Plenary 1, 11.00 Saturday 15th July: Shaun Nichols

Scepticism and the Acquisition of "knowledge"

The philosophically important concepts present in common-sense thought must be acquired in some fashion. Recent advances in statistical learning theory provide resources for providing candidate explanations for the acquisition of certain such concepts. Substantive theories of acquisition will include (i) a thesis (T) about the concept (typically concerning the content of the concept) (ii) a specification of the learner's hypothesis space (H), (iii) an empirical assessment of the evidence (D) available to the learner, and (iv) an explication of how statistical principles would make it appropriate for a learner with H and D to infer T. We apply this method to the acquisition of the concept of knowledge. Do you know you're not being massively deceived by an evil demon? That's a familiar sceptical challenge. But how do you have a conception of knowledge on which the evil demon constitutes a prima facie challenge? Why do people respond so quickly to outlandish sceptical scenarios involving sorcerers and mad scientists? Recently several philosophers have suggested that our responses to sceptical scenarios can be explained in terms of heuristics and biases. Drawing instead on learning theory, we argue that, given the evidence available to the learner, it would be rational for the learner to infer an infallibilist conception of knowledge.

Plenary 2, 17.00 Saturday 15th July: Eugen Fischer, Paul Engelhardt, and Aurélie Herbelot

Eyes as Windows to Minds: Psycholinguistics for Philosophy

Language shapes inferences. This talk discusses how methods from psycholinguistics and computational linguistics can be used to study automatic inferences that routinely go on in language comprehension – e.g., when philosophers read case-descriptions for thought experiments, or premises of arguments. Our research examines under what conditions the key process of stereotypical enrichment (Levinson 2000) leads to contextually inappropriate inferences, and how such inferences lead to unsound intuitions, and fallacies in philosophical arguments. We present a study on a classical paradox about perception, the 'argument from hallucination'. We used computational methods (distributional semantics analysis) to document salience differences between senses of 'to see'. We then combined pupillometry with plausibility ratings to examine the hypothesis that extensive salience differences promote inappropriate inferences from less salient uses of the verb. This paradigm can be used more generally to study whether automatic inferences from words are made in language comprehension and go on to affect further thought.

Plenary 3, 15.45 Sunday 16th July: Arianna Betti

History of Philosophy in Ones and Zeros

Experimental philosophers are increasingly interested in complementing experimental methods with empirical methods from the digital humanities. These crucially include computational tools such as text mining and data mining techniques. These methods are attractive, because valuable contributions can be obtained by applying even rather simple, well-known computational techniques to texts relevant to the work of philosophers (van Wierst et al. 2016). This talk substantiates the point by presenting a quantitative, computational analysis of philosophical texts in addressing two open questions in the history of 19th and 20th analytic philosophy, by relying on corpora of writings by Bernard Bolzano (1781-1848) and Willard van Orman Quine (1908-2000). I highlight practical and methodological issues arising from the use of these new methods, and formulate recommendations and plans for future work in this field.

Can the "Knobe Effect" be merely a cognitive illusion? - Katarzyna Kuś and Bartosz Maćkiewicz

Experimental philosophers aim at developing experimentally testable psychological explanations for the actual patterns of people's intuitions. Fischer and his colleagues (Fisher 2014, Fischer, Engelhardt, Herbelot 2015, Fischer, Engelhardt 2016) propose to look for the so called GRECI explanations that trace such intuitions back to generally reliable cognitive processes, especially psycholinguistic ones, that under specific circumstances predictably produce cognitive illusions.

At some very fundamental level the cognitive processes that are responsible for our intuitions are assumed to be universal and culture independent, hence many attempts to replicate experiments across different languages and cultures. One of the primary assumptions of this endeavour is that similar outcomes of the experiments follow from similar psychological causes. The Knobe Effect is one of the most successfully replicated experiments (e.g. Knobe, Burra 2006; Dalbauer, Hergovich 2013, Kuś, Maćkiewicz 2016) and it is often assumed that the psychological mechanisms responsible for the intuitions evoked in Knobe's scenarios are homogenous across all cultures and languages where the experiment reveals significant differences between two experimental conditions.

Our aim is to challenge this assumption using linguistic data. Firstly, linguistic corpora, a resource often overlooked by experimental philosophers, can be used to establish that Polish adverbs used in the replication of the original experiments are semantically and pragmatically different from their English counterparts. The data shows that there is a huge gap in expected frequencies of bigrams with an intentionality-indicating adverb when it comes to verbs depicting positive actions, whereas in English there is no such phenomenon. Moreover, various ways of measuring the strength of collocations indicate that intentionality-indicating verbs are more likely to occur with verbs depicting negative actions. In contrast, in English such a relation is not that clear, especially in case of other adverbs than "intentionally".

As far as the experimental data goes, we adopted Fisher's and colleagues' framework and formulated GRECI explanation of the Knobe effect which can be experimentally tested. Our hypothesis was that GRECI explanation will be language-dependant. In Polish the size of the effect observed in variations of the Knobe experiment employing different adverbs can be directly related to the strength and non-cancellability of appropriate stereotypes involved in GRECI explanation.

Relation between Care and Loyalty Foundation and Its Implication for the Nature of Moral Intuition - Hyo Eun Kim

This study aims to identify various patterns of the interaction between care and loyalty across 4 countries. Moral Foundation Theory (MFT) suggested that several innate psychological systems are the foundations of moral intuitions. Further it explained variations of moral intuitions across cultures. Our study begins with the idea that the nature of the moral judgments could be further investigated by focusing on the relation between Care and Loyalty moral foundation which has not yet been investigated. We assessed the effect of care and loyalty factors across four countries by testing folk judgments of moral wrongness of an agent's behaviour to a victim.

