Learning and Doing:
Toward a Unified Account of Rationality in Belief, Desire, and Action

John Locke Lectures 2018

Lecture 1:

"Desire, Value, and Learning"
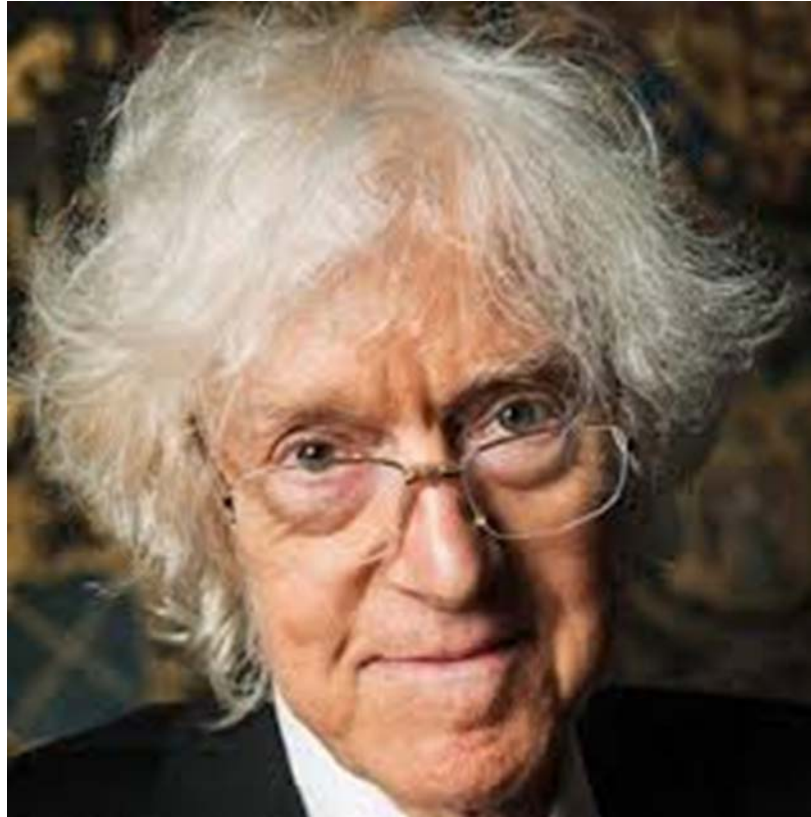
Peter Railton

(University of Michigan)

Oxford, May 2018

To **Derek Parfit** (1942-2017)

To **Derek Parfit** (1942-2017)

Who had passions for what matters,
lived them, and shared them freely

To **Derek Parfit** (1942-2017)

"He, on his side, had that faculty of beholding at a hint the face of his desire and the shape of his dream, without which the earth would know no lover and no adventurer." (Joseph Conrad)

"Don't make it a matter of course, but as a remarkable fact, that pictures and fictions give us pleasure, occupy our minds."

Wittgenstein, *Investigation*

"Learning and Doing"

- … could equally have been "Doing and Learning"
- About the intimate connection between learning and doing—the ways in which learning is for doing, and doing is, by its nature, a way of learning … whether we ask it to be or not.
  - The varieties of doing
  - The implicit or intuitive processes that underwrite doing
- Wait—aren't implicit or intuitive processes relatively simple, inflexible, heuristic-based, "push-button" responses (Greene, 2009), with "little understanding of logic and statistics" (Kahneman, 2011)?
- I'll argue that they are part of an intelligent, flexible, calculating infrastructure that operates in ways that are recognizably rational.

Learning and Doing

- I will be focusing on implicit, intuitive processes and the attitudes they involve not because I think that explicit, conscious deliberation and judgment are unimportant.
    - On the contrary, explicit, deliberative processes are central to making humans what they distinctively are.
- Rather, I don't see how such explicit, deliberative processes could operate in a manner that actually succeeds in being aptly responsive to reasons unless they were underwritten by intelligent implicit processes.
- That's the challenge, understanding how this is possible.

'What I can't build, I don't understand'

- So I set myself the task of trying to build core constituents of our capacity to be aptly responsive to reasons, such as:
  - desire
  - belief
  - intuition
  - inference
  - action
- … using whatever help I could find, including empirical psychology. In these lectures, I'll be trying to pull these together in a unified picture of mind.

Method?

- Field notes
- A job description—truisms, paradigm cases, dynamic and conceptual roles, etc.

- Then ask: What might fulfill these job descriptions and explain these field notes—as fully as possible, while not simply inventing or stipulating the facts.

Big picture? Naturalism vs. non-naturalism

- While I am a naturalist, this challenge could figure equally in the non-naturalist's project.
  - If we are to avoid skepticism about morality or rationality, we all need to see how apt responsiveness to reasons could be psychologically possible for humans.

- It took me a long time to see it, but Aristotle, Hume, and Kant set themselves similar projects.

Rationality

- *Narrow* – concerned with deliberative reasoning and control of thought and action by such reasoning
- *Broad* – concerned with apt responsiveness to reasons in thought and action, by non-deliberative as well as deliberative means; can include:
  - attention and perception
  - memory and imagination
  - inferential dispositions
  - emotion and motivation
  - implicit as well as explicit behavior
- We needn't fight over the word.

Why reasons?

- Talk of responding to reasons can sound odd.
- But 'reasons' is really a generic shorthand for speaking of how more substantive things relate to us as cognizers and agents.
- For example:
  - facts and evidence,
  - needs, goods and harms, our own or others'
  - relationships and practices,
  - communication and meanings.
- Reasons matter because these things matter.

