Think of the activity among the employees of a corporation or the activity of insects that together make up a “swarm agent”. What matters for the agency of the whole is the global properties of the behaviour of the collective; there is no need for this behaviour to be implemented by cooperative interactions between the parts. Hence, nothing in accounts of team reasoning rules out that participants intend to do their parts in bringing about the group goal by means of brute coercion or manipulation. This is compatible with the idea in the minds of the participants that they are acting as parts of a single group-level agent.

Team reasoning-based accounts would only rule out such forms of coercion and manipulation if they required that the parties engaged in team reasoning “all the way down” to intentions to perform basic actions (a basic action is an action that an agent intends to perform directly without intending to perform it by means of intending to perform some other action, everyday examples might be an agent grasping a cup or clenching her fist). Similarly, Gilbert’s joint commitment account would only rule out similar forms of coercion if it required that the parties to a shared intention had a joint commitment that specified into every last detail what the parties, as a single body, were to do.

I argue that an assumption that team reasoning or joint commitments reach all the way down to the output of practical reasoning—intentions to perform basic actions—is implausible. Because of this, appealing to participants’ conception of themselves and their co-participants as parts of a larger agent or body fails to capture the cooperative character of joint intentional action. There is nothing inherently cooperative about team reasoning or joint commitment. This does not mean, of course, that processes of team reasoning or joint commitment cannot play important other roles in cooperation or joint intentional action. But taken as accounts of shared intention, the accounts considered in this talk will be shown to be incomplete. Adding conditions that require interconnected intentions and plans can complete them, but this brings back the conceptual and cognitive complexity that one might have hoped to avoid in the first place (see especially Pacherie 2011, 2013).

In a final part of the talk, I also consider to what extent my argument against team reasoning accounts of shared intention can be transposed to attempts to use team reasoning to solve dynamic choice problems facing an agent who exercises his or her intentional agency across time (see e.g. Gold 2013).

### References:

- Bacharach, M. (2006), *Beyond Individual Choice: Teams and Frames in Game Theory*, Princeton University Press.
- Bratman, M. (2014), *Shared Agency: A Planning Theory of Acting Together*, Oxford University Press.
- Gilbert, M. (2009), Shared intention and personal intentions, *Philosophical Studies*, 144(1), 167–187.

- Gold, N. (2013), Team Reasoning, Framing, and Self-Control: An Aristotelian Account, in N. Levy, ed., 'Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience', Oxford University Press, chapter 4, pp. 48–66.
- Gold, N. and Sugden, R. (2007), 'Collective intentions and team agency', *Journal of Philosophy* CIV(3), 109–137.
- Ludwig, K. (2016), *From Individual to Plural Agency: Collective Action: Volume 1*, Oxford University Press.
- Pacherie, E. (2011), 'Framing Joint Action', *Review of Philosophy and Psychology* 2(2), 173–192.
- Pacherie, E. (2013), 'Intentional joint agency: shared intention lite', *Synthese* 190(10), 1817–1839.

## **Hubert Cambier (Independent Scholar). The ethical, political and metaphysical foundations of Karl Popper's methodological individualism**

Karl Popper is well-known for his defence of methodological individualism in social sciences. His position is not exempt of difficulties that have been highlighted by the controversies even among his close commentators.

In the paper, I intend to clarify Popper's position(s) with regards to the methodology of social sciences and individualism in particular. I will discuss in particular the choice of a nominalist approach that, for Popper, justifies the adoption of methodological individualism, the more so that he was never a nominalist philosopher.

I will argue that, under the name of nominalism, what Popper defended in fact is an ethical and political will to put the individuals at the centre of the social life, of history and of evolution. Methodological individualism should then be reformulated in a metaphysical individualism, emphasising the autonomy of the subject against all kind of heteronomy. This, in my opinion, raises the role of philosophy with regards science – should it provide foundations for sciences enable to establish them themselves or should the philosophy accept that sciences can take care of themselves.

## Ricardo Crespo (IAE /Universidad Austral/ and CONICET). Liberal Naturalism and non-epistemic values

The 'value free ideal', a requirement for science correctness enjoying a long tradition since the Eighteenth century, has been called into question for different reasons. This requirement does not include "epistemic values" such as explanatory and unifying power, simplicity, consistency and predictive accuracy, which are generally considered as characteristic of 'good science', and it rejects the so-called 'contextual', 'non-cognitive' or 'non-epistemic' values – all of them being personal, moral, or political values. Many arguments against the possibility of excluding non-epistemic values rely on Willard van Orman Quine (1951) challenge to the analytic-synthetic distinction. They were developed and sustained, for example, by Hilary Putnam (2004) Richard Rudner (1953), Helen Longino (1990) and John Dupré (2007) (as reviewed by Kincaid, Dupré and Wylie, 2007, and Reiss and Sprenger, 2014).

In this paper I will analyze a possible complementary argument against the value free ideal. Today's predominant naturalistic view has tended to 'naturalize' values by looking for physicalist explanations of them. This tendency has been resisted by defenders of normativism in the social sciences (see, for example, Risjord, 2016). At the same time, a contending naturalist stream has emerged, claiming that not all natural entities or processes, especially human and biological ones (Dupré, 2001; Nagel, 2012), can be explained by the methods and concepts of physical sciences.

According to John McDowell (2002 and 2004: 92), modern natural science has evolved as a mechanistic approach to natural processes – 'a disenchanting conception of the natural world' (2002: 174) – in which the knowing subject (the human being) threatens to withdraw from the natural world. It is tempting to identify nature with the subject matter of modern natural sciences (2004: 92); nevertheless, McDowell argues that this might likely be a mistake. He makes a distinction between a 'restrictive naturalism', aiming 'to naturalize the concepts of thinking and knowing by forcing the conceptual structure in which they belong into the framework of the realm of law [as opposed to the realm of reason, expressions previously used by Sellars (1956)]' (2004: 95), and a 'liberal naturalism' which does not require to integrate our capacities of thinking into this narrow scientific framework – 'our capacities to acquire knowledge are natural powers' (2004: 95). He also named it as 'relaxed naturalism' (1996: 89). For him, 'knowledge and intentions can only be considered in the framework of the space of reasons' (2004: 93). Hence, 'we can bring practical reason back into nature' (2002: 184). That is, nature would provide with more than what natural sciences consider: it leaves room for practical reasoning, the human ability to rationally choose ends. Restrictive naturalism 'interprets the natural strictly in terms of the scientific image of the world, narrowly or broadly conceived, whereas liberal naturalism – or some versions of it – offers a broader, more expansive conception of nature that makes room for a class of nonscientific, but nonetheless non supernatural, entities' (De Caro and Macarthur 2010: 3-4; see also De Caro and Macarthur, eds. 2004).

Some authors develop naturalistic positions in the same 'liberal' spirit: Jennifer Hornsby (who calls it 'naïve naturalism', 1997), Barry Stroud ('a more open-minded or expansive naturalism' 1996: 54), Peter Strawson ('liberal', 'catholic' or 'soft' naturalism, 1985: 1 and 42), Johannes Brandl ('a modest form of naturalism', 2007: 256) and James Griffin ('an expansive naturalism', 1988: 51). Thomas Nagel (2012: 8) favors 'a pervasive conception of natural order, very different from materialism' (2012: 15) – that is, a non-materialist naturalism, including mind, consciousness, meaning and value as fundamental parts of nature that cannot be reduced to matter (id.: 20; 44). Nagel believes that teleology is 'a naturalistic alternative' (id.: 91; 122; 124) and asserts that human action 'is explained not only by physiology or by desires, but by judgments' (id.: 114) made by practical reason.

In this way, within 'liberal naturalism' it can be postulated that non-epistemic values could be 'naturally' included in the field of human sciences. In the paper I will analyze the liberal naturalistic arguments as well as its legitimacy to include non-epistemic values in the social sciences.

### References:

- Brandl, J. (2007). 'The Unmysteriousness of Consciousness: A Case Study in Naturalistic Philosophy', in G. Gasser (ed.) *How Successful is Naturalism?*, Publications of the Austrian Ludwig Wittgenstein Society. NS/4. Frankfurt: Ontos Verlag.
- De Caro, M. and Macarthur, D. (2010). 'Introduction: The Nature of Naturalism'. In M. De Caro and D. Macarthur (eds.), *Naturalism in Question*, Cambridge (Mass.): Harvard University Press.
- De Caro, M. and Macarthur, D. (2010). 'Introduction. Science, Naturalism, and the Problem of Normativity'. In M. De Caro and D. Macarthur (eds.), *Naturalism and Normativity*. New York: Columbia University Press.
- Dupré, J. (2001). *Human Nature and the Limits of Science*. Oxford: Oxford University Press.
- Dupré, J. (2007). 'Fact and Value', in Kincaid, Dupré and Wylie (eds.), *Value-Free Science? Ideals and Illusions*, Oxford: Oxford University Press.
- Griffin, J. (1988). *Value Judgement*. Oxford: Oxford University Press.
- Hornsby, J. (1997). *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind*. Cambridge (Mass.): Harvard University Press.
- Kincaid, H, J. Dupré and A. Wylie (2007). 'Introduction', in Kincaid, Dupré and Wylie (eds.), *Value-Free Science? Ideals and Illusions*. Oxford: Oxford University Press.
- Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

- McDowell, J. (1996). *Mind and World*. Cambridge (Mass.): Harvard University Press.
- McDowell, J. (2002). 'Two Sorts of Naturalism'. In J. McDowell, *Reason, Value, and Reality*. Cambridge (Mass.): Harvard University Press.
- McDowell, J. (2004). 'Naturalism in the Philosophy of Mind'. In M. De Caro and D. MacArthur (eds.), *Naturalism in Question*, Cambridge (Mass.): Harvard University Press.
- Nagel, T. (2012). *Mind and Cosmos*. Oxford and New York: Oxford University Press.
- Putnam, H. (2004). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge (Mass.) and London: The MIT Press.
- Quine, W.v.O. (1951). 'Two Dogmas of Empiricism', *The Philosophical Review* 60/1: 20-43.
- Reiss, J and J. Sprenger, 2014. 'Scientific Objectivity', *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/scientific-objectivity/>. Retrieved 30 January 2017.
- Risjord, M. (2016). *Normativity and Naturalism in the Philosophy of the Social Sciences*, London: Routledge.
- Rudner, R. (1953). 'The Scientist Qua Scientist Makes Value Judgments', *Philosophy of Science* 20: 1-6.
- Sellars, W. (1956). 'Empiricism and the Philosophy of Mind'. In H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. University of Minnesota Press.
- Strawson, P.F. (1985). *Skepticism and Naturalis: Some Varieties*, London: Methuen.
- Stroud, B. (1996). 'The Charm of Naturalism', *Proceedings of the American Psychology Association* 70/2: 43-55.

## Willem van der Deijl (Erasmus University Rotterdam). Can welfare be measured using preference-satisfaction?

Welfare is a central normative concept in economics, and is most commonly understood in economics as preference-satisfaction. For most of the twentieth century it has been the standard view in economics that welfare as preference-satisfaction is not measurable in the form of an index, even though some inferences could be made about it in market contexts - e.g. *ceteris paribus*, welfare increases when income increases. In the past decade economists have started to rely more and more on indicators of subjective wellbeing (SWB), such as happiness and life-satisfaction measures, to estimate welfare and the impact of economic phenomena (such as growth and unemployment) on welfare. Such indicators of SWB have challenged the assumption that welfare indices are not feasible, however, they do so by deviating from the view that welfare should be measured in terms of preference-satisfaction.

In recent years, (partly) in response to these developments, a number of economists have developed preference-based index measures of wellbeing, novel in their kind. One influential development in this regard is Daniel Benjamin et al.'s "Beyond happiness and satisfaction: toward wellbeing indices based on stated preference." (*The American economic review* 104:9 (2014): 2698-2735) in which a general framework is proposed for the development of preference-based welfare indices in the policy context. At the same time, preference-based measures of health related quality of life have been used in health economics for decades, but these measures have recently been generalized to get at measures of quality of life in general. One notable example of such a general measure is the ICEpop CAPability measure for Adults (ICECAP-A; Al-Janabi, Hareth, Terry N. Flynn, and Joanna Coast. "Development of a Self-Report Measure of Capability Wellbeing for Adults: The ICECAP-A." *Quality of life research* 21:1 (2012): 167-176). While using a capability framework, it uses preferences to weigh the relative importance of different capabilities for welfare.

To what extent do these measures succeed in developing preference-based welfare measures that are both faithful to normative and theoretical commitments of preference-satisfactionism as well as relevant for policy? In order to answer this question I specify that the main preference-satisfactionist commitment is that it is a non-paternalist view on welfare. Rather than a substantive view on welfare, it is a formal one, which avoids that any substantive claims about welfare do not resonate with the subjects about which they are made. I call this the Resonance Attraction.

Furthermore, I present a framework in which I suggest that a preference-based measure of welfare captures to what extent the world coheres to how we would like it to be, described in a finite number of dimensions. I distinguish the issue of ordinality of preferences from comparability and describe how this relates to policy questions. The main argument that I present is that while there is no conceptual problem in formulating a preference measure of welfare, due to feasibility constraints that come with a policy-relevant instrument, a preference-based measure of welfare cannot simultaneously be policy-relevant and do justice to its normative commitment: the Resonance Attraction.

In order to demonstrate the argument I present a detailed study of, first and foremost, Benjamin et al.'s framework, and its proposed suggestions to make pragmatic tradeoffs between theoretical rigor and feasibility. I then compare this framework with the ICACAP-A measure, which opts for different solutions to similar methodological choices.

Benjamin et al.'s theoretical framework formulates an individual index of welfare, but one that only tracks within-individual ordinal changes that are not interpersonally comparable. I argue this implies that it is limitedly useful in terms of policy. In order to measure it, they propose to use stated-reference methods that make respondents report hypothetical choices. I discuss Benjamin et al.'s commitment to not restricting the space of possible aspects of life. They select as many as 129 aspects of life to make up wellbeing from a variety of different contexts, such as objective lists from philosophy and other empirical work (e.g. based on SWB). Due to this commitment, it becomes unfeasible to construct individual preference maps indicating how all possible values of such aspects of life are valued compared to each other. Consequently, they use population preferences rather than individual preferences to keep their method feasible. While the commitments Benjamin et al. make are normatively highly attractive in the preference-satisfaction framework, they strongly limit the usefulness and validity of the measure as a policy instrument.

Benjamin et al.'s method contrasts strongly with the ICECAP-A, originating from health economics. Its developers base themselves on structured interviews and focus groups to identify only 5 functionings/capabilities that make up wellbeing, compared to the 129 dimensions Benjamin et al. employ. Furthermore, they calibrate their scale from 0 (having no capabilities at all) to 1 (having all capabilities fully), allowing for interpersonal comparisons, making it useful for policy. However, even with 5 aspects, there are still a large number of different wellbeing states that can be compared. They rely on group preferences as indicators of how valuable each of the capabilities is for wellbeing. Due to the limited number of capabilities considered, and due to the usage of group preferences to weigh the different capabilities for individual welfare, it becomes possible that this measure of welfare does not resonate with a person's view on what matters for his or her wellbeing.

In a final section I analyze the tradeoffs that a measure of welfare needs to make. I argue that a measure cannot be highly policy-relevant, be faithful to the theoretical commitments of a preference-satisfaction view, and at the same time be empirically feasible. The reason for this is that comparability and cardinality on the one hand, and individualism of a measure on the other, both are highly data-

demanding. A feasible measure of welfare thus requires making a tradeoff between policy-relevance and being faithful to the preference-satisfaction view.