We modified Haidt and Graham's Moral Foundations Questionnaire (2011) for cross-cultural consistency, and use the modified scenarios to investigate folk intuitions. Altogether, twelve sets of scenarios were administered diverse group of participants (N=840 per country) under 2 by 2 by 2 within- subject design. Our idea begins with the fact regarding care morality that intuitions with regard to emotional vs. physical pain are divergent. There is a general recognition among contemporary philosophy and neuroscience about the moral implication of emotion and pain; the fact that a close relation between emotional pain and moral judgments serves to distinguish moral

judgments that provide a genuine moral reason for genuine moral reason and those judgments should be disregarded. Based on the relation between emotion and pain, and its relevance to moral judgments, our study begins with the specific question as to how and why people's moral judgments about causing emotional harm or pain differ from their moral judgments about causing physical harm or pain, and how their moral judgments correlate with loyalty factors.

We also look for cross-cultural differences in patterns of moral intuitions between U.S., Korea, Romania, and China. A series of factor analysis showed diverse patterns of moral judgments across culture. The effect of the loyalty factor was found to be significant across nations. The clear distinction emerged between the U.S., scoring high in correlation between judgments of moral wrongness and IOS scale, and Korea, showing no effect of psychological closeness to the victim or the agent in the scenarios. Two follow-up experiments in Romania and China indicate that the diverse moral judgment patterns is not due to merely the East-West, or Collectivism vs. Individualism dichotomy and may be linked to other factors besides the frame.

Kantian Parents and their Utilitarian Children - **Robin Kopecky and Michaela Košová**

In this paper, we are presenting the results of our research, which explored moral judgements in children presented with certain variants of trolley problem and the factors that are predicted by behavioural economy games. The method chosen for the research was a direct interview with the respondents. The study was motivated by previous studies, which suggested that the sex of the respondent is a relevant factor in utilitarian judgments.

Two studies, which preceded the one in our focus, were realised solely by means of a questionnaire exploring moral judgement in sacrificial dilemmas. The first pilot study was conducted by means of a paper questionnaire on 231 biology and anthropology undergraduates. In the second study, we used a web-based questionnaire (N=12 335, mean age 31,19; SD=10,61) on the Czech general public. We found that in comparison to women, men were more willing to sacrifice a person in order to save five in both the switch and the bridge scenarios.

Sex difference in trolley scenarios that is present even among professional philosophers (Bourget, Chalmers 2014) lead us to conduct a new field research in four major Czech cities where we tested children's judgements (5 – 15 years old, N= 217, and their 63 parents) in variants of trolley problem by means of a direct interview conducted by an experimenter. Our hypothesis was that there is a similar sex difference also among children. In the first experimental site we conducted parallel interviews with parents and their children so that we were able to test whether there is any difference within particular families. We also included set of experimental games from behavioural economy in order to link moral judgement with external factors. The following games were used: In "the dictator game" participants can allocate their candies to an envelope addressed to orphans, by means of which we measured concern for the greater good. Risk aversion was measured by willingness to roll a dice with favourable risk. We also measured participants' aversion to effort and their delayed gratification.

We have found that children's claimed willingness to act in the switch and the bridge case and their moral evaluation of an agent who switched/pushed are almost identical. We conclude that children do understand the difference between personal and impersonal scenarios – substantial difference lies in the question: „Would you mind playing and being friends with a child who switched/pushed?" Even though children mostly claim to prefer to push and describe the pushing agent as "good", they are less willing to play with them than with an agent who switched. Thus, it seems that children perceive the intention to kill one in order to save five as problematic only in social but not moral domain.

Our conclusions: First: there is no difference between boys and girls in judging the variants of trolley problems although there is a strong sex difference between their parents, just as among students in the first study and general public in the second study. Second: children who are less afraid to take favourable risk are less worried about being friend with a child who pushed a man from a bridge (k =-0,222, p<0,05). Third: Children who show greater generosity in the dictator game are less utilitarian in the switch scenario (k =-0,207, p<0,05), while they show no significant difference in the bridge scenario. Fourth: Children are more utilitarian than their parents in both cases. We found no effect of family upbringing since it wasn't confirmed that exclusively parents prone to utilitarian judgements have similarly utilitarian children. Parents show a range between

utilitarian and deontological decisions, leaning towards deontology in the bridge case, while their children lean towards utilitarian decisions regardless of the case.

Detach the body, keep the pain - Kevin Reuter, Michael Sinhold and Justin Sytsma

It has been argued that our concept of pain seems to be contradictory (Aydede 2006; Hill 2006). On the one hand, it both feels to us, and we also speak as if pains are states of the body. On the other hand, pains are said to be mental because they are subjective states, i.e., they cannot exist without being felt. A series of experiments (Sytsma 2011, Reuter & Sytsma (manuscript)) has revealed that between two-thirds and 90% of the tested participants disagree with the subjectivity hypothesis; e.g. they accept the existence of pains even if painkillers prevent them from feeling these pains. These results point towards the truth of the pain as bodily state - hypothesis. However, an alternative interpretation states that pains are necessarily mental but not necessarily conscious all the time (see e.g. Lycan 1995) – call this the unconscious mental pain hypothesis.

