Morality, Language, and Thought Workshop
July 24, 2017

Institut Jean Nicod
Ecole Normale Supérieure 29, rue d'Ulm
Pavillon Jardin
75005 Paris, France

Organized by:

Laura Niemi (lauraniemi@gmail.com)   Brent Strickland (stricklandbrent@gmail.com)
Associate, Department of Psychology, Harvard University   CRNS Researcher, Institut Jean Nicod
Postdoctoral Fellow, Department of Philosophy & Center
for Cognitive Neuroscience, Duke University

Talks: *20 min with 10 min for discussion/Q+A; please bring your presentation on your own computer and also a backup PDF of your presentation on a USB drive in case there are problems with display.*

PROGRAM	Room: *Salle Paul Langevin*

9:15am-9:30am   *Greeting, Breakfast Provided*

9:30am-10:15am   **Paul Egré**, IJN   Opening Keynote
          Subjectivity in gradable adjectives: the case of tall and heavy (Joint work with Steven Verheyen)

10:15am-10:30am   Coffee Break

10:30am-11:00am   **Tobias Gerstenberg**, MIT
          A counterfactual simulation model of causal judgment

11:15am-11:45am   **Julian DeFreitas,** Harvard University
          High Level Visual Processing Alters Moral Judgment

11:45-11:50pm   Break

11:50am-12:20pm   **Phil Wolff,** Emory University
          Intentionality and Causal Structure: Insights from Perceptual Judgments and Big Data Analyses of the
          English Lexicon (Joint work with Jason Shepard, Middle Georgia State University)

12:20pm-12:50pm   **Bridget Copley**, CNRS/Université Paris 8
          The grammatical lightness of being in charge

12:50-2:00pm   Lunch Provided

2:00pm-2:30pm   **Jonathan Phillips,** Harvard University
          Prescriptive norms constrain default representations of possibility

2:30pm-3:00pm   **Ana Gantman,** Princeton University
          Motivated Moral Perception

3:15pm–3:45pm   **Alon Hafri,** University of Pennsylvania
Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level processing

3:45pm–4:15pm   **Florian Cova,** University of Geneva, Swiss Center for Affective Sciences
Does the effect of moral considerations on mental states ascriptions tell us anything about theory of mind?

4:20pm–4:50pm   **Roger Giner-Sorolla**, University of Kent
Is Evil Contagious, or Just Bad Optics? (Joint work with Tom Kupfer)

4:50pm–5:20pm   **Robert May**, UC Davis
The Moral Content of Pejoratives

5:20pm–5:50pm   **Max Kleinman-Weiner**, MIT
Learning a Commonsense Moral Theory

6:00pm–7:00pm   **Jesse Snedeker**, Harvard University   **Keynote**
Clean Mapping: How conceptual structure might serve as the developmental and phylogenetic starting point of syntax

---

ABSTRACTS

Opening Keynote:

Subjectivity in gradable adjectives: the case of tall and heavy
Paul Egré, joint work with Steven Verheyen

We propose an investigation of the ways in which speakers' subjective perspectives are likely to affect the meaning of gradable adjectives like "tall" or "heavy". We present the results of a study showing that people tend to use themselves as a yardstick when ascribing these adjectives to human figures of variable measurements: subjects' height and weight requirements for applying tall and heavy are found to be positively correlated with their personal measurements. We draw more general lessons regarding the definition of subjectivity and evaluativity in language, and regarding the ways in which a standard of comparison and a significant deviation of that standard are specified.

A counterfactual simulation model of causal judgment
Tobias Gerstenberg, MIT

How do we make causal judgments? In this talk, I will present a counterfactual simulation model (CSM) of causal judgments that unifies different philosophical views on causation. The CSM predicts that people's causal judgments are influenced by the extent to which a candidate cause made a difference to whether the outcome occurred, and how it occurred. I will show how whether-causation and how-causation can be expressed in terms of different counterfactual contrasts defined over the same generative model of a domain. I will focus on applying the CSM to the domain of intuitive physics, asking people to make judgments about colliding billiard balls. The CSM accounts for participants' causal judgments to a high degree of quantitative accuracy. Causal judgments increase the more certain participants are that a ball was a whether-cause, a how-cause, as well as sufficient for bringing about the outcome. The CSM postulates that people make causal judgments by comparing what actually happened with what would have happened if the candidate cause had been removed from the scene. In direct support of this claim, I will show eye-tracking data of how people spontaneously simulate how the counterfactual world would have unfolded. I will conclude by discussing how the CSM can be extended to model causal judgments about omissions, judgments about physical support, and how the model may help us better understand the mapping between causal events in the world and the words we use to describe them.