Worries of regress

- As Aristotle (and later Hume and Kant) argued, we cannot ask reasoning to come up with the inputs to reasoning without risking a regress.

  - Somehow, we must get reason and reflection underway, and follow them through "spontaneously" or "intuitively", via forms of appreciation or understanding that are not simply deliberative.

- Let's call this notion of non-deliberative responsiveness *attunement.*

  - Given the vast variation and complexity of human life, and the changing contexts we inhabit and make, "blind dispositions" or "automatic" responses would not suffice.

Today

- … we will be focusing on desire and motivation.

- Questions of motivation have long played an outsized role in ethics and normative theory.

  - Today we'll mention two:

    - the dispute between cognitivists and non-cognitivists about the nature of moral judgment

    - the relation of desire to well-being.

'Judgment internalism'

- **… about moral judgments:** Moral judgments, it is argued, must have motivational force for the person making or accepting them, or else morality would fail to have a genuinely *practical* character—it would be possible to make a moral judgment sincerely while being indifferent to whether anyone, oneself included, acted accordingly (Stevenson, 1935; Gibbard, 2003).

  - This is the principal fault line in the debate between **cognitivism vs. non-cognitivism** that has dominated metaethics for the last half-century or more.

  - Why does involvement of *motivation* lead to *non-cognitivism*?

The "standard" or "Humean" model

- **belief   +   desire   →   action**
- *representational*
- *mind-to-world*
- *T/F*

The "standard" or "Humean" model

- **belief   +   desire   →   action**
- *representational*
- *mind-to-world*
- *T/F*
- *cognition*

The "standard" or "Humean" model

- **belief** + **desire** → **action**
- *representational*   *non-representational*
- *mind-to-world*   *world-to-mind*
- *T/F*   *not T/F*
- *cognition*

The "standard" or "Humean" model

- **belief   +   desire   →   action**
- *representational   non-representational*
- *mind-to-world   world-to-mind*
- *T/F   not T/F*
- *cognition   conation, not cognition*

However, there is something puzzling

- … about this way of seeing things, and the "standard" or "Humean" model has been increasingly criticized.

- Beliefs and desires might suffice to *cause* an action, but do they suffice to show that action was done for reason—even from the standpoint of the agent?

  - At one point, it was more or less the standard view in theory of action that beliefs and desires are *paradigmatic* reasons for actions—the agent's rationale is that she is seeking what she wants in light of what she believes.

- However, if desires are simply non-cognitive dispositions to act, what *rational intelligibility* can this confer on the action?

Motivation and rationality

- Warren Quinn (1993, 189): "Suppose I am in a strange functional state that disposes me to turn on radios that I see to be turned off. …
  - "I do not turn them on in order to *hear* anything. …
  - "My disposition is, I am supposing, basic"
- Quinn would be acting to bring about what he wants, given his beliefs, but he asks, "Does this state rationalize my choices?"
  - "Told nothing more than this, one may certainly doubt that it does."
- Or Derek Parfit, *On What Matters* (2:291):
  - "An irresistible impulse is not a normative reason."

Return to the model

- **belief   +   desire   →   action**
- *representational   non-representational*
- *mind-to-world   world-to-mind*
- *T/F   not T/F*
- *cognition   conation, not cognition*

Introducing rationality

- **belief**   **+**   **desire**   **→**   **action**
- *representational*   *non-representational*
- *mind-to-world*   *world-to-mind*
- *T/F*   *not T/F*
- *cognition*   *conation, not cognition*

- *rational*   *+*

Introducing ?????

- **belief   +   desire   →   action**
- *representational   non-representational*
- *mind-to-world   world-to-mind*
- *T/F   not T/F*
- *cognition   conation, not cognition*

- *rational   +   ?????*

A word from evolution …

- Could Mother Nature really have been so feckless in designing our motivational system as to leave it rudderless in the world with respect to considerations of attunement to reasons?

  – Even the standard argument that our perceptual and cognitive faculties would have been selected for representing the world around us with reasonable accuracy depends upon what we *do* with these reasonably accurate thoughts.

  – For that we need motivation, which would need to attune us to our needs and the conditions of our personal and social survival or flourishing, or those of our progeny and potential partners.

A further puzzle

- **belief** + **desire** → **action**
- *representational*   *non-representational*
- *mind-to-world*   *world-to-mind*
- *T/F*   *not T/F*
- *cognitive*   *conation, not cognition*

- *rational* + *?????* → *rational*

And yet another

- **belief   +   desire   →   action**
- *representational   non-representational   non-rep.*
- *mind-to-world   world-to-mind   world-to-mind*
- *T/F   not T/F   not T/F*
- *cognitive   conation, not cognition   volition,*

                                                        not cognition

- *rational   +   ????   →   rational*

"Irrational desire"

- Could there be rationality in desire? Aristotle thought so (*boulesis*). And commonsense often talks of desires as irrational—and not just in the means/end sense.
  - Obsessions
  - Compulsions and addictions
  - The ex-miser
- Desires—including intrinsic desires—can be seen as *mistaken, misplaced, misinformed, disproportionate,* or *excessively narrow.*

Desire and well-being

- I was trying to defend a cognitivist, realist approach in meta-ethics starting out from the notion of a person's well-being—how well her life is going, for herself. "Non-moral value".
  - We seem to know a fair amount about what makes lives go better or worse in this sense.
    - Paradigm cases: pain, anguish, happiness, health, meaningful connections with others, self-development and accomplishment, etc.
- Moreover, it seemed to be a truism that the individual's felt quality of life has a special role to play in her well-being—it would be an unacceptably alienated conception of someone's well-being if it could in no way be *experienced* as engaging or desirable.