In order to investigate these questions, we designed a vignette to disentangle these competing interpretations. Importantly, conducting the experiment would also allow us to specify more precisely the bodily conception of pain. Thus, 745 participants were instructed to imagine the year 2072, in which body parts and especially hands, can be replaced by reproduced body parts that consist of human tissue indistinguishable from ordinary hands. A man named Joshua complains about of sharp and persistent pain in his hand to his doctor. First, the doctor examines the detachable hand while it is still attached to Joshua. Second, Joshua detaches his hand, i.e., it is not connected to any person. Third, the doctor detaches her own hand and replaces it with Joshua's hand. Consequently, she shows typical pain behaviour, such as "Ouch!" Fourth, the doctor detaches Joshua's hand and puts it on the shelf. At each of the four steps, the participants were asked whether they believe that there is a pain in Joshua's damaged hand.

The results lead us to draw three conclusions that we will discuss in this paper: First, since the statement that pain is in the detached hand is affirmed by about as many participants as those who reject the statement, we can identify at least two different conception of pain. Second, the results strongly suggest that at least 45% of the respondents hold a bodily conception of pain, according to which pain refers to bodily damage itself. Note, that the responses cannot be squared with the unconscious mental pain hypothesis, given that the connection between the body part and its mental representation has been severed. Third, it is still an open question how to adequately interpret the responses from those participants who rejected to locate pains in detached body parts. Following interpretations are possible: (a) pains are thought of as representational states, i.e., brain states that merely represent disorders in the body; (b) pains are conceived of as bodily states but necessarily connected to a conscious mind; (c) the contradiction between the bodily and mental aspects of our concept of pain is merely apparent, conditional on a faulty Cartesian conception of the mind (Hyman 2003): Shifting focus from the criterion of localisability to distinguish mental from bodily aspects to the criterion of subjectivity, see e.g., Merleau-Ponty (1945), a pain can be clearly localised in a body part (as described by Hyman 2003) but nevertheless conceived of as mental.

Oops, I did it again – the effect of morality and typicality on causal judgement - Lara-Christina Kirfel

Current theories of causal judgement aim to incorporate the effect of different kinds of norm on judgments of actual causation. It has been shown that if an agent violates a moral norm, or does something untypical, they are seen as more causal for a corresponding outcome of the action than if they had done something moral or typical (Knobe & Fraser, 2008, Kahnemann & Miller, 1986). It has therefore been suggested that our causal judgments are sensitive to norms, e.g. moral or statistical norms, with norm-violating actions or features being judged as more causal than norm-conforming ones (Halpern & Hitchcock 2015, Icard & Knobe, 2016). The current paper investigates whether there is a unique effect of norms on causal judgments by examining the effect of the typicality of an action and its interplay with morality on causal attributions. Furthermore, we develop a novel way of testing typicality by using dynamic scenarios that allow a natural representation of typical actions and the flexible introduction of moral norms.

Experiment 1 investigated the influence of typicality and morality on causal judgments in a conjunctive causal structure, i.e. two actions are needed in order for the outcome to occur. The scenario we used presents a sequence of six days and successively gives information about which agent acted that day, hence allowing for a natural representation of typical vs. untypical actions. Furthermore, we varied whether moral norm in the story was either introduced right at the beginning of the scenario, or after five days, i.e. just before the final outcome. Our experiment confirmed the effect of moral norm violations on causal attribution, and showed that a moral norm violation is assigned even greater causality when it is performed typically. We also found a reversed norm influence for agent- level typical behaviour, with typical behaviour being judged more causal than untypical behaviour. The time point of the introduction of the moral norm did not make a difference to causal judgements.

Experiment 2 found an increase in causal attribution to typical behaviour for both an accumulative causal structure, i.e. where repeated behaviour gradually increases the causal contribution, and a conjunctive causal structure, where the frequency of an individual action does not make a difference to the outcome. The increased causal attribution to typical behaviour was furthermore independent of whether the agents knew about each other's behaviour.

The fact that people assign increased causality to a typical, repeated action goes against predictions of norm theories and shows the need for a new account of typicality in causal attribution. We argue that a probabilistic extension of current counterfactual theories is needed in order to capture the influence of typicality. In addition to counterfactually testing whether the undoing of an action makes a difference, we also need to test whether the variation of the typicality of the action would make difference to the likelihood of the outcome to occur. The influence of typicality of an action proves to be a crucial factor in the attribution of causality, and calls for new theoretical frameworks of causation.

Culturally sensitive Virtue Ethics: Creating a Set of Virtues and Vices for a Colombian sample - Sergio Barbosa de la Torre

Since its dawn, one of Experimental Philosophy's most discussed topics are lay conceptions of Ethics and Morality (Doris & Stich, 2005; Knobe, 2010; Sinnott-Armstrong, 2005). Research in experimental philosophy has increasingly been informed by psychological methodologies and results on lay conceptions of morality, harm and so on (Kahane, Everett, Earp, Farias, & Savulescu, 2015; Mikhail, 2007; Shenhav & Greene, 2010; Tse, 2008). These results have usually been interpreted as representative of a universal lay conception of morality. However, as cognitive psychologists have only recently admitted (Henrich, Heine, & Norenzayan, 2010), lay conceptions and cognition in general may be highly variable according to cultural and contextual variables. This study's main aim is to provide set of virtues and vices as well as behaviours that lay perceivers use as cues to ascribe these traits in a Colombian sample. We believe these could be used in future empirical studies on the lay conception of morality and character (Pizarro & Tannenbaum, 2011; Uhlmann, Pizarro, & Diermeier, 2015) and as basis for creation of similar materials in other cultures. A secondary aim is to use qualitative data analysis to draw hypothesis about the lay conception of Unity or Disunity of Virtue (Brickhouse & Smith, 1997; Sreenivasan, 2009; Toner, 2014; A. D. M. Walker, 1993; Wolf, 2007).