High Level Visual Processing Alters Moral Judgment
Julian De Freitas, Harvard University

How does vision interface with the rest of cognition? The interaction between visual cognition and moral cognition offers a unique opportunity to address this question, because the visual system appears to have mechanisms dedicated to inferring whether one object caused another to move (Michotte, 1963), and moral judgments often hinge on an assessment of cause-and-effect. However, it is unknown 1) whether perceived causality is an inference that only occurs when making causal judgments, or whether it is computed automatically, and 2) although it has been known for decades that causal inferences affect moral judgments (Cushman, 2008), it is unknown whether the visual system's representation of causality per se alters moral judgments. We find confirmatory evidence for both possibilities, suggesting that moral judgment does not only accept causal inferences from slow (i.e., deliberative) cognitive systems, but is also influenced by fast, automatic causal inferences made by the visual system.

Intentionality and Causal Structure: Insights from Perceptual Judgments and Big Data Analyses of the English Lexicon
Phillip Wolff, Emory University & Jason Shepard, Middle Georgia State University

Recent studies have found that ascriptions of intentional action are higher when people see an actor bring about an outcome that is bad rather than an outcome that is good. The results have been interpreted as indicating that ascriptions of intentional action are based on judgments of badness, or norm violations. An alternative explanation ties ascriptions of intentional action to a particular kind of causal structure, one that can arise when people act against a preventive force. Consistent with this latter account, we observed that actions leading to bad outcomes are associated with normative pressures (Experiment 1a), and that these pressures give rise to a specific kind of causal structure (Experiment 1b). When these causal structures are pitted against the badness of the outcome, ascriptions of intentional action track with causal structure and not badness (Experiment 2), and asymmetric ascriptions of intentional action can be extended to physical forces devoid of norm violations (Experiment 3, 4a, 4b, 5). In a Big Data analysis of the English lexicon involving the automatic generation of definitions, we further show that the meanings of verbs of intentionality are closely related to notions of force and causation, and force and causation are related to badness, but intentionality is not closely related to badness. While the badness of an outcome or the norm status of an action may have an indirect effect on judgments of intentional action, our results suggest that the factors that affect judgments of intentional action most directly are non-evaluative and descriptive.

The grammatical lightness of being in charge
Bridget Copley, CNRS/Université Paris 8

There are cases where a matrix subject has authority over an embedded subject. Where does this meaning of authority come from, and why is so little morphology associated with it? We will see that neither authority nor even animacy is not part of the meaning of such cases in English; rather, there is a simple causal structure contributed by the sentence and authority is the main (but not the only) way for this structure to be interpreted.

Prescriptive norms constrain default representations of possibility
Jonathan Phillips, Harvard University

As humans, we think not only about what is, but also what could be. These representations of non-actual possibilities support many important cognitive functions, such as predicting others' future actions, assigning responsibility for past events, and making moral judgments. Prior research on our understanding of non-actual possibilities asks how humans explicitly and deliberatively reason about what is possible. Less well understood is whether or how people have a default, implicit representation of which events are possible. I present three studies that characterize the role of implicit representations of non-actual possibility in cognition. Collectively, these studies differentiate explicit reasoning about possibilities from default implicit representations, demonstrate that human adults often default to treating immoral and irrational events as impossible, and show that high-level cognitive judgments rely on default implicit representations of possibility rather than explicit deliberation.

Motivated Moral Perception
Ana Gantman, Princeton University

Readily identifying morally relevant stimuli is vital for social living; how do we flexibly recognize moral stimuli in our environment? Evidence suggests that visual perception can be tuned to morally relevant stimuli. Specifically, people detect moral words (e.g., kill, moral, should) with greater frequency than non-moral words (e.g., die, useful, could) when they are presented at the threshold for visual awareness (i.e., ambiguously). This effect is motivationally sensitive; when justice needs are satiated, the detection of moral words is selectively diminished, emphasizing that perceptual processes for detecting moral content may be flexibly tuned by context. Taken together, this work, at the intersection of morality, language and perception, provides evidence for thinking about moral psychology not only in terms of what kinds of content fall into the moral domain, but how malleable domain-general processes are tuned and utilized to create a specific cognitive signature for moral cognition.

Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level processing
Alon Hafri, University of Pennsylvania

In order to make moral judgments about observed events, one must encode who did what to whom, i.e. the event roles: boy hitting girl requires different blame assignment from girl hitting boy. We often categorize Agents and Patients from visual input to describe what we see or to encode events in memory. However, even when attention is otherwise occupied, do we spontaneously encode event roles? Across several experiments, participants observed continuous sequences of two-person scenes and had to rapidly identify the side of the person with a certain property (the male/female or red/blue-shirted person). Crucially, even though role was never explicitly mentioned and was completely orthogonal to the task (gender or color search), participants responded more slowly when roles switched from trial to trial (e.g., the male went from Patient to Agent). Our results suggest that some of the first building blocks for moral judgments are spontaneously extracted at first glance.

Does the effect of moral considerations on mental states ascriptions tell us anything about theory of mind?
Florian Cova, University of Geneva, Swiss Center for Affective Sciences

Numerous studies suggest that moral considerations impact our attributions of mental states. But what does it mean? So far, two main options have dominated the literature: either this phenomenon is the product of core competences, and they lead to revise our understanding of folk psychology of mind, or it is only a bias, and it tells us nothing interesting about folk psychology and theory of mind. Here, I wish to pursue of third approach: though this phenomenon reflects the proper working of core competences, it tells us nothing about folk psychology and theory of mind. More precisely, I will present studies suggesting that moral considerations only impact the semantics of psychological terms, but have no impact on how we think about and predict others' behaviour.

Is Evil Contagious, or Just Bad Optics?
Tom Kupfer & Roger Giner-Sorolla, University of Kent

Although a substantial experimental literature has shown people's unwillingness to wear or touch objects associated with evil, we argue that many of these experiments take place with an implicit audience (the experimenter) and that aversion can be explained by reputation management rather than a literal sense of moral contagion. In a hypothetical study (N = 102) people preferred the choice of wearing a historical Nazi armband next to the skin but hidden by clothing, rather than on top of clothing. Those who wore it on top mainly cited other reasons than contagion avoidance for doing so. In a behavioural study replicating this choice with a historical Nazi armband (N = 90), participants chose to wear the band over a piece of clothing mainly in conditions unseen by the experimenter, and again cited reasons of normality or expression rather than contagion for their choice. To see whether observers create a valid reason for these reputational concerns, a third, vignette study showed that people judged characters who chose to wear the Nazi armband in sight as being more disgusting and immoral than those who chose to hide it, whether or not the hidden armband was worn next to the skin. Moral disgust and disapproval, then, seem to be much more strongly influenced by how an activity looks than whether it can transmit an irrational essence of evil physically.

The Moral Content of Pejoratives
Robert May, UC Davis

What do pejorative words mean; what concept do they express? In this talk, I argue that they express morally loaded concepts that are grounded in defective ideologies (racism, homophobia, anti-semitism, Islamaphobia, etc), and that in virtue of expressing these concepts, they are unsatisfiable. That is, there are no such things as what these words purport to describe, much like mythological terms.

Learning a Commonsense Moral Theory
Max Kleiman-Weiner, MIT

We introduce a computational framework for understanding the structure and dynamics of moral learning, with a focus on how people learn to trade off the interests and welfare of different individuals in their social groups and the larger society. We posit a minimal set of cognitive capacities that together can solve this learning problem: (1) an abstract and recursive utility calculus to quantitatively represent welfare trade-offs; (2) hierarchical Bayesian inference to understand the actions and judgments of others; and (3) meta-values for learning by value alignment both externally to the values of others and internally to make moral theories consistent with one's own attachments and feelings. Our model explains how children can build from sparse noisy observations of how a small set of individuals make moral decisions to a broad moral competence, able to support an infinite range of judgments and decisions that generalizes even to people they have never met and situations they have not been in or observed. It also provides insight into the causes and dynamics of moral change across time, including cases when moral change can be rapidly progressive, changing values significantly in just a few generations, and cases when it is likely to move more slowly.

Keynote:

Clean Mapping:
How conceptual structure might serve as the developmental and phylogenetic starting point of syntax
Jesse Snedeker, Harvard University

In this talk, I will entertain an old and powerful idea: that language reflects the structure of thought. This hypothesis has several attractive properties. First, it leads to a theory of language development that puts meaning (and thus communication) front and center, without attempting to reduce structure to function. This framework allows us to begin integrating work from infant cognition, semantic bootstrapping, syntactic bootstrapping, and statistical learning. This approach also provides a more satisfying evolutionary account, calling into question the underlying assumption of minimalism (that language is the result of just one evolutionary change). I will naively and optimistically conclude that this is a good direction to go looking in, and then you all will tell me why I'm wrong.