## **Guillaume Dezecache (Institut Jean Nicod - Ecole Normale Supérieure of Paris). The social semantics of individual reactions to danger in humans: suggestions for a new typology**

How do humans individually and collectively react to imminent danger? Since the seminal work of Gustave Le Bon and others, individual reactions to threat have been characterized as self-preservative and antisocial in essence (Schweingruber and Wohlstein, 2005). Indeed, academic and non-academic audiences, crowd management experts and non-experts assume that humans succumb to individualistic tendencies when they perceive a danger (Drury et al., 2013). In other words, they “panic”. If the concept of panic is in itself difficult to define (Quarantelli, 1979), it is often proposed to correspond to the combination of three main components. People in panic are in a state of extreme anxiety (the affective component), they believe that something dangerous is happening but remain determined that there is a chance to escape (the epistemic component); finally, they are moved by a will to reach a safer environment, albeit at the expense of their neighbours (the motivational component). Additionally, panic is thought to be immediate and uncontrollable (Dezecache, 2015).

Surprisingly, it has never been empirically shown that humans primarily display self-preservative behaviour when exposed to threat in the company of others. On the contrary, numerous studies based on verbal reports with survivors from a variety of disasters have repeatedly shown that calm and helpfulness are the most prevalent responses to danger (see Dezecache, 2015 for a review of the evidence). Paradoxically, the concept of “panic” is often used by the survivors themselves (Cocking and Drury, 2014) but is largely inapplicable to what is then described. In survivors’ testimonies, “panic” may denote a lack of information on the exact nature of the danger or used to denunciate the inefficacy of the police or other emergency responses (Fahy et al., 2009; Cocking and Drury, 2014).

One of the most interesting evidence of calm and helpfulness during emergency situations comes from a case-study with survivors of the World Trade Centre attacks in New York (11/09/2001) (Proulx and Fahy, 2004). Perception of risk was high among survivors, but many cases of prosociality were reported. Prosocial responses typically include protecting and reassuring others helping injured individuals and clearing obstructed access. The emergence and stability of prosocial responses in endangered people is typically accounted by 3 (potentially exclusive) classes of explanation: first, social norms (such as the respect for physically weaker people) do not disappear in emergency situation and continue to prevail (Johnson, 1987). Second, affiliation and physical contact with others (preferentially familiar individuals) is a primitive response to danger, in humans and non-human animals (Mawson, 2005). The last explanation was brought to account for the fact that prosociality is massive and not restricted to familiar individuals: according to the Social Identification model of crowd behaviour, the perception of a common fate causes the adoption of a common “social identity” which in turn fosters prosociality among group members (Drury et al., 2009).

However, one fundamental limitation of those previous accounts is the theoretical polarization between presumably antagonistic motivations in threatened individuals: people could either be self-preservative or antisocial. In fact, no proper distinction is offered between examples of cooperation where both the agent and the recipient benefit from the interaction, e.g., clearing an obstructed access) and acts of altruism (where the agents suffer a direct net cost at providing support). This is yet a central conceptual distinction in other fields of research dealing with human social behaviour (West et al., 2007). As a matter of fact, prosociality may actively serve the agent. Indeed, distressed individuals can betray their position and that of their neighbours in the contexts of hostage-taking. In a study consisting of interviews with 30 survivors from the Bataclan attacks (Dezecache et al. in preparation), it is not uncommon to hear reports of prosocial acts that could be deemed “instrumental”, for instance, comforting somebody to make her/him silent, or providing others about some information about the position of hostage-takers to cause movements in people obstructing the exits.

In this paper, I will examine testimonies from survivors from the attacks at Le Bataclan to offer a new typology of individual reactions to danger and their impact on collective behaviour. Drawing on previous work, I will suggest that “instrumentality” (which qualify actions which are explicitly denoted as having been performed to benefit the agent) is a key concept to integrate to new research on human social behaviour during collective exposure to danger. I will also argue that accounting for the balance between the adoption of social and non-social strategies in endangered people constitutes a fundamental theoretical challenge for the future.

## Hein Duijf (Utrecht University). Collective reasoning and collective responsibility gaps

How should we reason in interactive decision contexts? Can we specify the difference between non-cooperative and cooperative decision problems? I aim to shed some light these questions by contrasting several practical reasoning paradigms. Bacharach (2006, Ch. 1) and Sugden (2000, Sections 2, 3, 7 and 8) argue that traditional game and decision theory needs to be augmented with a collectivistic reasoning method to successfully address cooperation problems like the Hi-Lo game. They introduce the team reasoning account of cooperation, which appeals to the reasoning method by which an individual agent reasons about what to do. An individual agent engaged in team reasoning “works out the best feasible combination of actions for all the members of her team, then does her part in it” (Bacharach, 2006, p. 121). The team reasoning literature has adopted the reasoning schema in Figure 1 (where (1)–(6) denote the premises, and •states the conclusion).

This team reasoning schema has a problem: it fails to deliver the desired recommendations in the alternative Hi-Lo game depicted in Figure 2. It seems

1. We can choose (A,A), (A,B), (B,A) or (B,B)
2. If we choose (A,A) the outcome will be O1
3. If we choose (A,B) the outcome will be O2
4. If we choose (B,A) the outcome will be O3
5. If we choose (B,B) the outcome will be O4
6. (A,A) uniquely maximizes each of our preferences
•Each of us should choose her component of (A,A)

Figure 1: Team reasoning schema.

		P2	
		high	low
P1	X	(2,2)	(0,0)
	Y	(2,2)	(0,0)
	low	(0,0)	(1,1)

Figure 2: The alternative Hi-Lo game.

that a good theory of cooperation should recommend choosing high to P2 and should recommend P1 to choose either X or Y, rather than low. Team reasoning, unfortunately, fails to deliver any of these recommendations. The problem for the team reasoning schema is that premise (6) is undermined: there are multiple group actions that maximize each of our preferences.

Our diagnosis of the failure of team reasoning is twofold: First, premise (6) in the team reasoning schema (Figure 1) highlights that it relies on an unrealistic uniqueness assumption, namely that there is a unique group action that maximizes each of the members’ preferences. To address the alternative Hi-Lo game, we need to drop this problematic assumption. Second, the team reasoning schema blurs the relation between team reasoning on the one hand, and individual and collective reasoning on the other. I aim to contribute to theories of cooperation by augmenting the team reasoning paradigm to, what I call, participatory reasoning in order to address these problems.

The participatory reasoning schema depicted in Figure 3 marks my main contribution (where (I•) denotes the conclusion; (C•) states intermediate conclusions; and the others denote the premises).

It is important to note that the participatory reasoning schema delivers the desired recommendations in the alternative Hi-Lo game (Figure 2). If P2 adopted the participatory reasoning schema, her reasoning would go as follows: first she reasons at the collective level and concludes that we should choose (X,high) or (Y,high). Then she reasons at the individual level and concludes that she should choose high, because choosing high is compatible with a best group action whereas choosing low is not. Similarly, participatory reasoning recommends P1 to choose either X or Y. My participatory reasoning schema thus resolves this paradox of team reasoning.

Can this mode of practical reasoning shed light on responsibility gaps in collective action contexts? In case of a collective wrongdoing this triggers the study of which (if any) members can be held responsible for it.<sup>1</sup> I therefore focus on cases of collective failure, which, in our current framework, are cases

(Stage 1: the collective level)
(C1) We can choose (A,A), (A,B), (B,A) or (B,B)
(C2) If we choose (A,A) the outcome will be O1
(C3) If we choose (A,B) the outcome will be O2
(C4) If we choose (B,A) the outcome will be O3
(C5) If we choose (B,B) the outcome will be O4

(C6) We prefer O1 over O2, O3 and O4
(C•) We should choose (A,A)
(Stage 2: the individual level)
(I1) I can choose A or B
(I2) If I choose A the outcome will be that we choose (A,A) or that we choose (A,B)
(I3) If I choose B the outcome will be that we choose (B,A) or that we choose (B,B)
(C•) We should choose (A,A)
(I•) I should choose A

Figure 3: Participatory reasoning schema.

where an inferior group action is performed. If we think of responsibility as performing those actions that are admitted by the participatory reasoning schema, then we thus investigate the ways in which individuals that endorse the participatory reasoning schema to maximize their collective preference can still end up performing an inferior group action.

In such a case, my set-up provides a distinction between two possible diagnoses: (1) The collective wrongdoing could originate from diverging expectations regarding out-group members. In this case, the group members with inaccurate expectations seem to be at fault. There is no responsibility gap.<sup>2</sup> (2) The inferior group action may result from diverging expectations regarding in-group members. In this case, it could be impossible to say which, if any, group member is at fault. There is a responsibility gap: there is a collective wrongdoing although none of the members can be held responsible.

#### References:

- Michael Bacharach. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton University Press, Princeton.
- Matthew Braham and Martin van Hees. 2011. Responsibility voids. *The Philosophical Quarterly* 61, 242 (2011), 6–15.
- Tracy Isaacs. 2011. *Moral Responsibility in Collective Contexts*. Oxford University Press.
- Christopher Kutz. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge University Press, Cambridge.
- Robert Sugden. 2000. Team preferences. *Economics & Philosophy* 16, 2 (2000), 175–204.

---

<sup>1</sup>It is disputed whether this is possible, compare Kutz (2000) and Isaacs (2011).

<sup>2</sup>Braham and van Hees (2011, p. 13) provide a more nuanced assessment: “we may have to say that m2 and m3 are not responsible for the state of affairs as such – and hence that there is still an epistemic void – although both are blameworthy for holding unjustified beliefs.”

## Raul Hakli (University of Helsinki) and Kaarlo Miller (University of Helsinki) Group Preferences as Total Subjective Comparative Evaluations

Team reasoning is an alternative to traditional game-theoretical accounts of decision making and practical reasoning. Team reasoning in its standard form presupposes a notion of collective preference, but collective preferences are often left undefined and no account on how they are arrived at is offered. We study the possibility of defining collective preferences analogously to individuals' preferences. As argued by Daniel Hausman, in economics, individuals' preferences are usually understood as total comparative subjective evaluations. The question arises what this means in the case of groups. We argue that existing accounts of collective preferences are not compatible with this idea of preferences. We aim at giving an account of collective preferences that is compatible with preferences as total comparative subjective evaluations.

The outline of our talk is as follows:

- Daniel Hausman has argued that preferences should be understood as total subjective, comparative evaluations.
- On the other hand, there are suggestions that groups could be agents and have preferences of their own. In particular, recent theories of team reasoning presuppose a notion of group preferences.
- We study to what extent the above two ideas can be reconciled by reconstructing Hausman's criteria for preferences to apply to group preferences.
- We study existing accounts of group preferences, presented by David Copp, Robert Sugden, and Margaret Gilbert, and argue that they meet some of Hausman's criteria but fail to meet others.
- We conclude by presenting our own account, and we argue that it satisfies Hausman's criteria better than its rivals.

We work on the hypothesis that, analogously to individual agents, a group acts as an agent if it acts on its preferences. And we go on by assuming that a group acts on its preferences if the group members frame the decision problem from the point of view of the group agent and employ team reasoning when choosing their actions. Team reasoning is a method of selecting actions in interaction situations Sugden (1993); Bacharach (1999). It differs from standard game-theoretic methods. The main idea is that agents act as group members in the sense that they figure out which outcome is best for the group and then select their part actions that will lead to that outcome if the others do their parts. Theories of team reasoning presuppose an understanding of what is best for a group, that is, a notion of collective preference. The purpose of this talk is to elucidate this notion.

We assume that whatever functional role individuals' preferences have in accounting for individual action, group preferences have an analogous role in accounting for group action. Based on this assumption we try to formulate criteria for a general account of group preferences. We adopt the basic ideas proposed by Daniel M. Hausman (2012). He argues that, in the actual practice of economics, preferences are understood as total subjective comparative evaluations, and rightly so (Hausman, 2012, 35). This means that preferences are assumed to take into account all considerations that from the agent's point of view affect her evaluation of the outcomes. Hence, the agent's choice depends on her preferences together with her beliefs about the structure of the decision-making situation: From the available actions that the agent believes to be available (and actually are available) she chooses one that, according to her beliefs, leads to a preferred outcome. We apply this idea to group preferences.

The notion of collective preference, or sometimes group preference, is used in the literature in various ways. We can distinguish between three types of accounts, which we call aggregative accounts, team agency accounts, and collective acceptance accounts. Aggregative accounts take individuals' preferences as primary and define group preferences on the basis of individual preferences. Team reasoning accounts are inspired by revealed preference views and take the group-members' team-directed reasoning as primitive: Group preferences are defined in terms of its members' team-directed reasoning. Collective acceptance accounts take the group members' collective acceptance as primitive and define group preferences on the basis of the group's collectively accepted goals and other group attitudes. As to current authors, for example, David Copp (1995) would be classified as a proponent of an aggregative account, Robert Sugden (2000) would be a proponent of a team reasoning account, and Margaret Gilbert (2001) and Raimo Tuomela (2007) would be proponents of collective acceptance accounts.

We study proposed accounts of collective preferences, in particular by David Copp (1995), Robert Sugden (2000), Margaret Gilbert (2001), and argue that they are not compatible with the general idea of preferences as total subjective comparative evaluations. We conclude by presenting our account of collective preferences as derivable from collectively accepted goals and mutual beliefs of the group members, and argue that it satisfies the general criteria of preferences better than the competing accounts.

Such an account can be seen to accord with Hausman's (2012) characterization of preferences as subjective comparative evaluations with respect to everything relevant to value or choice. In a paradigm case, the evaluation proceeds as follows: First, the group has accepted goals or made decisions for itself. In this process the relevant group members have jointly evaluated various relevant options before reaching their conclusion. Second, an individual group member is inferring, or computing, to use Bacharach's term, the group's preference ranking from these group decisions and from other relevant information. Next, given the newly computed group preference,

the group member compares, all relevant things considered, whether her preferences match with those of the group, and, consequently, if she adopts the group's point of view, she identifies her preference ranking with that of her group thereby forming her preferences that correspond to what she takes to be the group preferences.

References:

- Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics* 53, 117–147.
- Copp, D. (1995). *Morality, Normativity, and Society*. Oxford University Press.
- Gilbert, M. (2001). Collective preferences, obligations, and rational choice. *Economics and Philosophy* 17 (1), 109–119.
- Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. New York: Cambridge University Press.
- Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy* 10, 69–89. 3
- Sugden, R. (2000). Team preferences. *Economics and Philosophy* 20, 175– 204.
- Tuomela, R. (2007). *The Philosophy of Sociality*. Oxford University Press.

## Matti Heinonen (University of Helsinki). Discovering the Mechanisms of Joint Action

Analytic philosophers (Bratman 2014; Searle 2010; Tuomela 2013), behavioral scientists (Brownell 2011; Tomasello 2014; Warneken et al. 2006), cognitive scientists (Knoblich et al. 2010; Sebanz et al. 2006), and neuroscientists (Obhi&Cross 2016) have become increasingly interested in studying the phenomenon of joint action during recent years. What is joint action? Roughly speaking, joint action occurs when two or more individuals do something together in a coordinated manner so as to realize their common goal or intention. Some often discussed garden variety examples of joint action include lifting a heavy table together, singing a duet together, and negotiating a business agreement with one's partner. Central theoretical questions that have to do with understanding and explaining joint action include the following: By means of what processes are two bodies synchronized during the execution of joint action? How may two minds meet to coordinate and plan their prospective joint activities? Under what circumstances is it rational for one individual to trust another individual to perform her share of a joint action?