Based on previous work (Cawley, Martin, & Johnson, 2000; Chadwick, Bromgard, Bromgard, & Trafimow, 2006; L. J. Walker & Hennig, 2004) we proposed a six phase study to determine which personality traits are considered to be virtues, vices and non-moral traits and which behaviours are associated to these traits in a Colombian student sample. First, we asked participants to recall any number of adjectives describing an agent's personality. After filtering this list, we asked a different sample to classify which of these adjectives are morally desirable, morally undesirable or have nothing to do with morality. Based on this phase we selected the most consistently classified adjectives as the lay Virtues, Vices and Amoral traits. Second, we used a similar procedure to elicit a number of behaviours that they believe are accurate cues to ascribe these traits. These behaviours were filtered out and then confronted in a classification task by an independent sample.

Based on our study we offer a list of Virtues, Vices and Amoral traits coupled with a list of behaviours that allows to quantitatively define how linked each behaviour is to each trait. We believe our study is a suitable basis to create culturally sensitive materials for research in moral psychology and experimental philosophy. In addition, using content analysis on the elicited

behaviours allowed us to draw certain hypothesis for future studies on the lay conception of Unity or Disunity of Virtue.

Best, Second Best and Good Enough Explanations: How they Matter to Reasoning - Patricia Mirabile

When trying to find an explanation for a given phenomenon, we are often faced with multiple competing explanations. What is, then, a rational choice rule that can be used to answer the question of which explanation, if any, we are to accept as being true? One proposal, dubbed IBE - for Inference to the Best Explanation - states that differences in explanation quality among competing explanations should be taken into account: we should infer to the best explanation, provided that explanation is good enough and provided that it is significantly better than its competitors.

Is there, however, a correlation to be found between explanatory quality and people's willingness to accept an explanation? And is the quality of the second best explanation also taken into consideration in reasoning? Three experiments were conducted in order to investigate whether IBE provides an adequate description of the way people reason about explanations. All participants were presented with six scenarios, involving different types of explanatory connections (physical, intentional and functional), and care was taken to develop content-rich material, evocative of situations that might present themselves to everyday reasoning.

The first experiment shows that the absolute quality of an explanation, when it is presented by itself, has a significant effect on whether people are ready to accept it as true. The second and third experiments show that the quality of the second best explanation also has an effect on the participants' dispositions to accept the best explanation as being true: namely, a strong competitor is likely to lead to a significant decrease in the acceptance rate of the best explanation, while a weaker competitor is much less likely to do so. The third experiment differs from the second one in two respects: Firstly, it uses a within-subject design to investigate how explanation quality influences the acceptance of explanations at the individual level; and secondly, it also asks participants for judgements on their metacognitive confidence, opening the way for further inquiries that could benefit from the use of methods borrowed from cognitive psychology.

SUNDAY 16th JULY:

Surveys and Heuristics - Daniel Dohrn

There is substantial disagreement on the experimental results in Gettier cases. While some surveys show great variations among different groups, these variations are not reproduced in others (Weinberg et al. 2001, Nagel et al. 2013). If there are variations, they call for an explanation. Empirically credible candidates for such an explanation depict the judgement as the outcome of heuristic reasoning. I shall consider several proposals in the literature what the heuristics in play looks like and then present my own view.

I start with Nagel's (2012) view that the Gettier judgement is the output of mindreading heuristics, and that the heuristics conforms to Koriat's self-consistency model (SCM). According to SCM, there is a common heuristic mechanism of sampling from a vast pool of potentially relevant information. An intuition will be the stronger the more the sampled information points in one direction. If there is sufficient overlap in the individual pools of information, intuitions are the more likely to intersubjectively converge the stronger they are. SCM has difficulties explaining the divergence of judgements, though. The only explanation is that the pools of potentially relevant information vary among subjects. But this seems implausible for an everyday category like knowledge.

Spicer (2007) posits a mental module for folk epistemology. The module provides a fast and frugal heuristic of knowledge ascriptions. Spicer also considers the possibility that several mechanisms may be in play on different occasions, including more sophisticated ones. This view has a certain potential to explain the divergence of judgements if we assume that variations in judgements are due to use of different heuristics. However, the question becomes why different heuristics should be in play.

Coming to my own proposal, I account for the Gettier judgement in terms of classificatory heuristics as researched in the heuristics and biases program. Justification and knowledge are the categories of interest. One considers whether the case as far as it is described sufficiently resembles representative items of justification / lack of justification and knowledge / lack of knowledge, respectively. However, similarity is not overall but weighted similarity.

From this proposal I derive the following explanation of inter-group variation of judgement: in forming the heuristics, representative items are tentatively grouped together such as to form reference classes associated with a category. The classes may vary due to differences in contextual salience over different groups. Moreover, contextual salience may lead to variations in how respects of similarity are weighted. These divergences are opposed to tendencies of unification arising from linguistic coordination and (folk) epistemology.

What can moral philosophy learn from population surveys? - Hugo Viciana

Within the growing interest in ethics for empirical and experimental methodology, research on representative surveys provides an opportunity as well as a largely untapped tool. If the empirical and experimental turn is neither strictly a novelty nor always philosophically relevant, it is nevertheless easy to show that the methodological aggiornamento has come to stay. By distinguishing between convenience sampling and probability sampling, in this presentation, I will show briefly what is to be understood by representative surveys and how they can overcome some methodological shortcomings of other preferred methods in experimental philosophy. One of these shortcomings is the putative "WEIRDness" of some findings, the fact that they might have only been tested among participants living in WEIRD settings: Western, Educated, Industrialised, Rich, Democratic environments. Firstly, I will respond to the objection "from within" consisting in questioning that a population survey methodology can offer something interesting about the moral intuitions of individuals. Faced with the objections of instrumental uselessness, I will show that the internal validity, as well as the external validity of the measures of moral attitudes and intuitions in population surveys can be satisfyingly controlled. Population-based survey experiments are a particular case in point.