"Informed desire"?

- An individual's well-being should be linked to what does or can *matter* to her, at least when she is widely experienced, well-informed, etc.
- There also seemed to be a truism allied with this one. Why "widely experienced and well-informed"?
    - People can learn from experience and general understanding whether their lives are going better or worse. We show deference to this.
    – These considerations seemed to point toward "informed desire" or "subjectivist" conception of a person's well-being, versions of which have enjoyed a long history.

Learning in desire

- But it seemed to me that the proper role of informed desire is as *evidence* of well-being.

- What *explains* why an individual's desires would, with experience, tend to favor certain ways of living?

- Such a grounding of her subjective preferences lies in objective, relational facts—like nutrition. A notion of "fit".

  - Some such facts will be generic—typical for beings like us.

  - Others will depend upon the nature of the individual and her particular capacities.

- Experience and information figure in the epistemology of well- or ill-being because they help us discover such a "fit".

Reduction?

- Normative concepts remain normative—*<reasons>, <well-being>, <reasonableness>, <health>*, etc.

    - But the *reason-making considerations* are matters of facts about us and the world.

    - The concept of *<reasons>* enables us to think about these considerations in a distinctive way, with special relevance to deliberation and action. But it is not an added ingredient needed to make them reason-making.

The nature of desire

- But what picture of desire would support such ideas?

Some field notes

Paris!

Paris!

Paris …

ÊTRE CHEZ SOI, CE N'EST PAS UN ENDROIT, C'EST UN SENTIMENT.
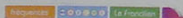
de • sire

dē- + *sider-*

apart, away + star

Kant's advertisement for the good will

- Even if it should happen that, owing to special disfavour of fortune, or the miserly provision of a step-motherly nature, this will should wholly lack power to accomplish its purpose, if with its greatest efforts it should yet achieve nothing, and there should remain only the good will (not, to be sure, a mere wish, but the summoning of all means in our power), then, **like a jewel, it would still shine by its own light**, as a thing which has its whole value in itself. [*Groundwork*, 4:394]

Desire: first pass
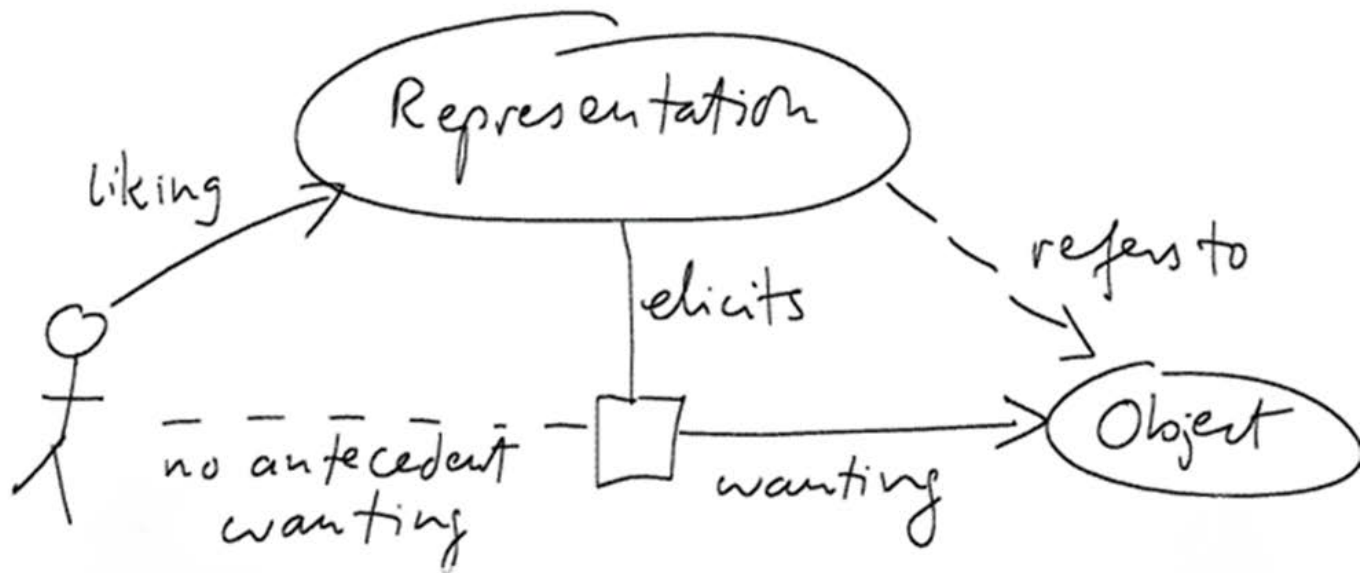
So why don't advertisers always win?