Given their common subject matter, surprisingly little systematic attention has been directed so far at the methodological and theoretical relations between different disciplinary approaches to joint action. Do the methods of different disciplines complement or compete with one another? Should their results be regarded as involving autonomous theoretical posits, or can the outputs of some disciplines be reduced to the theoretical posits and findings of other disciplines? In the absence of any overt conflict between different disciplinary paradigms, we seem to be warranted in making the charitable assumption that the insights that are provided by different disciplinary approaches to joint action are compatible with one another. However, this leaves open a host of important details about how the relevant disciplinary accounts of joint action latch together, where are the joints at which they interlock, and what are the nuts and bolts that tie them together into an overall coherent story. To answer such questions, a more comprehensive methodological perspective is called for.

I will focus on the study of joint action in the behavioral and cognitive sciences, particularly evolutionary anthropology, neuroscience and developmental psychology. In these disciplines, intense study of joint action is currently taking place, bold hypotheses are being advanced, and innovative experimental paradigms are being invented to test these hypotheses. In the behavioral sciences, the study of joint action has converged on two central questions: 1) Are there forms of joint action that are unique to the human species, as contrasted with our nearest primate relatives (e.g. Tomasello 2014; Warneken et al. 2006)? 2) How does the capacity for joint action develop during human ontogeny (e.g. Brownell 2011)? In the cognitive sciences, a central concern has been to find out whether there are dedicated cognitive mechanisms for joint action, which stand in contrast to the cognitive mechanisms that bring about the goal-directed activities of individuals acting on their own (Dolk&Prinz 2016). The influence on joint action of general-purpose mechanisms of social cognition, including mirror neurons, has also received its fair share of attention (e.g. Sartori 2016).

In my presentation, I will apply the framework of mechanistic explanation to explicate how the behavioral and cognitive sciences can make unique contributions to our understanding of the phenomenon of joint action, without either set of disciplines being eliminated or eradicated in favor of the other. I will argue that it is possible for these disciplines to complement one another, because mechanisms bridge levels of explanation, and they are individuated in part by the functions that they perform. In this sense, Bechtel and Abrahamsen (2005, 423) define a mechanism as "a structure performing a function in virtue of its component parts, component operations, and their organization". Their definition makes transparent that a crucial part of the mechanistic explanation of a particular phenomenon of interest is an adequate description of the phenomenon to be explained. Accordingly, it would not be possible to understand the behavior of the human heart, if one assumed that the function of the heart is to function as a giant boom-box, rather than to pump blood to the rest of the circulatory system. In a similar manner, an account of the neural processes that lead to coordinated joint activity would be found wanting in the absence of an adequate description of the forms of activity to be explained. However, it is precisely the functional form of these activities that is in question in many investigations of joint action in the behavioral sciences. Drawing on Bechtel and Abrahamsen's account of mechanistic explanation, I will argue that the behavioral and cognitive sciences are inextricably bound with one another in their explanatory endeavors, and neither can succeed in providing an adequate account of joint action without inputs from the other. I will support this thesis by considering relevant experimental paradigms that have been used for investigating joint action in the behavioral and cognitive sciences, and consider possible objections to my argument that stem from alternative accounts of mechanistic explanation in philosophy of science. I will also draw some tentative conclusions about how interdisciplinary research on joint action could be more successfully carried out.

### Bibliography

- Bechtel, W. & A. Abrahamsen. 2005. "Explanation: a Mechanist Alternative". *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 36: 421-441.
- Bratman, M. 2014. *Shared Agency: a Planning Theory of Acting Together*. New York: Oxford University Press.
- Dolk, T. & W. Prinz. 2016. "What it Takes to Share a Task: Sharing versus Shaping Task Representations". In Obhi&Cross 2016.
- Knoblich, G., S. Butterfill, and N. Sebanz. 2010. *Psychological Research on Joint Action: Theory and Data*. In *Psychology of Learning and Motivation*, vol. 51, edited by B. Ross, 121-54. Burlington: Academic Press.
- Obhi, S. & E. Cross. 2016. *Shared Representations: Sensorimotor Foundations of Social Life*. Cambridge: Cambridge University Press
- Sartori, L. 2016. "Complementary Actions". In Obhi&Cross 2016.
- Searle, J. 2010. *Making the Social World: the Structure of Human Civilization*. New York: Oxford University Press."

- Sebanz, N., H. Bekkering & G. Knoblich. 2006. "Joint Action: Bodies and Minds Moving Together". *Trends in Cognitive Sciences* 10 (2): 70-76.
- Tomasello, M. 2014. *A Natural History of Human Thinking*. Harvard (MA): Harvard University Press.
- Tuomela, R. 2013. *Social Ontology: Collective Intentionality and Group Agents*. New York: Oxford University Press.
- Warneken, F., F. Chen, and M. Tomasello. 2006. "Cooperative Activities in Young Children and Chimpanzees." *Child Development* 77 (3): 640-63.

## Mats Ingelström (Department of Philosophy, Stockholm University). A Dilemma for Validating Subjective Well-Being Measurement

Can subjective well-being (SWB) measurements give us valid measures of well-being? In a reply to Hausman (2015), Anna Alexandrova recently suggested that they can, at least in some contexts (2016). I will challenge this by posing a dilemma for proponents of SWB-measures.

To determine whether a SWB-measure is valid, we must first clarify the relation between the numeric representations the measurement procedure produce and the quantity it is a representation of. Only when we have an answer to this can we start investigating whether the representations are accurate and hence whether the measure is valid. Broadly speaking, there are two competing accounts of the nature of measurements: measurement-realism and measurement-nominalism.

According to measurement-realism, measurements aim to identify/discover properties or relations that are characterized independently of the measurement procedure:

Chang and Cartwright explains this account as: "For the realist, measurement is an activity aimed at discovering the true value of a specified quantity that exists independently of how we measure it, and the question of the correctness of method is certainly not vacuous." (2008, 368–369). A slightly more technical description is proposed by Savage and Ehrlich: "Realism in measurement theory is [...] the view that numerical relations such as order, difference, addition, multiplication, differentiation, and so forth represent real, nonnumerical empirical relations under appropriate scales of measurement ..." (2013, 11)

Measurement-nominalism is the rejection of measurement-realism. On this account, the measured properties or relations are not characterized independently of the measurement procedure, measurements rather define or regulate the usage of a concept within a scientific field.

Whether one opts for measurement-realism or measurement-nominalism may depend on the purpose of the measurement. However, a satisfactory answer to the challenge of assessing whether a measure is valid must assume only one of these two accounts, and each account leads to its own problems for SWB-measurements. My argument poses a dilemma for anyone who wants to defend SWB-measures as valid measures of well-being. Very briefly, the overall argument runs as follows:

1. Either measurement-realism or measurement-nominalism.

2. If measurement-realism, then Chang's problem of nomic measurement (Chang 2004) must be handled. The problem arises since we cannot directly observe the quantity we are interested in, but instead we need to rely on indirect observation. How can we then know that the indirect observation provides accurate information about the true value of the quantity we are interested in? The problem of nomic measurement is a general problem for psychometric measurement, but it is, as I will argue, particularly serious for SWB-measures. The main reason for this is that 'quantity of well-being' is not a theoretical concept internal to the science in which it is measured. Rather, it is a concept that already plays a central role in value theory and policy evaluation. The most promising way of avoiding the problem of nomic measurement is to abandon measurement-realism and opt for some form of measurement-nominalism.

3. If measurement-nominalism, then it is difficult to account for different explanations of the phenomena of response-shift in SWB-studies (Oort et al. 2009; McClimans et al. 2013). A plausible account of well-being will be sensitive to different explanations of the response-shifting. Thus, if measurement-nominalism is true, then SWB-studies face a potentially serious problem, if response-shift is a common phenomenon. Furthermore, if SWB-measures are not aiming at discovering the true quantity of people's well-being, then the evaluative relevance of SWB-measures is less clear than proponents often seem to assume. The most promising way of accounting for response-shift and to guarantee the evaluative relevance is to abandon measurement-nominalism and opt for some form of measurement-realism.

### References:

- Alexandrova, Anna. 2016. 'Is Well-Being Measurable After All?' *Public Health Ethics* phw015
- Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. OUP, USA.
- Chang, Hasok, and Nancy Cartwright. 2008. 'Measurement'. In *The Routledge Companion to Philosophy of Science*, edited by Stathis Psillos and Martin Curd, 1 edition, 367–384. London; New York: Routledge.
- Hausman, Daniel M. 2015. *Valuing Health: Well-Being, Freedom, and Suffering*. Oxford University Press.
- McClimans, Leah, Jerome Bickenbach, Marjan Westerman, Licia Carlson, David Wasserman, and Carolyn Schwartz. 2013. 'Philosophical Perspectives on Response Shift'. *Quality of Life Research* 22 (7): 1871–78.
- Oort, Frans J., Mechteld R. M. Visser, and Mirjam A. G. Sprangers. 2009. 'Formal Definitions of Measurement Bias and Explanation Bias Clarify Measurement and Conceptual Perspectives on Response Shift'. *Journal of Clinical Epidemiology* 62 (11): 1126–37.
- Savage, C. Wade, and Philip Ehrlich. 2013. *Philosophical and Foundational Issues in Measurement Theory*. Psychology Press.

## **Pawel Kawalec (John Paul II Catholic University of Lublin). Towards evolutionary theorizing on science dynamics: generation and production of scientific knowledge**

Science studies have prompted considerations of comprehensive science ontology (Kitcher 2001; Evans & Collins 2002; Mirowski 2011; Zamora-Bonilla 2016) underlying its 'mode of creation' (Gibbons et al. 1994). However, even the more recent 'post-Kuhnian' attempts (Ankeny and Leonelli 2016; Sample 2017) do not fully address science dynamics. My paper is an attempt to fill in this gap by adopting pertinent elements of 'evolutionary theorizing' from philosophy of economics (Nelson and Winter 1982; Winter 1994; Safarzyńska et al. 2012). Taking into account variations in e.g. data sources, inferential patterns, institutional settings, dissemination patterns, etc. I distinguish two basic forms of scientific knowledge creation (in terms of BV+SR modeling, e.g. Campbell 1960; Simonton 2012; Nickles 2016): generative vs. production models. While they both, as I argue, stem from research routines, the principled distinction between them follows T. Williamson's (2002) opposition between knowledge (as factive, prime and broad condition) and justified belief. These two are used to define payoff matrix in Bacharach's variable frame theory of coordination games (with the conditioning evidence set for more specific payoff matrix). When discussing the n level variant of the model I draw on the result that there may exist probability distributions, which force Pareto optimal team-coordination equilibria (focal points) to pick up the (epistemically sub-optimal) routine practices even in the case of team members's awareness of epistemic optimality of the genuine knowledge-seeking option. My solution to this - epistemically paradoxical, but behavioristically realistic - result draws upon institutional setup of team coordination, which may determine the probability distribution in order to balance the rational/deliberative vs salient solutions. In discussing the institutional rules I draw on empirical research on the best performing research teams (author 2016; Boyack et al. 2017). I also substantiate the conclusion with my recent experience in participating in EC Expert Group (<http://bookshop.europa.eu/en/changing-gear-in-r-i-pbKI0216237/>).

I further substantiate the difference in dynamics behind generative vs. production models – as captured by Simonton's criterion of creativity and deliberative team cooperation - by examining the underlying combinatorial stochastic processes in terms of mixed nonparametric Bayesian models (beta with Poisson and Bernoulli processes with EPPF vs. EPPF representations). Knowledge production model is shown to be less creative, more prone to sightedness and frame lock-in.

## **Kaisa Kärki (University of Jyväskylä). Not doings as resistance: Conceptual issues between social sciences and the philosophy of action**

What does it mean when an agent intentionally does not do something as resistance toward that action or something that the action represents? An agent can resist the army by refusing to be drafted, resist the government with a hunger strike or resist the social system by not voting. In this paper it is argued that in social science a general concept of refusing to act as resistance is needed to grasp the ordinariness of this kind of resistance.

How have social scientist previously conceptualized this kind of resistance? When talking about resistance in social science, sometimes a contra-attitude, such as distrust or dissatisfaction of the agent is the central content of the discussion. Some concepts of resistance, namely passive resistance and non-participation, refer to the action not done due to this contra-attitude. Sometimes resistance is talked about as referring to the expressions of this contra-attitudes, as when talking about the hidden transcripts of subordinate groups (Scott 1990), protests or conscientious objections. Some concepts of resistance, that is, non-compliance, civil disobedience or conscientious refusing, grasp the breaking of an expectation, norm or a rule. Are the concepts used in social science enough to pick the scope of this kind of resistance in its entirety?

To answer this question I present a categorization of different concepts used in social sciences of this kind of resistance. The idea is to provide a preliminary conceptual platform for unification of resistance concepts in social sciences and philosophy. I divide the concepts into expressive concepts (such as protest, hidden transcript, objection), disobedience concepts that are about breaking an expectation or rule (such as disobedience, noncompliance and conscientious objection), contra-attitude concepts (such as dissent, distrust and dissatisfaction) legal concepts (such as delegitimation, civil disobedience and breaking a contract) and action concepts (such as non-participation, passive resistance and refusing to act), in which a contra-attitude of the agent is transformed into an intentional omission.

Based on this categorization I argue in this paper that previous concepts used in social sciences, such as conscientious refusing, conscientious objection, civil disobedience and non-violence, cannot entirely pick out the ordinariness of this kind of resistance. Not doing something out of resistance is a normal part of everyday life of agents. Not doings as resistance are not necessarily public, articulated or organized, nor are they necessarily targeted against a specific law, a rule or even an expectation. This is why it is argued that a general notion of refusing is needed in social science to grasp the ordinariness of not doings as resistance.

Philosophy of action has not discussed not doings as a form of resistance, but the generality of the approach is needed when drafting this more general notion of not doings. This concept of refusing also builds on Albert Hirschman's analysis of the dynamic between exit and voice (1970) and James Scott's notion of hidden transcript which was able to grasp the diversity of resistance behaviors. In this paper I argue that what is needed for the most minimal concept of refusing to act as resistance is the agent intentionally not doing some action due to a contra-attitude toward that action or something the action represents. A mere preference is not enough to account for this kind of resistance either, because this kind of not doings cannot be reduced to the action the agent does instead of the action not done. A need for new concepts to understand modern resistance is often articulated by social scientist, which is what this paper is trying to answer in part.

The notion of refusing to act as a form of resistance links the discussions of resistance in social sciences to conceptual rigor of philosophy of action. The idea is to create a common ground for talking about these often hidden forms of resistance, that can be hard to distinguish from mere passivity, but would need to be taken into consideration in policy-making. If only the official, visible forms of resistance such as civil disobedience are recognized by the social scientist, a big part of societal resistance that can take a form of non-participation, for instance, is left outside theory.

### References:

- Hirschman, Albert O. (1970): "Exit, Voice, and Loyalty – Responses to Decline in Firms, Organizations, and States." Cambridge and London: Harvard University Press.
- Scott, James (1990): "Domination and the Arts of Resistance: Hidden Transcripts." Yale University Press.

## Inkeri Koskinen (University of Helsinki). Defending a contextual account of objectivity

It has become clear in the recent philosophical discussions of scientific objectivity that the notion is irreducibly complex. When we call X objective, X can be, for instance, a researcher, a research process, a knowledge claim, or a research community. And when doing so, we use the adjective "objective" in many different ways. Heather Douglas (2004, 2009) has recognised eight conceptually distinct senses of objectivity. For instance, in manipulable objectivity we call objective something we can manipulate: if we can use that something as a tool, we have good reasons to say that this something really exists – it is objective. In procedural objectivity the research process has been designed so that a researcher can always be changed to another, and that will not change the result. And in interactive objectivity a research community follows practices that ensure effective critical discussions and debates. The different senses of objectivity are clearly distinct. However, as Douglas argues, in all cases objectivity indicates a shared basis for trust in a knowledge claim. When we call something objective, we state that we trust it or rely on it, and that others should do so, too.