Secondly, I will respond to the objection "from the outside" consisting of questioning the substantive relevance of this type of empirical input on moral philosophy generally. In order to achieve this latter point I will review some of the most remarkable results in relation to moral values and meta-ethical attitudes and their cultural evolution. By paying attention to some cases

of social conditions that give rise to different folk ethical configurations, I will show how this source of data actually opens up new areas of normative inquiry. In sum, expanding the inductive basis of moral philosophy through a more frequent use of population surveys seems but one more step in the empirical and experimental turn.

Intuitive Expertise and Irrelevant Options - Joachim Horvath

In the so-called *push dilemma*, an out-of-control speed-train is about to run over five people and can only be stopped by pushing a heavy person onto the tracks (henceforth: push option). Most lay people and moral philosophers consider it morally wrong to kill the heavy person and prefer not to intervene (henceforth: doing nothing) in this classical two-option scenario. Unger (1992, 1996), however, argued that adding irrelevant options to the push dilemma would overturn this intuition.

In this paper, we test Unger's claim empirically by presenting participants with the classical two-option scenario (henceforth: Two Options) and an extended six-option scenario (henceforth: Six Options) that included the two options of the classical scenario, as well as four "intermediate" options (in terms of number of people saved and aversiveness of the action). Both scenarios were presented to participants in randomised order. Participants included 145 lay people as well as 134 professional moral philosophers. Including philosophical experts allowed us to investigate the so-called expertise defence against the challenge that empirical results from experimental philosophy pose for the trustworthiness of philosophical intuitions.

According to the expertise defence, the intuitions of well-trained philosophical experts are superior to the intuitions of lay people. Overall, we found that the additional options indeed increased the ratings for the push option, and that order of presentation strongly influenced subjects' choices. Moreover, we found that the intuitions of expert ethicists were no less susceptible to these two factors than lay people's intuitions. First, we found a complete reversal of the response pattern in expert subjects concerning the two extreme options of doing nothing versus the push option, solely as a function of the number of available options. Whereas the majority of our expert ethicists (68%) preferred to do nothing in Two Options, the majority (66%) chose the push option in Six Options.

Second, we also found very significant order effects in expert ethicists in both directions, from Two Options to Six Options and vice versa, concerning the two "extreme" options of doing nothing versus push. When Two Options is presented first, a clear majority of experts (68%) prefers doing nothing, but when Two Options is presented last, a narrow majority (55%) prefers the push option; and when Six Options is presented first, a clear majority (66%) prefers the push option, but when Six Options is presented last, a clear majority (59%) prefers doing nothing. We argue that both effects are due to the influence of morally irrelevant factors, and that this has unfavourable implications for the expertise defence against the challenge from experimental philosophy. In line with previous research, we found no evidence that the moral intuitions of expert ethicists are more resistant to the influence of morally irrelevant factors than those of lay people.

"Don't believe the hype: Conducting (and building on) fMRI research to inform philosophical issues - Rodrigo Diaz

Methodological issues have recently become a "hot topic" inside the field of experimental philosophy. In particular, there is an increased awareness about the problems related to the field's excessive reliance on questionnaires. The main problem with questionnaires is their susceptibility to framing effects: the influence of how a question is presented in the responses to it. Framing effects, and the related problem of experimenter bias (framing questions so to get the desired results), can call into question the reliability and significance of the data obtained (Strickland and Suben 2012; Andow 2016). Adopting new methods could increase the reliability of experimental-philosophical results, perhaps providing converging evidence about the issues of interest. However, new does not always means better.

The aim of this talk is to analyze the prospects of using, and building on research using, neuroimaging methods (fMRI in particular) for answering philosophical questions.

On the one hand, if we aim to uncover the cognitive processes underlying people's intuitions about philosophical issues (Knobe 2016), we will face a serious problem when using fMRI. Inferring cognitive processes from patterns of brain activation, or reverse inference, requires brain areas to be (at least to some extent) selective for specific cognitive processes (Poldrack 2006). However, decades of fMRI research have provided a brain-function mapping that is far from selective, with the same brain areas involved in many different cognitive processes and the same cognitive processes recruiting activity in many different brain areas (Poldrack 2010; Anderson et al. 2013). Although there have been attempts to "save" reverse inference by reformulating it (Machery 2013; Del Pinal and Nathan 2013), I will argue that this reformulations are still problematic, and they drastically reduce the significance and possibilities of using fMRI data to make inferences about cognitive processes.

On the other hand, however, the same collection of data that does not help us to infer cognitive processes from maps of brain activation can inform philosophy in a different way. The lack of a one-to-one correspondence between cognitive functions and brain areas raises the question of which is the right taxonomy for understanding cognition, or cognitive ontology (Anderson 2015). Perhaps this imperfect brain-function mapping is due to the fact that the categories we currently use to understand the mind have no "ontological reality", and we have to revise them. For example, the fact that emotions recruit activity in different brain areas that are in turn involved in other psychological processes has been used to argue against a "natural kinds" approach to emotions and in favour of psychological constructionism (Barret and Satpute 2013).

To sum up: philosophers should be interested in fMRI research because of the interesting philosophical questions it poses more than for the solutions it offers.

Empirical Philosophy, Biases, and Systematic Reviews - **Andrea Polonioli**

Naturalism comes in different forms, and a distinction is often made between experimental and empirical philosophy (Prinz 2008). Whilst a number of challenges for experimental philosophy have already been presented in the literature (e.g., Carmel 2011), possible limitations of the methodology used by empirical philosophers have not received the attention they deserve. This article argues that if empirical philosophers wish to reduce bias in their research, they should then supplement their traditional tools for literature search and review by including systematic methodologies currently in use in many scientific disciplines (Cooper 1998) and recently advocated also in non- empirical contexts (Strech and Sofaer 2012).