SUPRÈME BIÈRE
GANGLOFF
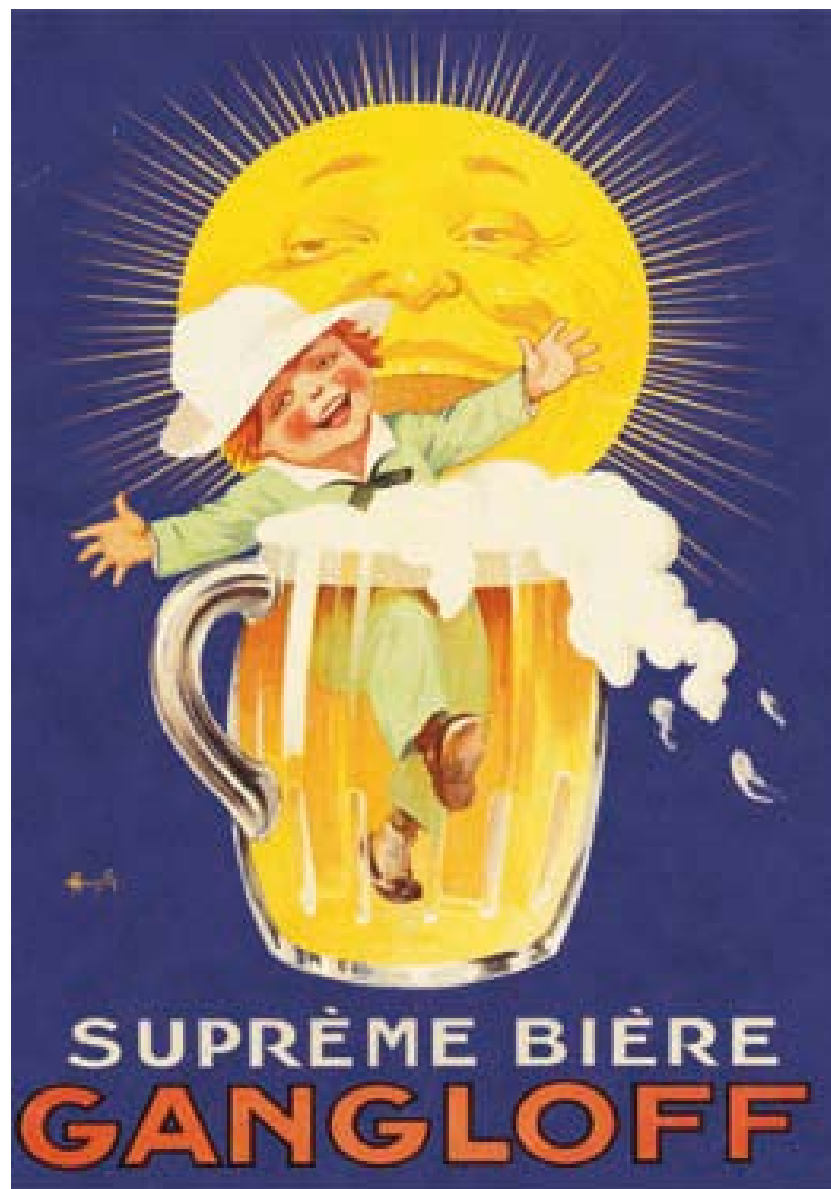
• "There are only two tragedies in life:

one is not getting what one wants,

Oscar Wilde

• "There are only two tragedies in life:

one is not getting what one wants, the other is getting it."

Oscar Wilde

Desire: second pass

Once as tragedy …

Once as farce …
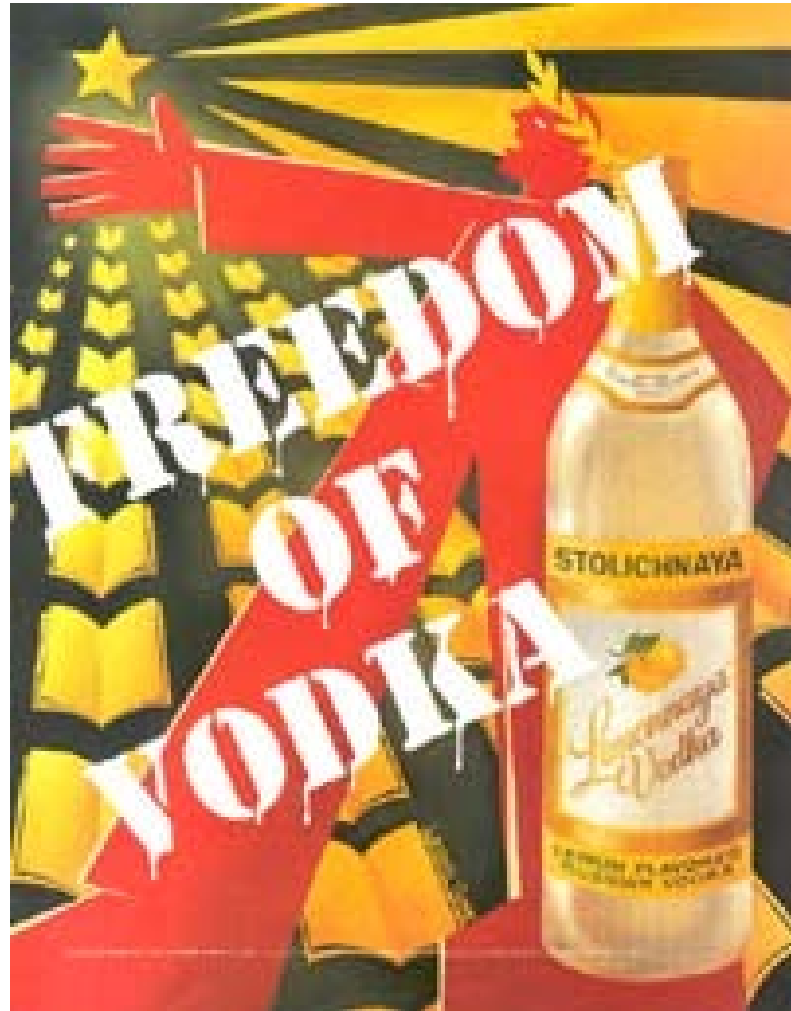
Desire: third pass

Desire, it seemed, was an open, learning state

- Rather than see motivation as a fixed battery or "set" of drives or sources of oomph,
  - … it should be seen as an exploratory, imagination-fueled investigation of the prospects … or perils … of the world.
  - These can be made possible by culture and new concepts—the distinctive and familiar ways in which human acquire new intrinsic desires through life.
- Affective appraisal (≠ *judgment* of value) has a certain primacy, so that recruiting of various forms of affect can yield desires not derived from prior desires.
- But affect is not the whole of desire—it is an integrated state involving the regulation of motivational resources by affective appraisal and responsiveness to feedback in guiding action.