Julian Reiss and Jan Sprenger (2016) have noted that objectivity cannot be identified with features that promote trust in science, as trust can be misguided. Of course no-one has suggested that felt trust be used as an indicator to recognise objectivity. However, the observation does point out a problem: Douglas has identified several bases for trust that do indeed seem to justify the assessment that X is objective – we have good reasons for trust. But we can trust also without having good reasons for doing so. Do the good reasons identified by Douglas have something in common?

I believe they do. I suggest that the different senses of objectivity Douglas mentions have more in common than she believes. This becomes clear when objectivity is seen as a negative term.

Lorraine Daston and Peter Galison (2007) have noted that when new senses of objectivity have emerged in science, they have been related to newly recognised epistemic threats. Recently Ian Hacking (2015) has claimed that "objective" is a negative adjective: it marks the absence of this or that vice. I would however like to point out that, firstly, not all epistemic threats count, and, secondly, all epistemic threats that do count are not vices.

For instance, in archaeology results can be skewed because the available data is skewed in random ways archaeologist cannot possibly control or detect. This is an epistemic threat, but it does not make the research nor the results biased. It is an epistemic threat that is not related to objectivity.

On the other hand, I believe it is somewhat misleading to talk about vices. To clarify this let us return to the different senses of objectivity Douglas has described. In all cases she mentions or at least clearly implies a specific epistemic threat that is being averted; a contrast to objectivity. These contrasts include illusions, subjectivity, idiosyncracies, and collective biases. Can they be called vices? Is it a vice to be fooled by an illusion? I prefer to say that they are all epistemic threats arising from our imperfections as epistemic agents. So I would like to suggest the following:

When we call X objective, we do indeed endorse it, as Douglas recognises: we say that we trust X or rely on X, and that others should do so, too. But the word "objective" is reserved to a specific type of confidence: it is based on the belief that some important epistemic threat arising from our imperfections as epistemic agents has been effectively averted.

I call this account of objectivity contextual for two reasons. First, we are imperfect as epistemic agents in many ways: we are prone to wishful thinking, we can be fooled by illusions, we easily adopt ideas and values shared in our communities and do not question them, etc. Our failings as epistemic agents lead to many different kinds of epistemic threats. It is plausible that some of them are particularly pressing in some contexts, other ones in others. So what is an important threat may depend on the context. Secondly, this account separates strategies developed for averting some threat from the threat itself. There is no need to think that a specific threat could not be averted in several different ways. Some are more appropriate in some contexts, others in others.

This brings some unity to the diverse senses of objectivity. As Douglas points out, none of the recognised senses of objectivity is reducible to the others. This becomes quite understandable when we realise that each sense identifies an efficient strategy for averting some important epistemic threat (or several threats) arising from our imperfections as epistemic agents. Even a single threat can be averted in several different ways. For instance, we can ensure that an object really is there, instead of being an illusion, by using it as a tool, as in manipulable objectivity, or by using several independent methods of observation, as in convergent objectivity. As our failings as epistemic agents give rise to several epistemic threats, the number of distinct senses of objectivity is bound to be high.

Daston, L. & Galison, P. 2007. *Objectivity*. New York: Zone.

Douglas, H. 2004 *The Irreducible Complexity of Objectivity*. *Synthese* 138:3, 453–473.

Douglas, H. 2009. *Science, Policy and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.

Hacking, I. 2015. *Let's Not Talk about Objectivity*. Teoksessa Padovani, Richardson & Tsou (eds.). *Objectivity in Science*. *New Perspectives from Science and Technology Studies*. Dordrecht: Springer, 19–33.

Reiss J. & J. Sprenger. 2016. *Scientific Objectivity*. In Zalta, E. N. (ed.). *The Stanford Encyclopedia of Philosophy* (Summer 2016

edition).

## **Edwin Koster (Vrije Universiteit Amsterdam) and Peter Versteeg (Vrije Universiteit Amsterdam). The insider/outsider problem: beyond methodological ludism?**

Issues about views and convictions, norms and values, and rules and beliefs have always played a role in the philosophy of the social sciences. These issues are, on the one hand, studies by the social sciences. But on the other hand (and this perspective is perhaps more interesting from a philosophical point of view) these issues are also involved regarding the stance of social scientists. Questions about values and objectivity could be raised or, for instance, one can reflect on the influence of ethical convictions and political ideologies on social scientific work. In my paper I will ask this type of questions with reference to anthropology of religion.

The most persistent methodological issue in the study of religion is the so called 'insider / outsider problem'. In a nutshell the problem is whether and to what extent someone can study, understand, or explain the beliefs, words, and actions of another person. The insider claims that the essence of a religious tradition can only be understood from within. For centuries this approach was not disputed: it was the traditional way to study religion and it was strongly supported by powerful institutions such as the church. Sometimes this approach is considered to be the only suitable perspective to study religion. Some Muslim writers for example claim that 'true knowledge' of Islam is only available from an insider's point of view. However, since Greek philosophy promoted the spirit of detached inquiry, a second approach to the study of religion came into being: a view from without. It took a long time to institutionalize the possibility to study religion without the control of the church or other religious institutions, but when this was finally realized in the 1960's, it gained more and more influence. Following this approach, the scholar of religion tries to build theories from the outside. It is based on the assumption that only what can be observed empirically is worth to study; its goal is to determine the causes and regularities of human actions and beliefs.

Some objections against these two positions have been made. Is the first position not too quick in validating the claims of the insider? And is the emphasis on the development of explanatory theories not an unjust reduction or even elimination of insider's claims? Illustrative is the struggle between Muslim and Western writers on Islam. In the eyes of Muslims (as insiders), Western scholars (as outsiders) present a distorted and one-dimensional picture of Islam, while Western researchers characterize Muslim writings as subjective and apologetic.

As is well-known, several solutions to this dilemma have been offered. Best known are the ones which suggest to remain 'neutral' (methodological agnosticism) and the ones which recommend to take a reflexive stance. Less known is an approach called 'methodological ludism', developed by the Dutch anthropologist André Droogers.

'Methodological ludism' refers to the human capacity to play: to deal simultaneously and subjunctively with two or more ways of classifying reality. Droogers means that scholars of the study of religion must not confine themselves to one point of view. Instead of exclusively opting for one approach, such as methodological agnosticism, Droogers suggests to view religion from a vantage point that is opposite to the one usually assumed.

This approach has some advantages, as will be explained. However, empirical studies show that it is not so easy for people belonging to certain religious traditions to identify with the position of believers of other traditions - and, strikingly, some atheists reported that they were influenced deeply by the faith of their subjects thus failing to maintain a neutral stance. The scholar's scientific attitude is thus affected by the method of participant observation. The results of these empirical studies have consequences which go beyond the classical insider / outsider problem.

In my paper I will show that methodological ludism has some interesting characteristics, but that it is, in the end, not realistic, due to the condition of 'situatedness'. Most of the time scholars do not consciously know their conditions of situatedness. They belong to the embodied and unconscious part of their identity. In spite of this, they still can be made productive. In explaining how this is possible I propose a revision of Droogers' methodological ludism.

From the perspective that in the study of religion different methodologies are necessary, it is desirable to approach religion from the outside looking for empirically testable explanations and to approach it from the inside, trying to explain religion in its own right. But because of the embodied and thus situated condition of every scholar, it is not possible to locate the ideal of a multitude of perspectives in the individual scholar, it must be put in practice in the community of scholars of religion. The situatedness of individual scholars can then be made productive: from their different perspectives ('emotional' (and scholarly)) they disclose different aspects of religion and by confronting these results with each other obscure subjective perspectives can be eliminated.

## Raphaël Künstler (Institut Jean Nicod). Introspective social sciences

If social sciences are subjective, they are not social sciences (Durkheim, 1896); if they do not involve subjectivity, they are not social sciences (Weber, 1978). Methodological doctrines are attempts to solve this dilemma. On one side of the spectrum, naturalism excludes subjectivity, on the other side, hermeneutics gets rid of objectivity (Mantzavinos, 2005).

Drawing on Matzavinos's "naturalistic hermeneutics", my paper claims that introspective inquiry should and could be a part of a scientific social inquiry. In order to vindicate this claim, two other claims should be defended:

- (1) Introspective knowledge is a necessary condition of social knowledge.
- (2) Introspective inquiry does not preclude objectivity.

There are two ways of understanding the first claim. Introspection can be used in order to identify and neutralize prejudices, desires, tastes, associations, values, which could prevent the researcher from being objective. This is why cultural anthropologists, following Malinowski (1967), use a diary during their inquiries. But introspection could also be used in a positive manner, that is, not to eliminate one's subjectivity, but rather to find in one's own experience an explanation to someone else's behavior.

In order to support such positive use of introspection, the paper draws on the case of Jonathan Cohen's distinction between belief and acceptance, which has received a lot of attention among social scientists. Without such distinction, it would be difficult to understand the behavior of a jury during a trial or the behavior of scientists towards their theories (Cohen, 1992), to analyze a terrorist's decision (Bouvier, 2016) or to analyze ecological militancy (Allouche, 2016). Now, this distinction originates from a rational introspective inquiry, namely, the cartesian philosophy. Therefore, without introspection, there are social phenomena which could not be explained.

In order to vindicate the second claim, it is necessary to distinguish a projective from a constructive use of subjectivity. Introspective projection leads to misunderstandings, as in the case of Eli Chinoy (1955) puzzlement in front of workers' satisfaction with their situation (Boudon, 1986). Constructive introspection leans from subjective experience how to distinguish between cognitive states which would otherwise remain undetectable, as in Cohen's case. It does not immediately apply to other's behavior. It is then the naturalistic hermeneutics which enables scientists to separate grain from shaft.

### References

- Allouche, A. (2016), "L'effet surgénérateur des attentes mutuelles d'assentiment sur les protestations collectives", in Bouvier & Künstler (2016)
- Bouvier, A. (2016), "Croyances collectives, assentiments collectifs et argumentation en contexte religieux", in Bouvier & Künstler (2016).
- Bouvier, A. & Künstler, R. (2016), *Croire ou accepter*, Hermann, Paris.
- Boudon, R. (1986), *L'idéologie, ou l'origine des idées reçues*, Fayard, Paris.
- Chinoy, E. (1955), *Automobile Workers and the American Dream*, University of Illinois Press, Chicago.
- Cohen, J., 1992, *An Essay on Belief and Acceptance*, Oxford University Press, Oxford.
- Malinowski, B. (1967) *A Diary in a Strict Sense of the Term*, Athlone Press, London.
- Mantzavinos, C. (2005), *Naturalistic Hermeneutics*, Cambridge University Press, Cambridge.
- Weber, M. (1978), *Economy and Society*, University of California Press, Berkeley.

## Charles Lowe (University of Osnabrück). Self-fulfilling theories, meta-theories, and values in science

Heated debate in the literature surrounds the questions of whether some scientific theories are self-fulfilling phenomena and what consequences this fact would have for our understanding of science in theory and practice. For example, studies suggest that economics students who are taught that human action is motivated primarily by self-interest are more likely to act in a manner conforming to this theory than their peers. This has led to claims (e.g. Ferraro et al. 2009) that such theories may actually prescribe, shape or even cause the behavior they are usually taken to merely describe, and that researchers are obliged to take this possibility into consideration when developing and disseminating their theories. Others (e.g. Felin and Foss 2009) have strongly criticized such claims.

Bergenholtz and Busch (2016) attempt to “cool the fire” in this debate by introducing the concept of a meta-theory. A meta-theory “takes theories about empirical phenomena as its subject of explanation”, “predicts if there is a self-fulfilling impact of the adoptions of (first-order) theories”, and “identifies the specific theoretical mechanisms that constitute this impact” (p. 36). Using this concept, the authors attempt to allay fears that the existence of self-fulfilling theories undermines either scientific realism (as they understand it) or the integrity of individual scientists. The introduction of meta-theories is a step in the right direction toward untangling some of the complexities surrounding the significance of self-fulfilling theories. However, Bergenholtz and Busch’s treatment misses the mark in two important ways. First, their conception of meta-theories is both overly restrictive and underdeveloped. More importantly, their characterization of the ‘problems’ self-fulfilling theories raise and their proposed ‘solutions’ via meta-theories is formulated at a level that is so abstract as to obscure several more pressing issues.

In this paper, I develop a more nuanced account of meta-theories as they relate to self-fulfilling theories and some of the issues they raise from the perspective of philosophy of science. In developing this account, I examine a claim recently put forward by Matthew Kopec (forthcoming) to the effect that the apparently successful use of the game theoretic model of the ‘tragedy of the commons’ to predict the behavior of nation-actors in international climate negotiations may be due to the model’s ability to generate self-fulfilling predictions.

I begin by identifying a number of questions we might ask about self-fulfilling theories and the states of affairs or situations they pertain to. Some of these are concerned with the current situation and how we got here, for example: “Does the theory adequately account for current observations?”, “Did the currently observable situation come about because of some previous application of the theory in question?”, and “If the current situation is a product of self-fulfillment, what mechanisms were involved in this process?” Other questions deal with future prospects, for example: “Is the future development of the current situation (or lack thereof) dependent in some way upon continued belief in the adequacy of the theory?” and “Would such belief be significantly affected if scientists were to adopt a different set of attitudes toward the theory?”

While we may reasonably classify theories that provide answers to questions like these as ‘meta-theories’, it is important to be aware of the diversity of the relevant objects of investigation and methods involved. Examination of these issues shows the need for additional conceptual categories for describing theories that are often lumped together under the term ‘self-fulfilling’. For example, we should distinguish between theories that have self-fulfilled, in the sense that their current explanatory and predictive success came about due to their previous application, from those that are self-sustaining, in the sense that their continued success is dependent upon continued belief in the theory’s adequacy.

In the second half of the paper, I employ this more nuanced account of meta-theories to elucidate two pressing issues that self-fulfilling theories raise from a philosophy of science perspective. The first is the question of how meta-theories should affect (epistemic) evaluations of the ‘first-order’ theories they pertain to. The fact that the current predictive success of the ‘tragedy of the commons’ model may be a product of self-fulfillment should not diminish our confidence in its predictive success. However, it may call for a reevaluation of what state of affairs it is, exactly, that the theory in question is taken to describe and explain.

The second issue concerns the difficult situation scientists are put into when dealing with theories that are clearly epistemically validated from the current point of view but which they have good reason to believe may be self-fulfilling or self-sustaining. The problem is particularly thorny when an undesirable future state of affairs is likely to be the result of a self-fulfilling or self-sustaining theory. Kopec’s example is a case in point: the ‘tragedy of the commons’ model is widely considered to be epistemically superior to other attempts to predict behavior in climate negotiations, yet it also predicts the inevitable failure of such negotiations to help us avoid future climate crisis. If this model is truly self-sustaining, it seems we have good reasons both to accept and employ the model as well as to reject it. To conclude my paper, I suggest how this issue links up with current debates (Elliot & McKaughan 2014) about the role of and potential conflicts between epistemic and non-epistemic values in science.

### Literature

Bergenholtz, C. and Busch, J. (2016), “Self-Fulfillment of Social Science Theories: Cooling the Fire”, *Philosophy of the Social Sciences*, 46(1): 24-43.

Elliot, K. C. and McKaughan, D. J. (2014), “Nonepistemic Values and the Multiple Goals of Science”, *Philosophy of Science*, 81: 1-21.

Felin, T. and Foss, N. J. (2009), "Social Reality, the Boundaries of Self-Fulfilling Prophecy, and Economics", *Organization Science*, 20(3): 654-668.