More precisely, the paper presents and defends the following claims. First, empirical philosophers do not philosophise in a vacuum and, in fact, rely on literature search and review in a number of ways and for several purposes. Second, biases and cognitive limitations are likely to affect literature search and review in many critical ways. Over the past decades, psychologists have described numerous ways in which judgment formation and information search can be biased, and there are no reasons to doubt that also literature search and review should be biased in important ways (Baumeister and Leary 1997), and even in the field of philosophy. Third, scientists have come to widely adopt systematic reviews to minimise bias in the activity of literature search and review, and these tools should also be of wide interest to empirical philosophers, who have instead heavily relied on narrative approaches to literature review.

Systematic research review is a highly structured approach to cumulating knowledge. Progress in knowledge acquisition is the result of the integration of efforts, and literature reviews are vehicles for summarising research. For systematic reviews, a clear set of rules exists for searching studies and for determining which should be included in or excluded from the analysis. The reproducibility of an experimental result is a fundamental assumption in science and in a similar fashion systematic reviews aim to allow for high reproducibility of conclusions by minimising bias and maximising transparency.

Distinguishing the Phenomenal from the Cognitive - **Kevin Reuter and Rodrigo Diaz**

The divide between identifying emotions with either the feeling aspects or the cognitive aspects "is one of the major fault lines within emotion research" (Prinz 2004: 22). Whereas feeling theories take the phenomenal aspects to be at the core of emotions, cognitivists argue that emotions are

conceptualized in order to explain the representational properties of emotions, and thus believe that they are reducible to (or at least based on) judgments or some other cognitive states.

Fascinating as the debate between feeling theories and cognitive theories may be, its motivation hinges largely on there being a lack of conceptual sophistication to separately pick out the feeling aspects and the cognitive aspects of the emotions. Imagine for a moment that instead of having a single concept of happiness that refers to the emotional state of happiness simpliciter, we would be in possession of two different concepts, happiness1 referring to the experiential or feeling aspects of happiness, and happiness2 referring to the cognitive aspect, i.e. a propositional attitude that in conjunction with an intentional object expresses an evaluative judgment. Now, most philosophers readily assume that people possess just a single concept of happiness, anger, sadness, etc. Others, e.g. Peter Hacker, have explicitly pointed out that "there is no difference between having an emotion and feeling an emotion (being jealous and feeling jealous)" Hacker (2009: 18). Unfortunately, no reason is given to support this conjecture.

In order to test whether people distinguish the phenomenal from the cognitive aspects of emotions, we conducted studies using both traditional questionnaires as well as corpus-linguistic analyses.

In our corpus analysis study we investigated possible differences in the use of the expressions 'feeling x' and 'being x' on the world-wide web. The results suggest the falsity of Hacker's claim that there is no difference between having an emotion and feeling an emotion. English speakers primarily use 'being happy', 'being sad', as well as 'being angry' when they refer to an emotional state that is directed at an intentional object, whereas they state more often that they 'feel happy', 'feel sad', and 'feel angry' when an external event has caused them to be in an emotional state

In a second experiment, we directly asked 119 participants to read one of two vignettes in which it was specified that a person named Tom:

i. evaluated a situation as dangerous but did not experience any uneasiness, or

ii. did not evaluate a situation as dangerous but experienced uneasiness.

The participants were then prompted to select one of the following four response options: "Tom is afraid.", "Tom feels afraid.", "Both of the above.", "Neither/Other".

The data confirmed the results we obtained in the corpus analysis. A majority of the participants who were assigned to condition (i) selected "Tom is afraid", while a majority of the participants who were drawn to condition (ii) thought "Tom feels afraid" made the most sense.

In sum, the results indicate that, pace the philosophical consensus, people seem to distinguish between the phenomenal and the intentional aspects of emotions.

Individual Differences in Moral Taste Explain Condemnation of Mind Upload Technologies - Michael Laakasuo

Research in cognitive sciences and evolutionary psychology suggests that people are ontological substance dualists: we intuitively separate the "mind stuff" from matter. Dualistic ontology and the concept of "souls" have been important in the development of our moral cognition; moral issues regarding the fate of the soul date back to the ancient Egypt and the Book of the Dead; anthropological evidence suggests that this could extend even to our hunter-gatherer past. Individuals' moral life is tied to the fate of the soul in the afterlife in almost all major religions from Buddhism to Christianity. Morality of the soul is thus connected to moral issues of the surrounding societies as well.

In our paper we present three studies (Ns 268, 221 and 160) investigating moral condemnation of the act of uploading one's consciousness or "mind" (i.e. one's "essence"/"soul") into different platforms. In our first study we observed that Purity Norm adherence (from Haidt's Moral Foundations scale) predicts condemnation of uploading one's mind into a computer. In our second study we found that condemnation was not affected by whether the target of the mind upload was an android, a computer, a gorilla, or another biological human brain. However, we replicated the effect of Purity Norm adherence across all conditions – even after controlling for religiosity, technological literacy, gender and education. In our third study we evaluated whether condemnation was affected by the body itself staying "alive" after the mind upload (i.e. whether or not the heart kept pounding afterwards), but found no effects.