Learning is hard

- This regulative structure makes action into experimentation, trial and error as affective expectations are compared to experienced outcomes.

  - Attunement of affective appraisal and expectation are possible through experience.

- This is often a hard and incomplete way to learn.

  - Many variables are involved in any actual action or outcome, and often outcomes become clear only much later. There are "received views" to overcome.

  - One must have available options and resources to explore them, including the experience of others and joint action.

But these difficulties go with the territory

- … of experience-based learning—whether in desire or belief.
  - To have the beginnings of a credible epistemology is to have a picture of how the dynamics might work:
    - to promote expectation-guided action that provides new and varied information
    - to make possible learning from action
    - to reduce the influence of arbitrary starting points and "received views"

Desire: fitting the job description

- (i) Desire is **non-voluntary, representationally-mediated,** and **motivating.** Because desire involves positive affect, it cannot be brought about voluntarily, and can "strike" us without deliberation or decision. Desire can take a variety of objects—acts, objects, outcomes, etc. –yielding motivational energy directed toward a "satisfaction condition" set by the representation of its object.
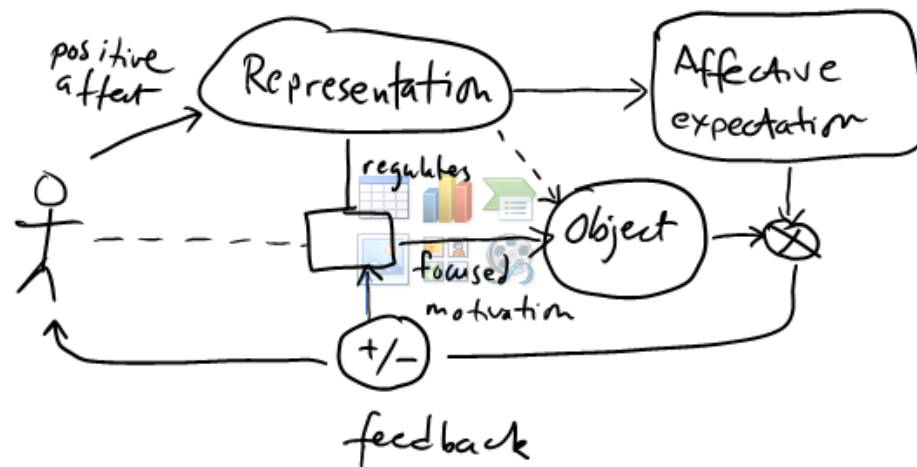
Desire: fitting the job description

- (ii) Desire is **prospective, comes in degrees,** and **responds to experience.** Affective states in general function to coordinate thought- and action-guiding capacities *prospectively*, sustaining expectations with specific contents and in varying degrees of strength, which then spontaneously respond to subsequent experience.

- $Like_t [A_{t+1} \rightarrow O_{t+2}] = d \rightarrow Expect_t [Like_{t+2} [O_{t+2}]] = d$
    $\rightarrow Perform [A_{t+1}] \rightarrow Outcome [O_{t+2}]$
        $\rightarrow Detect [Expect_t [Like_{t+2} [O_{t+2}]] - Like_{t+2} [O_{t+2}]] = r$
            $\rightarrow Like_{t+3} [A_{t+4} \rightarrow O_{t+5}] = d \pm r \rightarrow \ldots.$

Desire: fitting the job description

- (iii) **Desirability characteristic, bittersweet phenomenology, narrative structure:** While the *content* of a desire need not involve evaluative concepts, desire *presents* its content as attractive, sustaining positive expectation—the agent "sees something" favorable in her action or its outcome, yielding intelligibility. At the same time, the wanting and striving toward the object of desire sustains tension or longing, which can yield "practical pleasure" if the striving is successful and frustration if not. Desire thus has neither the phenomenology of pure positive affect nor that of pure wanting, but a structured combination that permits desire to be the foundation for narrative.