Ferraro, F., Pfeffer, J. and Sutton, R. I. (2009), "How and Why Theories Matter: A Comment on Felin and Foss (2009)", *Organizational Science*, 20(3): 669-675.

Kopec, M. (forthcoming), "Game Theory and the Self-Fulfilling Climate Tragedy", *Environmental Values*.

## Andres Luco (Philosophy Programme/Nanyang Technological University).The Evolution of Inclusive Morality

Human moral psychology is a suite of innate cognitive capacities and motivations that prompt human agents to learn, comply with, and enforce norms. In addition, there is broad consensus that human moral psychology includes motivations—such as sympathy and empathy—that induce altruistic behavior (Chudek et al. 2013). Several researchers have suggested that human moral psychology evolved in the cauldron of war (Sober & Wilson 1998; Bowles & Gintis 2011). As a result, moral psychology is said to contain strong xenophobic and ethnocentric biases. Such biases are thought to imply that so-called “inclusive moralities” could not have evolved either as social-psychological adaptations to selective pressures, nor even as byproducts of adaptations (Buchanan & Powell 2015). Allen Buchanan and Russell Powell define inclusive moralities as systems of norms that “extend moral standing to all human beings and even to some nonhuman animals, irrespective of their group membership or strategic capacities (i.e., their ability to contribute to or disrupt cooperation)” (Powell & Buchanan 2016: 987). In two recent articles, Buchanan and Powell argue that none of the “received” evolutionary explanations of moral psychology can explain the rise of inclusive moralities (cf. Buchanan & Powell 2015, 2016). The reason, Buchanan and Powell contend, is that these explanations imply that the evolutionary history of moral psychology is “essentially an in-group affair,” which is to say that moral psychology evolved as a solution to the adaptive problem of motivating members of a social group to cooperate with one another so that, together, they could compete more successfully against other groups (Buchanan & Powell 2015: 42). If the received evolutionary explanations are correct, then our evolved moral psychologies would never have extended impartial moral concern to individuals who are not members of our perceived “in-group,” because doing so would have made our ancestors relatively weak competitors in inter-group struggles for existence. Therefore, Buchanan and Powell maintain, inclusive moralities cannot be explained by the mechanisms postulated by the received evolutionary explanations.

This essay makes the case that an evolutionary explanation may well account for the development of inclusive morality. Two major claims will be defended. First: the evidence that human moral psychology was an adaptation to inter-group warfare is ambiguous, and does not rule out an evolutionary explanation of inclusive morality. And second: cultural evolutionary explanations can explain the evolution of inclusive moralities. I shall elaborate the rationales for both these claims below.

Samuel Bowles and Herbert Gintis have argued that parochial altruism is a feature of human moral psychology that evolved by group selection (Bowles & Gintis 2011). Group selection is the selection mechanism by which individually costly behaviors, and their psychological underpinnings, can evolve on the condition that the group-level fitness advantage generated by the behaviors exceeds the individual-level fitness cost of the behaviors. Parochial altruism is the individually costly tendency for an agent to act altruistically toward members of the same group, but aggressively towards outsider groups. Bowles and Gintis suggest that most observed altruistic behavior in humans is an expression of parochial altruism. They also contend that parochial altruism evolved via group selection, whereby groups that were more parochially altruistic were more successful in defeating other groups in warfare. However, experimental studies have revealed that, contrary to Bowles and Gintis’s prediction, altruistic behaviors and personality traits are not correlated with bias favoring one’s in-group (Corr et al. 2015). In other words, it is not the case that the more altruistic a person is, the more parochial they are. Also, behavioral experiments have found that when people can choose between (1) promoting benefits to their in-group without affecting any out-group, and (2) promoting benefits to their in-group and reducing benefits to out-groups, they tend to select the first option (Yamagishi & Mifune 2015). This, too, goes against the claim that human altruism is by and large an expression of parochial altruism.

Furthermore, while group selection may have been a crucial factor driving the evolution of human moral psychology, the available evidence does not establish that group selection operated primarily through hostile military conflicts between groups. Our evolutionary ancestors depended on peaceable interactions with outsider groups. These interactions were both frequent and fitness-relevant. Observations of hunter-gatherer societies whose ways of life are thought to be similar to those of our ancestors reveal that these foraging societies often engage outsiders in military alliances, material trade, resource-sharing, offering temporary sanctuary, gift-giving intermarriage, and child adoption (Sterelny 2012: 186 – 190; Chudek et al. 2013: 434 – 435; Richerson & Henrich 2012: 55). Given these observations, it stands to reason that group selection pressures may have favored a more inclusive moral psychology, while disfavoring a rigid, inflexible antipathy toward outsiders.

The second main claim of this essay is that cultural evolutionary explanations can explain the evolution of inclusive moralities. Buchanan and Powell explicitly deny even this claim (Buchanan & Powell 2016: 1000). Evolutionary explanations are not limited to accounts of the evolution of genetically inherited traits; they include accounts of the evolution of culturally transmitted traits, or cultural variants. Accordingly, even if human beings have innate xenophobic and ethnocentric biases, such biases do not necessarily preclude the cultural evolution of inclusive moralities. In *The Descent of Man* (1871), Darwin expressed the idea that the operation of reason explains how people’s sympathetic concern may expand from the self, to the in-group, and potentially to all of humankind. This speculative hypothesis about the emergence of inclusive morality is compatible with a cultural-evolutionary explanation. For the operation of reason may be classified as a decision-making cultural evolutionary force, such as direct bias (Boyd & Richerson 2005). In direct bias, a cultural variant, such as a norm or a practice, is preferentially adopted because of its intrinsic qualities. Inclusive moral norms that require the ascription of equal moral standing to all persons may be adopted just because there is no salient difference between those insiders with whom our moral psychology inclines us to sympathize by default, and those persons we have never

interacted with or never will encounter. Although our xenophobic biases may induce us to sympathize immediately with members of our in-group, deliberate reflection on the plight of outsiders may lead us to recognize similarities between them and our in-group which are so psychologically prominent that our sympathies naturally extend beyond their default confines. We may follow Darwin and others in calling this process of sympathetic extension an operation of reason. But whatever one calls it, it is a process that can be captured by the conceptual repertoire of cultural evolutionary theory.

## Mariusz Maziarz (Wroclaw University of Economics). Is a Unified Philosophy of Economics Possible?

Short summary:

In the current philosophy of economics, there are several research programs which seem contradictory or incomprehensible to each other. In my presentation, I address the question whether a unified philosophy of economics is possible. To do so, I reconstruct epistemic guidance and ontological commitment delivered for economists by each of the six schools in the philosophy of economics. Considering that philosophies of science can be either descriptive or normative and either concerned with ontology or with epistemology, I divide the six seemingly-divergent paradigms into the above-mentioned four fields of the philosophy of economics and consider the remaining discrepancies with an aim to establish a unified philosophy of economics. The systematization I coined shows that the discrepancy between scientific realism and critical realism widely discussed in the philosophy of economics is spurious.

Full abstract:

In spite of several attempts at delivering a unification (e.g. Mäki 1988), the contemporary philosophy of economics is a branch of the philosophy of science that seems to be practiced by philosophers and methodologists with contradictory viewpoints. Therefore, there are a few (six, according to my classification) schools of thought or research paradigms, to use the Kuhnian expression, that are believed to deliver contradictory answers to problems of philosophy of science. In my presentation, I address the question whether logical positivism, scientific realism, critical realism, falsificationism, instrumentalism, and epistemological anarchism (as peculiarly developed in the philosophy of economics) are indeed contradictory or a unified philosophy of economics is a possibility in a foreseeable future. Philosophy of science is usually divided into (1) ontology and epistemology (Furlong and Marsh 2010), and (2) normative and descriptive analyses (Scerri 2006). In my research, I reconstruct the epistemic guidance delivered by the six schools of the philosophy of economics and systematize them along the two dimensions of division into four groups, cf. Table 1.

It should be noted that my reconstruction of the six paradigms present in the philosophy-of-economics literature is in line with how these paradigms are understood by philosophers of economics instead of philosophers of science. Specifically, logical positivism is identified with works of Gartley (1935), Hutchison (1938; 2000) and Hill (1968). The scientific-realist project in the methodology of economics is elaborated most notably by Mäki (1990; 2001; 2005; 2008) and Cartwright (1983; 1994). Lawson (e.g. 2006) introduced Bhaskar's critical realism to the philosophy-of-economics literature. Falsificationism was originally advocated by Blaug (1992) in line with Agassi's reconstruction of the Popperian thought. Instrumentalism in the philosophy of economics is identified with Knight (1935) and Friedman's famous essay (1953) inspired by Dewey's (1938) thought. Epistemological anarchism in the philosophy of economics was introduced by McCloskey's (1998) book focused on rhetorical devices employed by economists interpreted in line with Paul Feyerabend's anarchism.

Table 1: A systematization of the six schools in the philosophy of economics

	Ontology	Epistemology
Normative	<ul style="list-style-type: none"> <li>critical realism</li> <li>logical positivism</li> <li>falsificationism</li> </ul>	<ul style="list-style-type: none"> <li>critical realism</li> <li>Friedman's instrumentalism</li> <li>logical positivism</li> <li>falsificationism</li> </ul>
Descriptive	<ul style="list-style-type: none"> <li>scientific realism</li> <li>falsificationism</li> </ul>	<ul style="list-style-type: none"> <li>epistemological anarchism</li> <li>scientific realism</li> </ul>

Employing my classification to addressing the question whether a unified philosophy of science is possible solves a few hitherto unsolved issues. For instance, distinguishing between the normative and descriptive dimensions of ontological inquiry in the philosophy of economics finishes the discrepancy between two seemingly contradictory realist research programs (i.e. scientific realism and critical realism). Scientific realism constructs ontological worldview considering the practice of economists. In contrary, critical realism assumes certain features of the economic world (e.g. being an open system, free will of economic agents etc.) and grounds its epistemic guidance in these assumptions.

However, there still are several discrepancies to be solved, most of which can be seen in the 'normative epistemology' part of Table1. For example, the epistemic guidance formulated by philosophers of economics working in the critical realist paradigm contradicts the other three schools of thought inscribed into 'normative epistemology' part of Table 1 (i.e. Friedman's instrumentalism, logical positivism, and falsificationism), which are coherent to some degree. In order to solve this issue, I advise rejecting the critical-realist project on the grounds that (1) the Lawsonian worldview is formulated a priori and is mistaken (e.g. free will of economic agents contradicts the current status of neuroscientific, psychological and philosophical discussions), (2) his postulate of rejecting econometrics and mathematical analyses is, to some degree, fulfilled by the institutional economics that seems to have poorer results than the mainstream economics.

References:

- Mäki, U. (1988). How to combine rhetoric and realism in the methodology of economics. *Economics and Philosophy*, 4(01), 89-109.
- Furlong, P., & Marsh, D. (2010). A skin not a sweater: Ontology and epistemology in political science. *Theory and methods in political science*, 184-211.
- Scerri, E. R. (2006). Normative and descriptive philosophy of science and the role of chemistry. In *Philosophy Of Chemistry* (pp. 119-128). Springer Netherlands.
- Mäki, U. (1990), Scientific realism and Austrian explanation, *Review of Political Economy*, 3(2), 310-344.
- Mäki, U. (2001). The way the world works (www): towards an ontology of theory choice. In: *The Economic World View. Studies in the Ontology of Economics*, Mäki, U. (ed.), London: Cambridge University Press, 369-384.
- Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, 12(2), 303-315.
- Mäki, U. (2008). Realism. In: Hausman D. (ed.), *The Philosophy of Economics. An Anthology*, Cambridge: Cambridge University Press, 431-439.
- Cartwright, N. (1983). *How the Laws of Physics Lie*, Oxford: Clarendon Press.
- Cartwright, N. (1994). *Nature's Capacities and their Measurement*. London: Oxford University Press.
- Lawson, T. (2006). *Economics and reality*. Routledge.
- Blaug, M. (1992). *The methodology of economics: Or, how economists explain*. Cambridge University Press.

## **Dominika Motak (Jagiellonian University, Cracow). Capitalism as Religion? Rethinking Walter Benjamin's Thesis**

This paper will reflect on the much discussed claim about religious character of modern capitalism. One of the best-known classical formulations of this claim is contained in a rather cryptic fragment left by Walter Benjamin, entitled „Capitalism as Religion”. According to Benjamin, capitalism is „not only a religiously conditioned construction”, but „an essentially religious phenomenon”: „Christianity did not encourage the emergence of capitalism, but rather changed itself into capitalism.”

Benjamin's thesis is usually interpreted as an inversion of Max Weber's conceptualisation of the relationship between Christianity and the capitalist system. There is, however, some evidence that the stimulus for this inflection of arguments was given by Weber himself. On the other hand, „Capitalism as Religion” might as well be seen as a continuation of Georg Simmel's study of the structural homology of religion and monetary economy.

Georg Simmel is widely recognised as a prominent turn-of-the-century German philosopher of culture and one of the founding fathers of sociology, but he could quite properly also be described as a classical theorist of religion. Despite his eminent position in the history of Western thought, Simmel's work on religion continues to be neglected – especially by comparison with that of Durkheim and Weber (Lechner 1990: 169). His concept of religion was not yet reflected in its full complexity, but was only occasionally referred to in the discussions concerning particular topics (Krech 1998: 1), and the task of incorporating it into the contemporary academic discourse remains unfulfilled. I will try to demonstrate that this still not properly appreciated part of Simmel's legacy is most intimately connected with the two much better known areas of his research: monetary culture and general sociology. A rough sketch of Simmel's concept of religion sheds light on Simmel's theory of society and allows to question the prevailing view that the Simmelian 'impressionist' sociology stands in sharp contrast to the classical French 'positivist' tradition of sociological inquiry, as epitomised by Émile Durkheim. The main subject of this chapter – Simmel's thesis on homology of money, society and the idea of God – provides a good argument that, even many decades later, Robert A. Nisbet's (1959: 81) appraisal of Simmel as the most relevant of all the pioneers of sociological reflexion still holds true.

The thesis on homology of money and the idea of God is a recurrent theme in Simmel's writings. In his philosophically informed sociology money serves as a symbol of modernity, expressing its contradictory character. Moreover, being a symbol of unity, money has remarkable affinity with the Judeo-Christian concept of God. At the same time, God constitutes a conceptual equivalent of society; and a notion of society shares certain characteristics with money.

The main common feature of money and society lies in fact that they both are transindividual constructs of universal and yet not abstract character. What is even more important, both money and society are based on acts of exchange. In *Philosophy of Money* Simmel argues that exchange does not simply promote sociation, but exchange itself is a form of sociation: a relationship which transforms a sum of individuals into a social group (Simmel 2004: 175). As Frisby (1992: 12) observed, Simmel took exchange to be 'both paradigmatic and symbolic of society as a whole'. Being a form of sociation, exchange reduces 'the human tragedy of competition' (Simmel 2004: 291). Money serves as a universal medium of exchange. It can be said that money (as a petrified form) is equivalent to exchange (as an action); in his own words: 'money belongs to the category of reified social functions. Like a flag incarnates the unity of a regiment, so money incarnates the function of exchange as a direct interaction between individuals' (Simmel 2004: 175). Already in *Money in Modern Culture* (1896) Simmel stated that 'money ties people together, for now everyone is working for the other'; 'money provides a common basis of direct mutual understanding'; it also 'serves as an ideal adhesive' in a certain type of social organisations, which 'represent one of the most enormous advances of culture' because 'it offers the only opportunity for a unity which eliminates everything personal' (for instance, trade union has only become possible by virtue of money). Therefore he concludes that 'money creates an extremely strong bond among the members of an economic circle. Precisely because it cannot be consumed directly, it refers people to others' (Simmel 1997a: 246ff.). To sum up: for Simmel, money is 'entirely a sociological phenomenon, a form of human interaction' (Simmel 2004: 172) and 'a claim upon society' (Frisby 1981: 96), whereas society (*Gesellschaft*) is understood as 'a synthesis or the general term for the totality of ... interactions' (Simmel 2004: 175). The next important equation of Simmel's conceptualisation of the mutual relationship between religion and monetary culture is the thesis on homology of society with the idea of God.