Moreover, individual differences in Death Anxiety predicted positive attitudes towards mind upload. The more anxious our participants were about the possibility of physical death, the more approving they were of mind upload. Males had more positive attitudes towards mind upload than females. In all three studies we also controlled for the level of science fiction hobbyism – a new scale currently under development. We found that familiarity with science fiction predicted approval of mind upload technologies, and fully mediated all gender differences in all three studies (males were more familiar with science fiction than females, which also fully explained their higher approval of mind upload). In each study the statistical model explanatory values were extremely high for standard moral psychology experiments (r2s = ~0.3).

In conclusion, given that adherence to purity norms predicted mind upload condemnation, there could be an evolutionarily old link between intuitive dualism and human group living. This has not been thoroughly investigated previously. Moreover, understanding the underpinning of moral cognition and its links to natural dualism might clarify how our philosophical intuitions function and, also, help us understand modern day and future moral debates with respect to implementing new technologies. For example, there are several companies in Silicon Valley that promise to transfer human consciousness into more durable platforms within the next 20 years. Against this backdrop, our studies provide an opening for research in the intersection of moral psychology and intelligent machines and AI technology.

Experimental Philosophy: Methods from Economics - **Boudewijn de Bruin**

One way to go beyond the questionnaire may come from economics. In this presentation I discuss a number of methods from economics that may help experimental philosophers. I do this (i) by describing in general terms what the added value is of using economic and econometric techniques, and (ii) by giving examples of potential applications. My aim is not to present economics as an alternative to psychology, neuroscience, digital humanities, or other X-phi methods, but rather as a complement. The main idea is that looking at experimental philosophy through the lens of economics allows us to find – and hopefully answer – new questions.

I show that what makes economic methods generally attractive are the following features:

- Real-life data. While psychologists and neuroscientists largely work with data gained through experimentation, economists generally use existing data (stock price, household wealth, purchasing behaviour, GDP, etc.), or data from 'natural experiments' (e.g., a policy change without an experimenter intervention). This enlarges the chance that they have 'real bite' and 'ecological validity' (validity in the real world).

- Long-term data. Psychologists can generate long-term data, but economists generally only need to collect them, so they use them more often. This makes it easier for them to study trends, to spot extremes, and to increase robustness as compared to one-shot studies.

- Aggregated data. Economists often have datasets with large numbers of observations, often international. This decreases the risk that one's findings only apply to a small segment of the world population (say, philosophy students or English speaking Amazon MTurk participants).

- I discuss the following examples. With the intended conference audience in mind, my aim is here specifically to illustrate particular methodologies rather than to highlight philosophical claims.

- Epistemic injustice captures a form of injustice resulting when structural negative prejudices make people discredit the value of testimonies of members of disadvantaged groups. While epistemic injustice has been demonstrated in experimental settings, is it unclear whether this is a large-scale phenomenon, and what it effects are. I show that – perhaps surprisingly – economic research on racial disparity in business lending offers relevant insights.

- Free market beliefs. What explains why people adopt particular normative beliefs? How do they cohere with other beliefs? A result on what economists call 'motivated beliefs' showed persuasively in a 'natural experiment' that people adopted 'free market' beliefs after their environment changed (exogenously through a policy change) from 'cooperative' to 'capitalist'. I argue that this type of research offers promising methods to empirically investigate doxastic (in)voluntarism.

- Situationism. The last example tongue-in-cheek returns to the questionnaire. But I show how we can gain some of the above mentioned advantages (real-life, long-term, etc.) by looking at experiments in business economics conducted in large multinational companies, ensuring high cross-cultural validity (translation of questionnaire is non-trivial). While my main contribution is methodological,

I briefly advertise that my data offer some empirical evidence against situationism, which to my knowledge is new.

Priming Effects in Experimental Contextualism - **Su Wu**

In epistemology, contextualism is an influential response to skepticism. It contends that standards for knowledge are sensitive to contexts. In skeptical contexts, the standards for knowledge are too high to regard any statements as knowledge while in our daily life the standards for knowledge are not that strict. Therefore, it is still possible for us to have knowledge in spite of the challenges from skepticism.

Many philosophers agree that people's standards for knowledge vary across cases. However, they disagree on how to explain this variation. On the one hand, contextualism holds that it is the contexts of the attributors of knowledge that explain the variation. On the other hand, Interest Relevant Invariantism (IRI, henceforth) argues that it is the practical factors the knowing subjects need to consider that make the difference (Stanley, 2005). With the development of experimental philosophy, both proposals can be tested empirically. In effect, this debate has become a heated topic in experimental epistemology.

Nevertheless, pervious experiments based on scenarios presented by Derose (1992) or their variants have little hope to resolve the dispute between contextualism and IRI. As mentioned above, the disagreement between contextualism and IRI is whether the context sensitivity of knowledge attribution should be explained by the contextual factors confronting the attributors or practical factors facing the subjects. Previous experiments focus on the effects of factors such as stakes or mention of errors that are presented in the scenarios. It is hard to tell whether these factors should be attributed to the protagonists in the scenarios (as the subjects) or to the research participants (as the attributors), since both attributors and subjects are aware of the factors mentioned in the scenarios. Even though it is emphasised in the narrations that the subjects are ignorant of stakes or possibilities of errors, participants may still tend to think that the subjects have as much access to the information as they do because of epistemic egocentrism (Alexander, Gonnerman, & Waterman, 2014). Therefore, once there is a difference in people's inclination to knowledge attribution, it is difficult to decide which theory wins the game, since both can be compatible with the data.

To isolate the contextual influence on the attributors, our study appealed to priming techniques, which have been widely used in psychology. We tried to investigate if stories like 'the brain in vat' could serve as primes to make the skeptic contexts salient and raise the standards for knowledge for participants. In the experiment we used a brief description of "the brain in a vat" and a brief introduction to "split-brain" as primes for experimental group and control group respectively. 433 effective responses from Chinese participants were recycled online and a statistically significant difference has been found. Participants primed by "the brain in vat" are less inclined to attribute knowledge to protagonists in all scenarios we used. Given that priming can only influence the attributors, the difference cannot be explained merely by contextual factors confronting protagonists. Thus, this study provides an independent support for contextualism.