La planète désir

www.galerieslafayette.fr

Louis B. Mayer

- "I can't tell you how to make a great motion picture,
- "but I *can* tell you how to make a successful one. …"
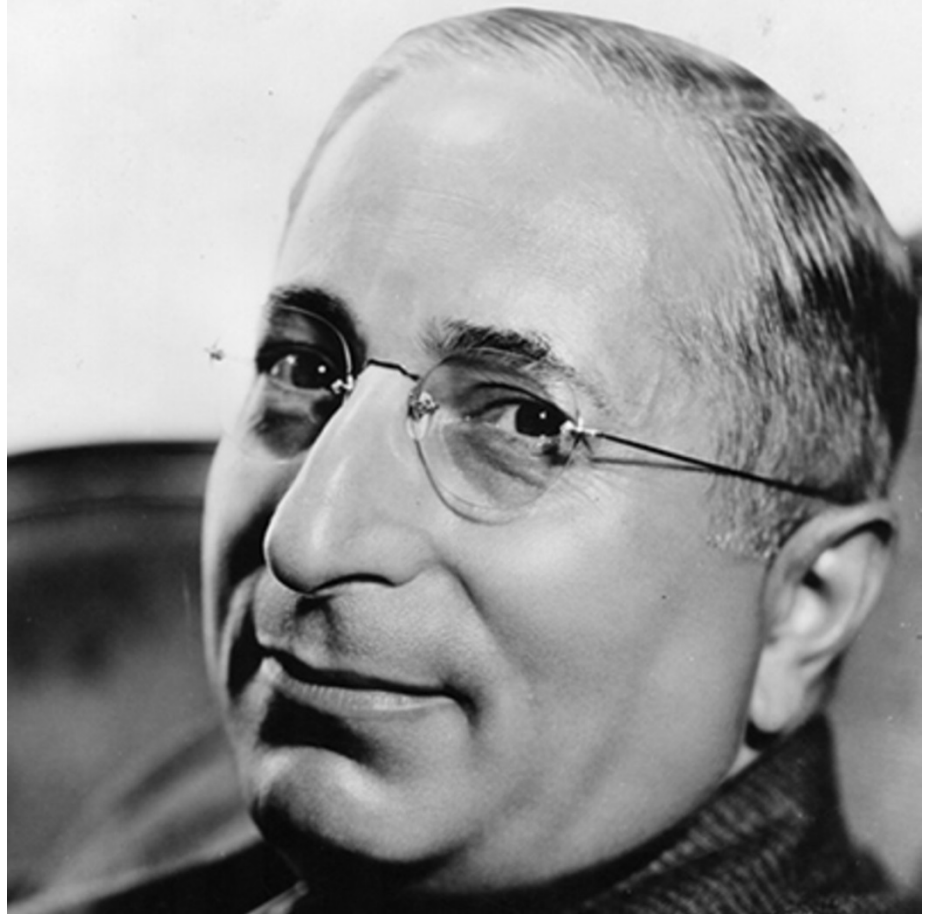
Louis B. Mayer

- "I can't tell you how to make a great motion picture,

- "but I *can* tell you how to make a successful one. …"

- "… You just make the audience *really* want something,

Louis B. Mayer

- "I can't tell you how to make a great motion picture,
- "but I *can* tell you how to make a successful one. …"

- "… You just make the audience *really* want something, then give it to them."
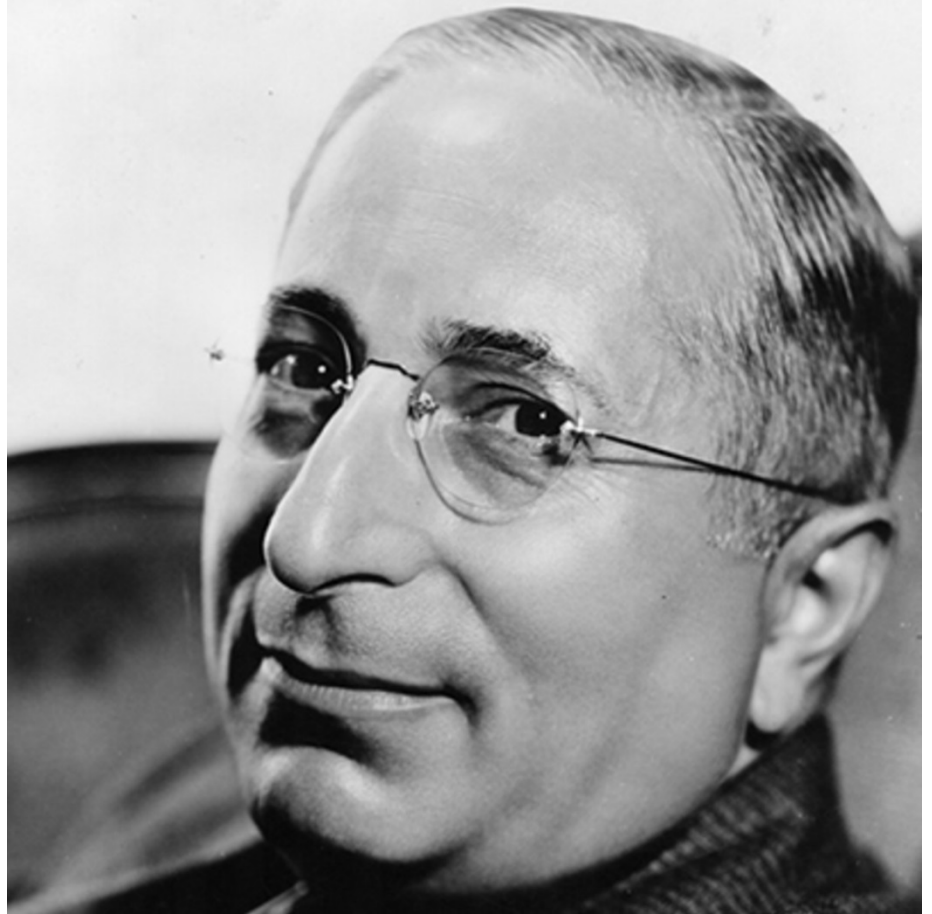
Desire: fitting the job description

- (iv) **Desire underwrites *intentional* action**. In contrast to the **casual product** picture:

  - **belief + desire → action**

- … in the **evaluative/regulative** picture of desire, the end is held "in view", so that the action is under its dynamic control—not just "pushed" by a thought of the end.
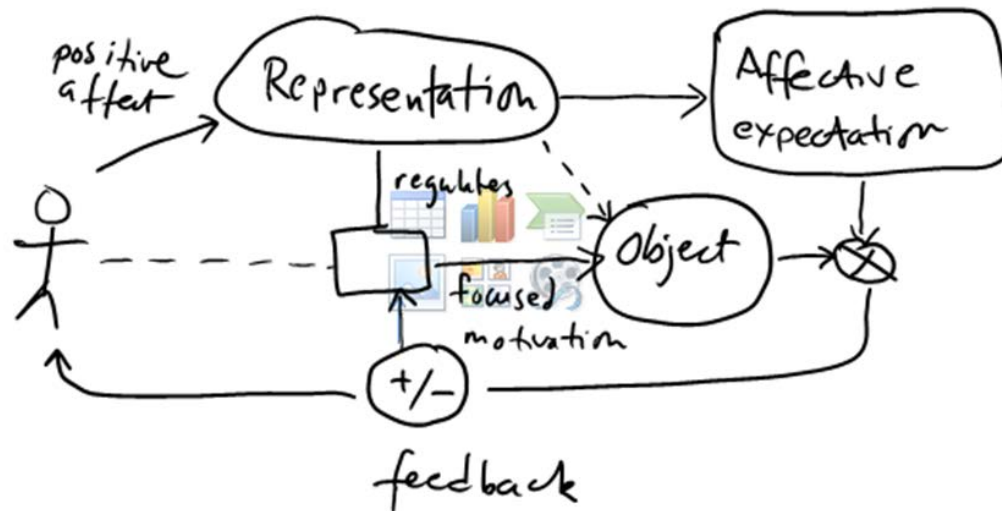
Desire: fitting the job description

- (v) Desire offers up a representation for **means/end reasoning** and **transfer of motivation** from ends to means. Affective transfer makes it possible that, as Aristotle puts it, practical deliberation *begins* with motivation and *concludes* with "the beginning of action"—"choice is either appetitive deliberation or deliberative appetition; and man is a principle of this kind" (Aristotle, *NE* 1139a):
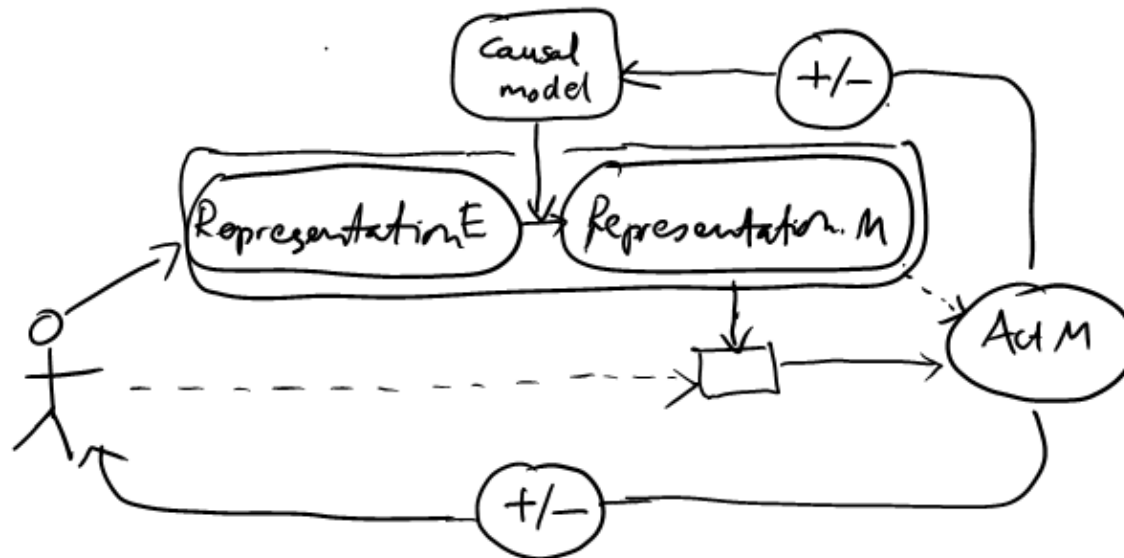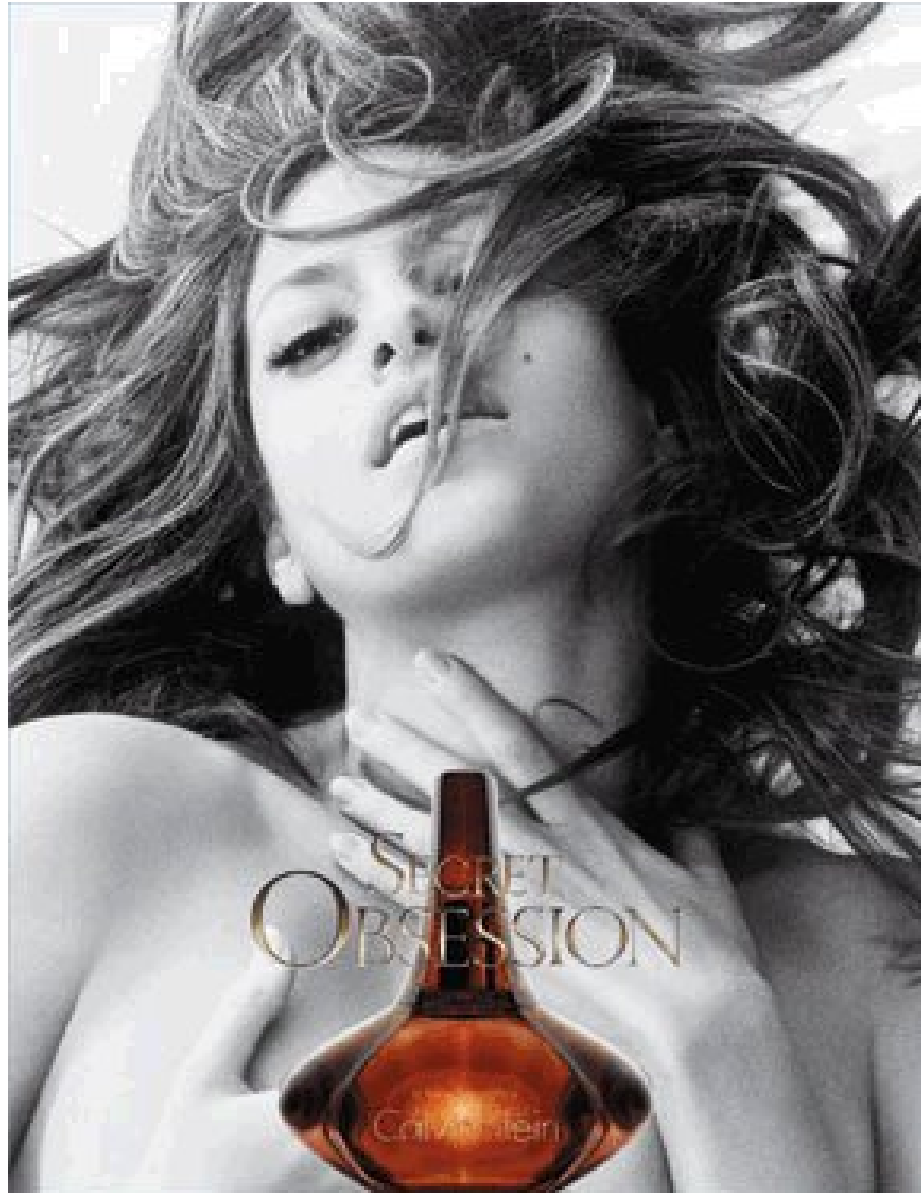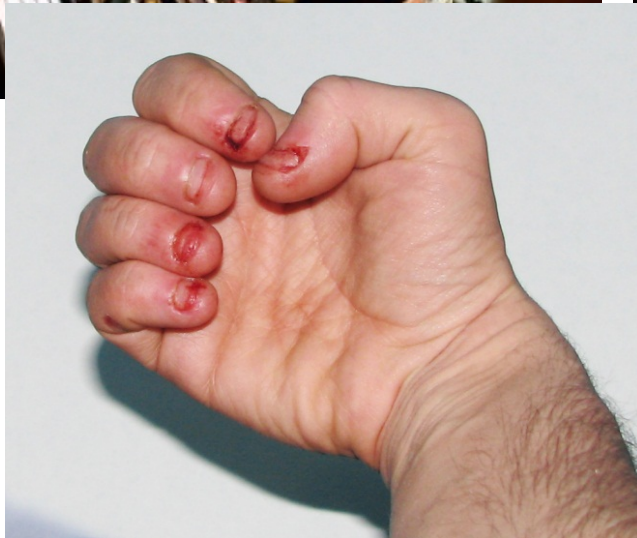
Desire: fitting the job description