It should come as no surprise, then, that in Simmel's eyes there is at least a kernel of truth in the lament that 'money has become the God of our time'. All these terms – money, God, society – are thus inextricably intertwined. It should come as no surprise, then, that in Simmel's eyes there is at least a kernel of truth in the lament that „money has become the God of our time”. The question appears: what are the implications of this claim? The paper will reflect on the relevance of this classical conceptualisation for the ongoing debate over religious character of the prevailing global economical system.

## Valeria Motta (University of Birmingham) Socially Constructed Psychological Events. The Case of Loneliness as a Conceptual Act

Can we talk about socially constructed psychological events which are both (and at the same time) physiologically and culturally determined? In the attempt to find an answer, in this paper I provide an analysis of the Emotions as Conceptual Acts model by assessing its adequacy for explaining the phenomenon of loneliness. Psychologists have noticed that when humans are asked about the most important aspect of their life, they often describe their affiliations. According to recent studies (what is determined as a state of) loneliness is increasingly growing at a speed that could be described as an epidemic. A recent study from the Office of National Statistics shows that in Britain the greatest numbers of people have reported they "often feel lonely" compared to other European countries. At the scale at which the phenomenon is growing, it could be argued that it extends beyond being just subjects experiences and is becoming a phenomenon that describes our contemporary social reality. L. Heinrich and E. Gullone argue that loneliness is a crucial marker of social relationship deficits which is not just an adjunct phenomenon to other problems such as depression. Even when little is known about the exact mechanisms by which social bonds impact our immune system, it is commonly argued that the capacity to give and receive love and to maintain long-term social relationships are vital for human thriving, well-being, better aging and good physical and mental health. Medical studies reflect that the emotional state of loneliness is associated with several health problems and that it can also be associated with an increase in mortality. It is for this reason that the phenomenon of loneliness deserves attention in its own right. But, what exactly is loneliness? Psychiatrist Frieda Fromm-Reichmann (1959) first to raise awareness of loneliness and noted the need for a rigorous, scientific definition. In the decades that followed, loneliness as a psychological condition was characterized, and measures for quantifying individual differences were introduced (e.g., Lynch & Convey, 1979; Peplau et al., 1979; Russell, Peplau, & Cutrona, 1980; Weiss, 1973). Some authors have defined loneliness as the aversive state experienced when a discrepancy exists between the interpersonal relationships one wishes to have, and those that one perceives one currently has. But the phenomenon of loneliness is multifaceted. It has affective, cognitive and behavioral aspects. The literature is filled with rich descriptions but no theoretical framework with which to organize and explain the observed individual differences. I therefore address the need for a theoretical model and provide an analysis for the Emotions as Conceptual Acts model. This model is a psychological constructionist approach proposed by Lisa Barret Feldman and developed by several other researchers in different papers. If what it is proposed is correct, then the conceptual acts model should address the question of whether loneliness is a socially constructed psychological event. One that is both (and at the same time) physiologically and culturally determined. This theory promises to integrate the cognitive and behavioural factors of our experiences and to explain emotion granularity. The model suggests that the variety of emotional life extends from events that are called emotion to other classes of psychological events such as cognitions. The model proposes that emotion words such as 'fear' provide a name for a commonsense category that corresponds to a range of mental events.

I analyse attachment theories and empirical research on the neurobiology of affiliation for the study of loneliness (Feldman, 2016). These studies provide strong evidence for non-linguistic/pre-linguistic forms of emotion formation. Following these results, I argue that the phenomenon of loneliness is a challenge to the conceptual act model proposition that language is fundamental for the formation and function of social categories; and propose that loneliness is a case of a psychological event which is both physiologically and socially constructed even when it is a non-linguistic act.

### References:

- Averill, Stuart C. Recovery of the lost good object. *Bulletin of the Menninger Clinic*, Vol 61(3), 1997, 288-296.
- Barrett, L. F. (2005). Feelings or words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology*, 87, 266\_281.
- Barrett, L. F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science*, 1, 28\_58.
- Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, 10, 20\_46.
- Barrett, L. F., & Lindquist, K. (2008). The embodiment of emotion. In G. Semin & E. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscience approaches* (pp. 237\_ 262). New York: Cambridge University Press.
- Barrett, L. F., Lindquist, K., & Gendron, M. (2007b). Language as a context for emotion perception. *Trends in Cognitive Sciences*, 11, 327\_332.
- Barrett, L. F., Lindquist, K., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., et al. (2007a). Of mice and men: Natural kinds of emotion in the mammalian brain? *Perspectives on Psychological Science*, 2, 297\_312.
- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007c). The experience of emotion. *Annual Review of Psychology*, 58, 373\_403.
- Cacioppo, J. T., & Berntson, G. G. (2001). The affect system and racial prejudice. In J. Bargh & Clore, G. L., & Ortony, A. (2008). *Appraisal theories: How cognition shapes affect into emotion*. In M. Lewis, J.M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 628\_642). New York: Guilford Press.
- Cacioppo, S., Grippo, A., London, S., Goossens, L., Cacioppo, J. T., "Loneliness: Clinical import and interventions" *Perspectives on Psychological Science*, 2015, Vol.10(2), pp.238-249
- Campos, J. J., Mumme, D. L., Kermoian, R., & Campos, R. G Jones, W. H. (1981). Loneliness and social contact. *Journal of Social Psychology*, 113, 295-296.
- Feldman, Ruth. "The Neurobiology of Human Attachments." *Trends in Cognitive Sciences* (2016).

- Fromm-Reichmann, F. *Loneliness*, Pergamon, 1959
- Heinrich, L. M., & Gullone, E. 2006, The clinical significance of loneliness: A literature review. *Clinical psychology review*, 26(6), 695-718.
- Jones, W. H., & Hebb, L. (2003). The experience of loneliness: Objective and subjective factors. *The International Scope Review*, 5(9), 41-68.
- Jones, W. H., & Moore, T. L. (1987). Loneliness and social support. *Journal of Social Behavior and Personality*, 2(2, Pt 2), 145-156.
- Jones, W. H., Sansone, C., & Helm, B. (1983). Loneliness and interpersonal judgments. *Personality and Social Psychology Bulletin*, 9, 437-441.
- Jones, W. H. (1982). Loneliness and social behavior. In L. A. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 238-252). New York: Wiley.
- Lynch, J., Convey, W. H., Loneliness, disease, and death: Alternative approaches. *Psychosomatics*. 1979;20:702-708
- Peplau LA, Perlman D. Perspectives on loneliness. In: Peplau LA, Perlman D, editors. *Loneliness: A sourcebook of current theory, research and therapy*. New York: Wiley; 1982. pp. 1-8.
- Peplau LA, Russell D, Heim M. The experience of loneliness. In: Frieze IH, Bar-Tal D, Carroll JS, editors. *New approaches to social problems: Applications of attribution theory*. San Francisco, CA: Jossey-Bass; 1979. pp. 53-78.
- Russell D, Peplau LA, Cutrona CE. The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*. 1980;39:472-480. doi: 10.1037//0022-3514.39.3.472
- Shaver, Philip, and Cindy Hazan. "Being lonely, falling in love." *Journal of Social Behavior and Personality* 2.2 (1987): 105.
- Weiss RS, editor. *Loneliness: The experience of emotional and social isolation*. Cambridge, MA: MIT Press; 1973.

## **Albertina Oliverio (Università G. D'Annunzio Chieti-Pescara). Our Understanding of False Beliefs. An Epistemological Evaluation of Misperceptions Sources in Health Care: The Case of Childhood Vaccinations**

The explanation of the origin, diffusion, and reinforcement of false beliefs constitutes a central problem for social research. Paradoxically, progress in scientific knowledge are not always necessary followed by a reduction of misperceptions; moreover they may even contribute to consolidate these misperceptions, amplifying them or creating new ones. This involves the relationship between rationality and beliefs: is it possible to understand and rationally explain false beliefs? This article considers the case of false beliefs related to childhood vaccinations, which is one of the most serious pragmatic example of misperceptions distortive and dangerous effects on people's choices and behaviours.

Despite evidence largely proofing vaccines effectiveness and safety, hesitancy based on misperceptions toward them is now recognized to be a key and complex public health challenge (available data show for example that the Italian case is emblematic). Therefore, in order to ameliorate our understanding of false beliefs also in relation to health care, this article proposes a state of the art of the literature about the sources of misperceptions and aims to deepen the analysis and theoretical evaluation of misperceptions relation to rationality showing that the standard model of rationality referring to rational choice theory is in this case mainly useless in explanatory terms.

Three main explanatory axes are highlighted: the psychological one (cognition), the social one (social influence and virtual-interpersonal relations), the emotional one (brain and sense construction). These different explanatory perspectives present a number of elements that can be combined in helping to provide a theoretical explanatory framework for understanding vaccination hesitancy and its false beliefs source. It is moreover claimed that only by taking into account different explanatory levels recommendations to public health managers may be provided in order to more effectively correct myths and both public policy and individual choices.

### References

Boudon R., *L'art de se persuader*, Paris, Fayard, 1990.

Dubé E., MacDonald N. E. (2016) «Managing the risks of vaccine hesitancy and refusals, *The Lancet Infectious Diseases*», 16(5): 518–519.

Nyhan B., Reifler J. (2015) «Does Correcting Myths about the Flu Vaccine Work? An Experimental Evaluation of the Effects of Corrective Information». *Vaccine*, 33(3): 459–464.

Tversky A., Kahneman D. (1981) «The framing of decisions and the psychology of choice». *Science*, 211: 453–458.

Yaqub O., Castle-Clarke S., Sevdalis N., Chataway J. (2014) «Attitudes to vaccination: A critical review». *Social Sciences & Medicine*, 112: 1–11.

## Fran Osrecki (University of Osnabrück). On the uses of counter-intuitivity in the social sciences

Counter-intuitivity is often regarded as a key feature in social scientific research. However, in current social sciences, there is hardly any meaningful debate about what constitutes intuition, counter-intuitivity and their relevance for actual research. Several reasons for this lack of theorizing counter-intuitivity are presented, as well as a preliminary and dynamic model for counter-intuitive reasoning in sociology.

There is a tacit understanding that in the social sciences statements about the world should be counter-intuitive to a certain degree. A key argument for counter-intuitivity in the social sciences usually is that, because of their low level of formalization, social sciences compete with everyday concepts about the world. In order to avoid repetitiveness and redundancy, social sciences have to work with models of explanation that are deliberately at odds with and counter-intuitive to everyday notions about social phenomena (Merton 1957). Countless examples of counter-intuitive reasoning can be found in classic texts of academic sociology, a discipline that can be seen as a prototype of the modern social sciences. Karl Marx' materialist interpretation of philosophy; Max Weber's idea that modern capitalism is not built on greed, but on religious beliefs; Emile Durkheim's concept of the functionality of crime; Robert Merton's model of self-fulfilling prophecies; Pierre Bourdieu's insight that even the well educated upper classes are not self-determined in their uses of culture. And this is just a tiny fraction of how sociological perspectives provide perspectives that cut across everyday understandings of social processes.

But despite its enormous practical significance, counter-intuitivity is both a vague and sometimes even a contested concept in the social sciences in general and in sociology in particular. The concept is vague, as in sociology there currently exists no model of what actually constitutes intuition or counter-intuition and how to handle them in actual research. While in modern philosophy intuition is heavily debated, in sociology the concept is not theorized and so most sociologists (at best) have an intuitive approach to what constitutes counter-intuitive argumentation.

One reason for this lack of attention is that debates about what constitutes "good sociology" usually are very technical. It is mostly discussed how methodology is to be applied correctly or what cases call for what kind of method. But there is hardly any meaningful discussion about what, apart from coherent argumentation and compatibility with existing data, constitutes good models and theories (for some rare exemptions see e.g. Abbott 2004; Martin 2014). Another reason for why counterintuitivity has not been theorized in sociology is that meta-debates in this discipline, and they occur every now and then, are almost exclusively centered around questions of normativity. Here, it is debated to what extent sociology should emulate models of scientific objectivity borrowed from the natural sciences or rather openly take normative, i.e. political stances. These discussions come in various sizes and shapes, e.g. in the fierce debate about "public sociology" that shook American sociology about a decade ago (Burawoy 2005; Clawson et al. 2007). The entrenched nature of such disputes has in a way overshadowed discussions that do not fit into the objectivitypartisanship divide.

Now counter-intuitivity is not only a vague concept in sociology, it is sometimes implicitly opposed. Note that both in historical view and in its current state, sociology is a highly activist discipline, aiming at some kind of an extra-scientific application. This can relate to providing very technical and hands-on expertise for e.g. social policy makers or to the formulation of more abstract concepts of social critique for e.g. social movements or intellectuals. In any case, the activist approach to their discipline has committed many sociologists not to counter-intuitive reasoning, but to the production of concepts and data that are (implicitly or explicitly) designed to fit well with everyday concepts of the social world. An example for this "intuitive" strand of sociology is e.g. the very detailed collection and analysis of statistical or otherwise aggregated data on social inequality that preoccupies many, if not most, sociologists today. Here, counterintuitivity often relates only to miniscule changes in overall distributions of certain statistical factors but not to "everyday" concepts of social inequality.

This example shows that counter-intuitivity is a very demanding concept that, its desirability notwithstanding, often is at odds with the process of "normal science". Large parts of (social) scientific work cannot and need not put existing concepts upside-down, but rather draw their legitimacy from giving existing concepts a better empirical (or conceptual) foundation. Put differently, is "counter-intuitivity" a necessary prerequisite for "good social science" in general or is it an additional factor that is "nice to have"?

I will give a preliminary answer to this question by presenting a multi-level model of counter-intuitive reasoning in the social sciences. Here, I want to argue, first, that counter-intuitivity should be conceptualized as a gradual aspect in sociological reasoning where the relevant factor is not if an argument is counter-intuitive, but for how large an audience. Second, counter-intuitivity should be related to the diffusion rate of the argument it carries. Some counter-intuitive findings in sociology have successfully entered everyday concepts of the social world and have thereby lost their status as counter-intuitive. Third, I will discuss if and to what extent counter-intuitivity can be treated as a "non-epistemic" value in the social sciences.

### References:

- Abbott, Andrew. 2004. *Methods of Discovery: Heuristics for the Social Sciences*. New York: W. W. Norton & Co.
- Burawoy, Michael. 2005. "For Public Sociology." *American Sociological Review* 70 (1): 4–28.
- Clawson, Dan, Robert Zussman, Joya Misra, Naomi Gerstel, Randall Stokes, Douglas L. Anderton, and Michael Burawoy, eds. 2007.

Public Sociology: Fifteen Eminent Sociologists Debate Politics and the Profession in the Twenty-First Century. Berkeley, CA: University of California Press.

Martin, John Levi. 2014. *Thinking through Theory*. New York: W. W. Norton & Co.

Merton, Robert King. 1957. *Social Theory and Social Structure*. Glencoe, IL: Free Press.