Natural talent, work, and desert - **Aurélien Allard**

There has been a long-standing debate about the moral significance of natural talents for social justice. Classical philosophers have debated whether natural talents should come into account in the determination of how deserving someone is, either adopting a perspective sympathetic to efforts alone (Kant 2012) or attributing some importance to contribution as well (Smith 2010). In contemporary debates, this has become a central issue. So-called luck egalitarians (Arneson 1989) have famously denied that luck (including natural talents) should have any importance in a fair distribution of goods. Even if the rejection of natural talents as a desert basis is dominant in

contemporary moral philosophy, some theorists of desert (Miller 2001) have insisted that contribution, independently of efforts, is the proper basis of desert.

This philosophical debate meets classical research in moral psychology. In the 1970s, a major branch of research has tried to evaluate whether the folk did take into account output and talents, independently of effort (Leventhal et Michaels 1971). These experiments have shown a clear tendency to devalue idle talents, if not put to good use. However, the relative importance of contribution and efforts has remained unclear.

More recently, two experimental philosophers, Freiman and Nichols (2011), have tried to show that folk intuitions regarding the moral relevance of natural talents depended on whether abstract of concrete cases were considered. In an abstract condition, they presented to participants the following sentence: "Suppose that some people make more money than others solely because they have genetic advantages." In a concrete condition, a singer made more money because, solely due to genetic advantages, she had a better voice. Participants were then asked whether they considered that these people deserved to make more money. Participants gave higher ratings of desert in the concrete condition than in the abstract condition, which was interpreted by Freiman and Nichols as showing that the rejection of natural talents depended on the abstractness of the cases considered.

Against this interpretation, I argue that the abstract formulation was likely to elicit misunderstanding, as it suggested that some people might deserve to earn more money uniquely due to superior genes, independently of any work actually performed. To test this prediction, I have conducted a replication-extension of the experiment, asking participants for justifications and running an additional condition with the following formulation: "Suppose that some people make more money than others because, solely due to their genetic advantages, their work is more productive or more appreciated." As predicted, participants' justifications suggest that the original formulation is likely to have produced misunderstanding, as some participants interpreted the question as asking whether white people deserved to earn more solely because of their skin colour. Furthermore, the more precise wording significantly increased participants' willingness to judge that the genetically-advantaged people deserved to earn more. On a methodological level, the results shed light into the importance of precisely studying participants' justifications of their choices, and of keeping all factors constant across both abstract and precise vignettes. On a substantive level, the results suggest that the folk might be more sympathetic to contribution views of desert than previously assumed, indicating a bigger gap with philosophers' traditional rejection of this view.

The Potential of Longitudinal Studies - Ethan Landes

Experimental philosophers have overwhelmingly relied on data generated by individuals at one point in time, even when investigating diachronic phenomena such as philosophical expertise (e.g., Livengood, Sytsma, Feltz, Scheines, & Machery, 2010; Schwitzgebel & Cushman, 2015; Tobia, Chapman, & Stich, 2013). While a simple way to study philosophical expertise or the reliability of intuitions, cross-sectional designs leave important questions unanswered. In this talk, I explore the potential of longitudinal studies—discussing some initial data of my own. I argue that, not only can longitudinal studies answer a number of unexplored questions about the epistemic status of intuitions, but they can be easier to run than commonly thought.

Rini (2015) recently proposed a new line of the experimental attack focused on diachronic instability and self-knowledge. She argues that if intuitions are diachronically unstable, and philosophers do not recognise that their intuitions change over time, then philosophers would be unknowingly basing their views o of unreliable mental states. She points out, rightly, that such a study on professional philosophers would be prohibitively expensive and difficult. Despite Rini's pessimism, experimental philosophers can make ground on the question of temporal stability in easier, short-term studies on undergraduates. I have begun such a project, having developed a framework and completed an initial study focused on stability of intuitions over exposure to related class- room instruction.

Thee data I have collected, while not perfect, suggest that intuitions are remarkably stable over time—at least in undergraduates—even when exposed to classroom instruction related to the thought experiments. Students were shown a Gettier case, a variation of Mr. Truetemp, and a variation of Kripke's Gödel case at the beginning and end of a five week unit on epistemology. No

significant directional change occurred and, in the case of Mr. Truetemp and the Gettier case, 80% of students did not change their answer after five weeks. This offers additional support to the anti-expertise defence work of Machery (2015); Schwitzgebel and Cushman (2012); Tobia, Chapman, and Stich (2013) (among others), as well as claims that epistemic intuitions come from fairly encapsulated cognitive systems (e.g., Nagel, 2012; Spicer, 2008).

This presentation will also defend the potential power of longitudinal studies by discussing the questions planned future research will try to answer. In particular, it is unclear from current studies whether truth-irrelevant sensitivities only affect a portion of population or affect the whole population a portion of the time. These two possibilities raise very different challenges to the case method. One merely threatens the intuitions of certain individuals, but the other threatens all individuals. By replicating parts of Tobia, Buckwalter, and Stich (2013) at multiple points in time, I hope to separate the two possibilities. Tobia et al. found that the actor/observer bias had a large effect on intuitions of moral cases. By analysing how an individual's answers to moral thought experiments change over time—as opposed to previous studies which have merely studied group averages—the study will provide insight into how stable intuitions are over time and provide new insight into the epistemic value and cognitive source of intuitions.