## **Christian Piller (University of York). Interpersonal Comparisons of Utility and Well-Being: What is the Problem? What is the Solution?**

For Jeremy Bentham, the originator and advocate of utility, interpersonal comparisons of utility, which, given Bentham's conception of utility, are very closely related to interpersonal comparisons of well-being, are both necessary and puzzling. Social policy ought to be guided by considerations of utility, which led to the development of Bentham's utilitarian calculus. Without comparable utilities, social policy decisions (and not only they) would lack a rational foundation. This renders such comparisons necessary. Bentham's puzzlement about the very possibility of such comparisons is understandably less prominent. We find it, for example, in the following passage. 'Tis vain to talk of adding quantities which after the addition will continue to be as distinct as they were before; one man's happiness will never be another man's happiness: a gain to one man is no gain to another: you might as well pretend to add 20 apples to 20 pears.'

The advent of cardinal utility theory (Von Neumann and Morgenstern) has not fundamentally changed the situation. A utility function can represent a person's preferences if these preferences satisfy certain axioms. The connection between utility, thus conceived, and a person's well-being becomes more tenuous for two reasons. First, our preferences for things do not exclusively rest on how these things affect us and, secondly, these utilities are introduced on the basis of preferences between risky prospects and the notion of risk seems alien or at least secondary when we develop accounts of well-being. To allow for comparisons, measurability on an interval scale needs to be fixed to the world, i.e. we need to assume, for example, that the freezing point of water is of the same temperature whatever scale we use. Once we make this assumption, which is, of course reasonable in the temperature case, we can easily translate one scale into another. But can we assume that the same fixation is possible for well-being comparisons? Why should a lottery win, a toothache, one's death (or whatever else you chose) mark the same well-being level for everyone? Utility comparisons remain, thus, problematic.

Utility and well-being comparisons are necessary. Any ethical theory will need them. Despite their general conceptual mysteriousness, it is important to note that some such comparisons can easily be done. For example: It is better to be rich and healthy than to be poor and sick. This comparison of utility/well-being levels is correct, at least other things being equal. And so is the following comparison of utility or well-being differences: It is better to lose one's umbrella than to lose one's arm.

In this paper I offer an account that explains both the general mystery and the ease with which we make some utility and well-being comparisons. I then contrast this account with the explanations given by classic critics of such comparisons, like Jevons and Robbins. I finish by applying the lessons of my account to Broome's explanation of how to make interpersonal comparisons and, if time allows, by applying it to the extended sympathy account advocated by Harsanyi and Binmore.

In a nutshell, I argue that judgements of the form that something is good need to be explained in terms of judgements of the form that something is good for someone or something. Interpersonal comparison judgements are goodness (or betterness) judgements. It is better to lose one's umbrella than it is to lose one's arm. This judgement holds because it is better for anyone to lose his or her umbrella than it is to lose his or her arm. If goodness and betterness judgements are simply short for universal goodness-for judgements we will be able to make some such judgements whilst for many pairs of items no comparison (as universalised goodness-for judgement) can be made. On this view, some welfare comparisons are easy, whilst others remain impossible.

Classic critics of interpersonal comparisons, in contrast, appeal to a form of Cartesian scepticism about the mental states of others and their feelings of satisfaction. 'The reader will find again, that there is never, in any single instance, an attempt to compare the amount of feeling in one mind with that in another. I see no means by which such comparisons can be accomplished. The susceptibility of one mind may, for what we know, be a thousand times greater than that of another. But provided that the susceptibility was different in a like ratio in all directions, we should never be able to discover the difference. Every mind is thus inscrutable to every other mind, and no common denominator of feeling seems to be possible...' (Jevons 1871). This form of scepticism has been largely discredited in contemporary epistemology.

John Broome in his *Weighing Lives* (2004) suggests the following solution to the problem of interpersonal comparability. (1) How things are for a person I shall call her life. (2) Because a life includes everything that can affect a person's well-being, the person who lives a particular life has exactly the same well-being as any other person would have if she lived that life. (3) The goodness of a life is independent of who lives it. Thus, (4) each person's individual scale of goodness turns out to be a universal scale, measuring the goodness of lives for everyone.

In the paper, I will argue that Broome's perspective mirrors GE Moore's claim that goodness-for needs to be explained in terms of goodness. I have suggested the reverse order of explanation and I will use the plausibility of this thesis as an argument against Broome's solution.

Bentham, in his critical mood, said that different people's happiness simply create separate evaluative spheres. In the end, my view turns out to be very much the same as Bentham's.

References:

- K. Arrow. Extended sympathy and the problem of social choice. *Philosophia*, 7:233–237, 1978.
- J. Bentham. An introduction to the principles of morals and legislation. In *Utilitarianism and Other Essays*. Penguin, Harmondsworth, UK, 1987. (Introduction by A. Ryan. Essay first published 1789).
- K. Binmore. *Natural Justice*. Oxford University Press, New York, 2005.
- J. Broome. *Weighing lives*. OUP, 2006.
- J. Elster and J. Roemer. *Interpersonal Comparisons of Well-Being*. Cambridge University Press, Cambridge, 1991.
- J. Griffin. *Well-Being: Its Meaning, Measurement and Moral Importance*. Clarendon Press, Oxford, 1986.
- P. Hammond. Interpersonal comparisons of utility: Why and how they are and should be made. In J. Elster and J. Roemer, editors, *Interpersonal Comparisons of Well-Being*. Cambridge University Press, London, 1991.
- J. Harsanyi. Cardinal welfare, individualistic ethics, and the interpersonal comparison of utility. *Journal of Political Economy*, 63:309–321, 1955.
- D. Hausman. The impossibility of interpersonal utility comparisons. *Mind* 1997.
- W.S. Jevons. *The Theory of Political Economy*, Macmillan & Co, 1871.
- L. Robbins. Inter-personal comparisons of utility. *Economic Journal*, 48:635–641, 1938.

## Gianluca Pozzoni (University of Milan). Social mechanisms, scientific realism, and the metaphysics of causation

In recent decades, mechanism-based explanation has been a major focus of debate for philosophers of social science and theoretically-minded social scientists alike. A well-established notion in (the philosophy of) biology and in the (philosophy of the) cognitive sciences, the idea of a mechanism-based explanation has increasingly gained currency in many strands of research within the social sciences as well. Indeed, one may go so far as to say that in much part of social research the term “mechanism” has become near-synonymous with “explanation” altogether. This fact, of course, has been the source of much confusion about the notion of mechanism-based explanation: As some have noted, there is a lack of consensus around an unequivocal definition of a “mechanism”, especially as far as social mechanisms are concerned (Hedström and Swedberg 1998; Mahoney 2001; Hedström 2005; Hedström and Bearman 2009; Hedström and Ylikoski 2010), and the most widespread accounts of mechanism-based explanation often imply mutually incompatible views of social explanation.

Nonetheless, practitioners and advocates of the use of mechanisms for social explanation do agree in presenting the mechanism-based model as by and large opposed to the covering-law model of explanation inspired by the positivistic philosophy of logical empiricism. While the actual influence of the latter model on empirical social research can be questioned (Hedström and Bearman 2009, 22 n. 3), this paper will argue that a conflict between mechanism-based and covering-law explanation can indeed be identified in the divergent views on the nature of causation they presuppose. More specifically, the covering-law model of explanation as embodied in empirical social science – mainly correlational and variable-oriented analysis – can be shown to be tied to a reductionist and anti-realist view of causation that interprets causal relations in terms of probabilistic regularities and provides social explanations accordingly.

On the contrary, a main thrust of the mechanism-based model of explanation as presented by its proponents lies in its capacity to “open up the black box” of law-like correlations and detail the causal chain of events that lead up to the effect to be explained (see e.g. Elster 2007). As such, the very idea of a mechanism implies a realist approach that sees causation as a real-world process. Although this fact is largely acknowledged in the debate on social mechanisms and causal realism is openly defended by advocates of mechanism-based explanation (Mahoney 2001, Hedström 2005), no specific account of causation is invoked to underpin this epistemological position. In fact, some consider it as an advantage of the mechanism approach that it is not tied to any specific metaphysics of causation (Hedström and Ylikoski 2010). This paper will argue instead that important constraints on acceptable causal metaphysics need to be met if social mechanisms are to be viewed as overcoming the positivistic model of the social sciences in any meaningful sense.

For this purpose, the paper will draw on a useful distinction between “causation as a relation of dependence” and “causation as production” (Hall 2004). The former view essentially describes causation in terms of counterfactual dependence, i.e. as a relation that holds in all possible settings in which the causing event may occur. As even proponents of this view acknowledge, however, causation as dependence amounts to a sophisticated regularity view, since dependence in all possible settings can be straightforwardly reformulated in terms of regular co-occurrences of a cause and its effect in all possible settings (Mackie 1974, Beebe 2006). As such, causation as dependence has inevitable anti-realist undertones that make it unsuitable to support mechanism-based explanation on the basis of causal realism.

Productive causation, on the other hand, understands causation as a real-world process capable of genuinely generating outcomes. While this view is not qualified any further as an alternative to causation-as-dependence, some have altogether equated it with the “power-based” view of causation once prominent in metaphysical debate (e.g. Strawson 1987, Groff 2016). On this view, causation is seen as the materialization of dispositional properties by virtue of which the entities that populate the world are capable of producing effects. While this position has deep roots in the history of philosophy, it did not resurface until relatively recently in metaphysics and in the philosophy of science.

The argument above seeks to contribute to the debate on mechanisms by showing the relevance of the metaphysics of causal powers to mechanism-based explanation. Indeed, the main claim of this paper is that only on the basis of a power-based account of causation can mechanism-based explanation be defended as a more adequate paradigm with respect to the covering-law account prescribed by positivistic approaches to explanation.

### References

- Beebe, Helen. 2006. “Does Anything Hold the Universe Together?” In “Metaphysics in Science,” *Synthese* 149 (3): 509–533.
- Elster, Jon. 2007. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge, UK: Cambridge University Press.
- Groff, Ruth. 2016. “Causal Mechanisms and the Philosophy of Causation.” *Journal for the Theory of Social Behaviour* 46 (3): 1–20.
- Hall, Ned. 2004. “Two Concepts of Causation.” In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L.A. Paul, 225–276. Cambridge, MA and London: The MIT Press.
- Hedström, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge, UK: Cambridge University Press.
- Hedström, Peter, and Peter Bearman. 2009. “What Is Analytical Sociology All About?: An Introductory Essay.” In *The Oxford Handbook of Analytical Sociology*, edited by Peter Hedström and Peter Bearman, 3–24. Oxford: Oxford University Press.
- Hedström, Peter, and Richard Swedberg. 1998. “Social Mechanisms: An Introductory Essay.” In *Social Mechanisms: An Analytical*

- Approach to Social Theory, edited by Peter Hedström and Richard Swedberg, 1–31. Cambridge, UK: Cambridge University Press.
- Hedström, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36:49–67.
- Mackie, J.L. (1974) 1980. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.
- Mahoney, James. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16 (3): 575–593.
- Strawson, Galen. 1987. "Realism and Causation." *Philosophical Quarterly* 37 (148): 253–277.

## Dunja Seselja (Ruhr-University Bochum).The Epistemic Function of Agent-Based Models of Science: the case of scientific interaction

Recent inquiries into social aspects of scientific methodology, such as the division of cognitive labor, have been increasingly employing agent-based models (ABMs). A number of articles presenting results of such computerbased simulations have suggested that a high degree of information flow among scientists is not necessarily beneficial for their efficiency in acquiring knowledge (Zollman, 2007; Zollman, 2010, Grim, 2009, Grim et al., 2013, Kummerfeld and Zollman, 2015, Weisberg and Muldoon, 2009). These conclusions have been used as the basis for normative generalizations concerning actual scientific inquiry. For instance, it has been suggested that the social structure of 17th century science, characterized by an inhibited information flow, was more adequate for the efficiency of the scientific community than the social structure of nowadays science (Grim, 2009). Moreover, generalizations drawn from the results of these models have been adopted not only by other scholars investigating formal models of science, but also by philosophers of science and social epistemologists (e.g. Wray, 2011, Goldman and Blanchard, 2016).

Despite their popularity, the question which epistemic function such ABMs perform has largely remained open. On the one hand, such models have been used as explanatory of concrete historical case studies (e.g. Zollman, 2010). On the other hand, it has been suggested that they provide only how-possibly explanations, which show how under certain conditions a given phenomenon may occur (Rosenstock, O'Connor, and Bruner, 2016). Nevertheless, neither have these conditions been sufficiently elaborated, nor has it been clear which epistemic role such (how-possibly) explanations have (Riel, 2015).

The aim of this paper is to evaluate and specify the epistemic function of ABMs of scientific interaction. To this end, I will focus on three most prominent frameworks: (Zollman, 2010), (Grim, 2009) and (Weisberg and Muldoon, 2009). Due to their highly idealized nature, these models raise similar questions as a related class of models in economics and biology. In view of conceptions developed in the literature on such models (e.g. Alexandrova, 2008; Grüne-Yanoff, 2009; Reiss, 2012; Batterman and Rice, 2014; Rohwer and Rice, 2016) I argue that, on the one hand, ABMs of scientific interaction fail to be explanatory of actual scientific inquiry. On the other hand, I show in which respect they can be characterized as exploratory (Gelfert, 2016). By identifying explanatory deficiencies of such models, my analysis provides guidelines for devising ABMs of scientific interaction that are more relevant for social epistemology of science.

### References

- Alexandrova, Anna (2008). "Making models count". In: *Philosophy of Science* 75.3, pp. 383–404.
- Batterman, Robert W and Collin C Rice (2014). "Minimal model explanations". In: *Philosophy of Science* 81.3, pp. 349–376.
- Gelfert, Axel (2016). *How to do science with models: a philosophical primer*. Springer.
- Goldman, Alvin and Thomas Blanchard (2016). "Social Epistemology". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2016. Stanford University.
- Grim, Patrick (2009). "Threshold Phenomena in Epistemic Networks." In: *AAAI Fall Symposium: Complex Adaptive Systems and the Threshold Effect*, pp. 53–60.
- Grim, Patrick et al. (2013). "Scientific networks on data landscapes: question difficulty, epistemic success, and convergence". In: *Episteme* 10.04, pp. 441–464.
- Grüne-Yanoff, Till (2009). "Learning from minimal economic models". In: *Erkenntnis* 70.1, pp. 81–99.
- Kummerfeld, Erich and Kevin JS Zollman (2015). "Conservatism and the scientific state of nature". In: *The British Journal for the Philosophy of Science*, axv013.
- Reiss, Julian (2012). "The explanation paradox". In: *Journal of Economic Methodology* 19.1, pp. 43–62.
- Riel, Raphael van (2015). "The content of model-based information". In: *Synthese* 192.12, pp. 3839–3858.
- Rohwer, Yasha and Collin Rice (2016). "How are Models and Explanations Related?" In: *Erkenntnis* 81.5, pp. 1127–1148.
- Rosenstock, Sarita, Cailin O'Connor, and Justin Bruner (2016). "In Epistemic Networks, is Less Really More?" In: *Philosophy of Science*.

## David Strohmaier (University of Sheffield). Nested Groups: Membership and Parthood

Recently, the ontology of groups has received increasing attention within philosophy of the social sciences (Uzquiano 2004, Ritchie 2013, 2015, Epstein forthcoming). The question typically asked regards the reducibility of groups to other kinds of entities—i.e., sets or fusions. Here I seek to address two issues of group ontology often overlooked in the literature.

One issue is that many groups are nested (see Thomasson forthcoming). Nested social groups bear specific relations to one another such that, at least metaphorically, one group is within another. My aim is to clarify what precisely the nesting relation is and I will argue that there are in fact two nesting relations: membership and parthood.

The second issue relates to the scope of group membership—in particular, whether only individuals can be members. I demonstrate that our most plausible account of nested groups leads to an expansion of the scope of membership.

Clarifying such issues of group ontology allows us to consider more complex candidates for groups, such as states. Such cases are difficult to analyse without looking at how the group relates to other groups. How does the United Kingdom, for example, relate to the Ministry of Defence? Is this, metaphysically speaking, the same relation the UK has to NATO? And who qualifies for membership in a group like the United Kingdom anyway?

In the remainder of this abstract I sketch the two kinds of nesting, why they are necessary, and how my argument affects the scope of membership.

Here is an example: A marketing department is nested in the hypothetical corporation CORP. I call a lower-level group, like CORP's marketing department, a "sub-group", and a higher-level group, like CORP, a "super-group". In addition CORP belongs to a cartel that fixes prices. Relative to the cartel, CORP is a sub-group and relative to CORP the cartel is a super-group. Three nested levels of groups are at work here.

There are relations which the marketing department, CORP, and the cartel bear to one another which make them sub- and super-groups. In what follows, I call a relation, which makes two groups nested, the nesting relation between sub- and super-group. I propose that two different nesting relations hold, the first being mereological parthood, i.e., the marketing department is a part of CORP. The second relations is that of group membership, i.e., CORP is a member of the cartel.

The above conflicts with a popular conception of membership, which constrains membership to individuals (List & Pettit 2011, Effingham 2010: 259). In this view, the relation between sub- and super-groups is that of parthood, membership being limited to individuals. The UK would be – metaphysically speaking – a part of the NATO, rather than a member.

I argue against this view by showing that there is a difference between two types of nesting-cases. There is one class of nested groups where, necessarily, if  $x$  is a member of the sub-group,  $x$  is a member of the super-group. There exists, however, another class, where it is not the case that, necessarily, if  $x$  is a member of the sub-group,  $x$  is a member of the super-group. In other words, the classes of nested groups differ in whether sub-group members are necessarily super-group members or not. Here is an example for the two classes: A member of CORP's marketing departments, let's call him Henry, is automatically a member of CORP itself. However, a member in CORP is not automatically a members of the cartel. This difference, I suggest, is best explained by parthood being the nesting relation between the groups in one case and membership in the other. In this way, I show that limiting nesting relations to parthood is a mistake and we should broaden the scope of membership so that the UK can be a member of NATO.

To show the details of my argument let me look at the example of Henry a bit more. Because Henry is a member of CORP's marketing department, he has to be a member of CORP. There is no way around that. I suggest as an explanation that the nesting relation between marketing department and CORP ensures that every member of the sub-group is a member of the super-group. Let's call the nesting relation  $R$ -lower.

There is also a higher-level nesting relation between CORP and the cartel, which I call  $R$ -higher. Can  $R$ -lower and  $R$ -higher be the same relation? No, they cannot and for the following reason: Henry is a member of CORP, but by all appearance he is not a member of the cartel. If the cartel members were convicted of cartel activity, then Henry would not be convicted as one of them. He might have some responsibility for CORP being a cartel member, but he is not a member himself. So  $R$ -higher does not ensure that the members of the sub-group are members of the super-group. But  $R$ -lower ensured just that! It follows, that  $R$ -lower and  $R$ -higher must be different relations. Accordingly, the view that all sub-groups stand to the super-group in the relation of parthood fails. My own view that nesting relations can be parthood and membership relations remains as the better solution.

I therefore conclude that we should adapt our analysis of the membership relation to allow it to hold between groups so as to account for nested groups. Such an account of nested groups can make sense of the difference between  $R$ -lower and  $R$ -higher, in contrast to accounts that limit the nesting relation to parthood. This result is a step forward in analysing complex groups: The UK can be a member

of NATO, with all that entails.

## Philippe Verreault-Julien (Erasmus Institute for Philosophy and Economics). How possibly could economic models be 'how-possibly explanations'?

One puzzle, or even paradox (Reiss 2012), concerning highly idealized economic models is whether and under what conditions they can explain the world. Models misrepresent reality, and yet appear to be explanatory despite the fact that our best theories of scientific explanation require faithful representation (Woodward 2003; Strevens 2008). For some commentators, one way out of this conundrum is to view these models as providing 'how-possibly explanations' (HPEs) (e.g. Grüne-Yanoff 2013b; Grüne-Yanoff 2013c; Rohwer and Rice 2013; Ylikoski and Aydinonat 2014). They contrast HPEs to 'how-actually explanations' (HAEs), the latter being empirically confirmed whereas the former are not, or at least to a lesser degree. However, there is a disagreement on how viewing economic models as HPEs actually solves the paradox. Indeed, whereas someone like Grüne-Yanoff holds that HPEs simply serve a different explanatory purpose and do not faithfully represent the world, Ylikoski and Aydinonat claim that HPEs afford, like HAEs, understanding of the world by representing faithfully certain relations of dependence. The former account solves the paradox by denying that economic models explain in the usual sense while the latter does it by showing how HPEs achieve faithful representation. Regarding economic models as HPEs therefore does not straightforwardly provide a solution paradox. Rather, it raises an important question about the nature of HPEs, namely what demarcates them from HAEs? Despite the current taxonomical oversupply, there are basically two types of distinction between HPEs and HAEs on offer in the literature. While some (e.g. Dray 1968; Forber 2010) consider HPEs and HAEs to simply be different types of explanations, others (e.g. Hempel 1965; Brandon 1990; Bokulich 2014) distinguish them by their degree of empirical support. These distinctions, I contend, do not carve explanations at their joints.

My paper purports to provide an account of HPEs that clarifies their nature in the context of solving the puzzle of model-based explanation. More precisely, I argue, following how HPEs and HAEs are called suggest, that the modal notions of 'actuality' and 'possibility' provide the relevant dividing lines between HPEs and HAEs. The crucial feature that distinguishes HAEs from HPEs is neither the type of question they answer nor the empirical truth of the former and the falsehood of the latter, but instead the sort of knowledge they provide. Whereas HAEs provide knowledge of actual explanantia and explananda, HPEs provide knowledge of possible explanantia or explananda. Having a HAE implies having knowledge about an actual state of affairs whereas having a HPE implies having knowledge of a possible one under a given modality. For instance, widely discussed (see e.g. Sugden 2000; Sugden 2009; Aydinonat 2007; Grüne-Yanoff 2013b) models such as Banerjee's (1992) model of herb behaviour or Schelling's (1971; 1978) model of residential segregation do not provide knowledge about the actual causes of their targeted phenomenon. Rather, they provide knowledge of a possible mechanism. Models justify the beliefs we have concerning certain possibility claims. This suggests that HPEs are indeed a species of explanation in that they justify beliefs we have about possibility claims. Insofar as HPEs are 'explanatory', they are only explanatory with respect to possibility, not actuality.

This characterization sheds light on the contrasting views about HPEs in two different ways. Firstly, it implies that it is in a sense right to deny that HPEs faithfully represent (e.g. Grüne-Yanoff 2013a). Indeed, HPEs do not faithfully represent actual explananda and explanantia, only possibilities. It is thus correct and valuable to deny they represent in that sense. However, it is also right to consider that they faithfully represent (e.g. Ylikoski and Aydinonat 2014), as long as the domain is restricted to possibilities. This suggests that unqualified claims of (un)faithful representation can be misleading and that we better be clear about what exactly is represented. Secondly, it allows to clarify the exact epistemic contribution of models. Some models provide knowledge of actuality, others of (im)possibility about a given explanandum or explanans. This implies that models that do not offer HAEs are not necessarily epistemically vacuous, as is sometimes argued (Hausman 1992; Alexandrova 2008). Such models can be HPEs and thus contribute to our knowledge, albeit differently. It also implies that depending on the knowledge provided by a given HPE, it may, for instance, contribute to learning (Grüne-Yanoff 2009; Grüne-Yanoff 2013a) or understanding (Ylikoski and Aydinonat 2014).

My paper hence makes two main contributions. Firstly, it contributes to the literature on economic models by providing a framework with which we can appraise both our philosophical views about economic models and the contribution of models themselves. One important puzzle about them is whether or not they 'explain' and a precise account of HPEs is one step in the direction of solving this puzzle. Secondly, it contributes more generally to the literature on HPEs by proposing a novel account that does not rely on the usual distinctions as the criteria of demarcation between HPEs and HAEs. My proposal is instead to use modality--i.e. actuality and possibility--to distinguish them.

## **Juliette de Wit (University of Groningen) and Chiara Lisciandra (University of Groningen). Measuring social norms in economics. A philosophy of science perspective.**

Economists are increasingly interested in social constructs. Building upon the results from other social sciences, economists have examined social constructs both theoretically and empirically (Gershman, 2016), applying the methodological toolkit of the economist. For instance, concepts such as well-being, trust, and culture are now measured and incorporated into economic theory (Knack & Keefer, 1997; Rabin, 1993). The richness of social survey data has enabled and facilitated these developments. Philosophers have started investigating these developments, asking questions regarding the ontological status of social constructs (Guala, 2007), the objectivity of value-laden social constructs (Alexandrova, 2015; 2016), and the measurement of fuzzy social constructs (Cartwright & Runhardt, 2014).

Measuring social constructs is difficult, for four reasons: First, because there are no strict laws governing their behaviour; second, because they are value-laden; third, because they cannot be directly observed or exactly measured; and fourth, because they are fuzzy (Chang & Cartwright, 2013). In spite of these difficulties Alexandrova (2016), amongst others, argues that social constructs should have a place in science. They can be used in an objective way in so-called mixed claims: propositions that combine the normative with the empirical. That is on the condition that we remain attentive to value presuppositions and check whether “the empirical claim is robust to disagreement about [underlying] values” (Alexandrova, 2016, p.21). Chang and Cartwright (2013) argue in favour of a theory of measurement and a representation theorem to overcome the difficulties that are encountered when measuring social constructs. The former reflects the act of defining a construct, and the latter the creation and support of the empirical procedures that are developed to represent the construct. Issues of validity are involved in the latter.

In spite of the increasing presence of social constructs in economics (see for example the work of Hofstede, Hofstede, & Minkov (2010) or Knack & Keefer (1997)), the way they are conceptualized, operationalized and measured is not widely addressed. Reiss shows that empirical researchers make mistakes when interpreting these constructs (2008). When it comes to social norms, a construct that increasingly appears in macroeconomic models, there is a lot of ground to be gained.

Social norms, defined as behavioural rules that arise through human interaction (Bicchieri, 2006; Froncek, 2015), affect economic decision making. Their impact is not only observed at the microeconomic level, where individual decisions are made, but also at the meso- or macroeconomic level. Within microeconomic models, social norms function either as another constraint to the utility function or, alternatively, they can be included in the utility function (Hausman, 2008). The methods that are used in this strand of the literature are relatively well developed; one could even speak of a methodological consensus.

In contrast, we observe that such a consensus is missing for the majority of studies that assess social norms at the macroeconomic level. These are studies that, for example, examine the impact of social norms on institutional quality across countries (Beine & Sekkat, 2013). Here, the methods that are used to measure (types of) social norms differ considerably. They can be described as ad hoc, missing conceptual depth, rigour and operational validity. This is problematic for several reasons. Julian Reiss notes for example that “it is a fallacy to think different methods measure the same concept just because we attach the same name to them” (2008, p.71). Moreover, incommensurability of studies becomes an issue once the way in which social norms are conceptualized and operationalized varies to a great extent. Drawing on these observations, we argue that it is a theory of measurement and a representation theorem that is currently missing in the literature. For this reason, scholars make mistakes when social norms are incorporated into empirical, macroeconomic studies. In this paper, we therefore develop a new, macroeconomic measure of social norms.

We depart from Christina Bicchieri’s theory of social norms, as proposed in her book *The grammar of society* (2006). Here, Bicchieri develops an operational definition that identifies three components of social norms: behavioural, empirical and normative. The behavioural component reflects the behaviour an individual displays when conforming to a norm, the empirical component reflects expectations of that individual regarding others’ norm conforming behaviour. In addition to these expectations, the individual has beliefs regarding the expectations that others may have regarding his or her norm conforming behaviour. These beliefs are captured by the normative component of social norms. While Bicchieri’s defines norms at the micro-level, analysing how individual interactions are shaped by social norms, her definition can guide our thinking at supra-individual levels. The fact that social norms are emergent phenomena, arising from interactions at the individual level at the regional or national level (Sawyer, 2005), facilitates this transposition.

Using social survey data at the regional and national level (provided by the World Values Survey and the European Values Study), we develop a cross-country measure of social norms that follows Bicchieri’s definition. The richness of the data enables us to operationalize the fuzziness of social norms. Moreover, as different themes are addressed in these surveys, different types of norms can be approximated. For example, a norm of honesty or work ethic. In this exercise, specific attention is given to values that may underlie this process, following Alexandrova’s proposition to make value presuppositions explicit (2016). Several tests for its robustness are performed, showing that our measure performs quite well across countries and improves upon current measures encountered in the economic literature. In this way, we conclude that while the measurement of social constructs is difficult, especially in comparison to measurement in natural sciences, it is an undertaking that is worth our efforts.

## Jack Wright (University of Cambridge). If objectivity is plural how should it be used in economics?

When economists offer advice, write newspaper columns, or appear in the media, it is normally implied that what they say is valuable and more than the perspective of one citizen. This is often justified by the assumption, explicit or otherwise, that what they say is objective. Whether or not economics is or economists are objective lies at the centre of many debates within and about economics. When Raj Chetty defended the scientific status of economics in 2013, for example, he concluded the emergence of certain empirical methods (randomised control trials, big data analysis, and natural experiments) means that economics is about «evidence instead of ideology».

The issue is that objectivity can mean different things. Following a strict method, having a dispassionate disposition, being impartial, or having special access to the 'objective' world can all lead to declarations of objectivity. Sometimes people are taken to be objective, sometimes its experimental processes, and other times its statements or states of the world. All these things give rise to different meanings for objectivity; meanings that can come into conflict—methods might clash, methods do not necessarily imply impartiality, impartiality and dispassionate dispositions do not necessarily imply each other, etc.—and that are often mixed together. This is important because declarations of objectivity carry weight. As the Bank of England's internal document *Our Code* (2016) recognises:

"Impartiality and objectivity are crucial to the decisive behaviours that form a core part of our Bank values. We know that our reputation for impartiality and independence is vital to our effectiveness, and, if lost, would be hard to recover. Like other colleagues in the public sector, we know we must be seen to be apolitical and must never allow ourselves to become open to the perception that our decisions have been inappropriately influenced." (p. 11)

Being (or being seen to be) objective, impartial, and apolitical is important for many institutions and actors to have the influence they do. But if what is meant by such declarations is unclear then there is room for confusion and manipulation. It is worth asking, therefore, what role objectivity can and should have in economics?

In this paper I will make two claims; that the term objectivity is useful even if it is plural, and that a plurality of meanings for objectivity should imply a contextual sensitivity of methods for economics. To do that I will first outline some different possible meanings of objectivity. I will give a few reasons not to assume that this plurality is an illusion, but rather than arguing in detail (as Douglas (2004), Fine (1998), Megill (1991), and others have done), I will focus on the implications of that plurality. This will lead me to my first claim; I will argue that the plurality of objectivity does not imply that the noun 'objectivity' should be eliminated (as Hacking (2015) argues). The lack of fixed meaning for a term does not imply it is useless, and there are reasons to believe the general concept 'objectivity' is useful. This is pragmatic argument to counter Hacking's opposite pragmatic argument. I will then discuss what this means for economics. I will argue that if objectivity is a plural term, it is so because it is context sensitive. This implies that economists should be clear about what they mean when they invoke 'objectivity' and the context in which they mean it, and that the ways that economics constrain subjectivity and abstract from particular perspectives—i.e. The methodologies they follow—should also be sensitive to contextual factors. This results in a picture of economics in which methods are chosen relative to their context of application.