- (vi) There are t**wo commonsense notions of *wanting* and *strength of wanting*.** The compound account makes this intelligible, and its regulative character helps explain how "weakness of will" is experienced as a loss of control.
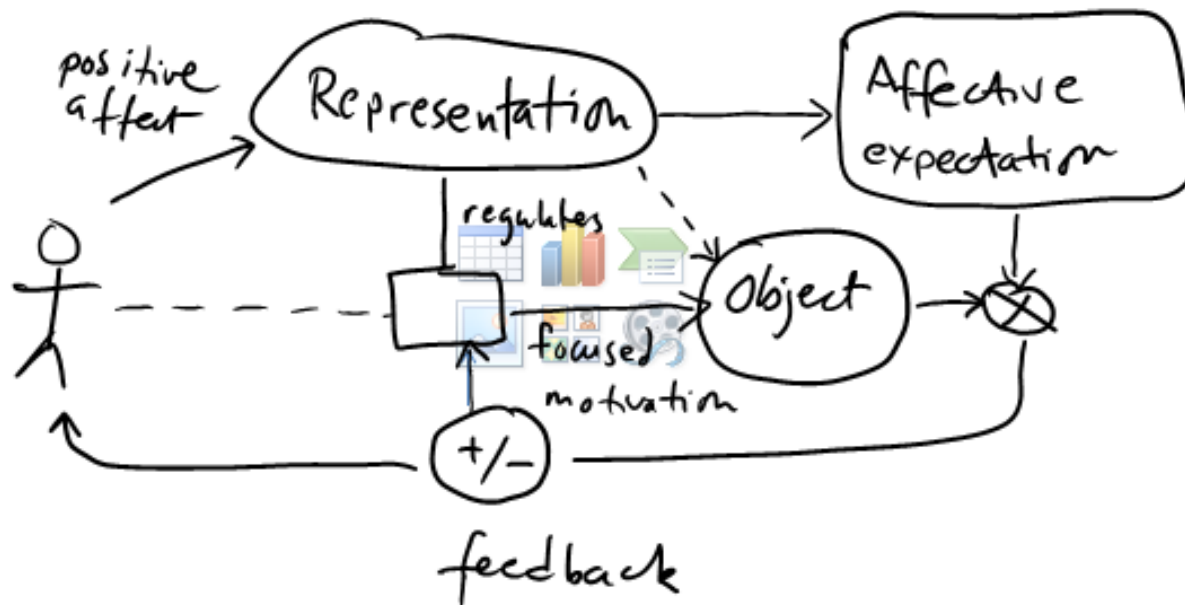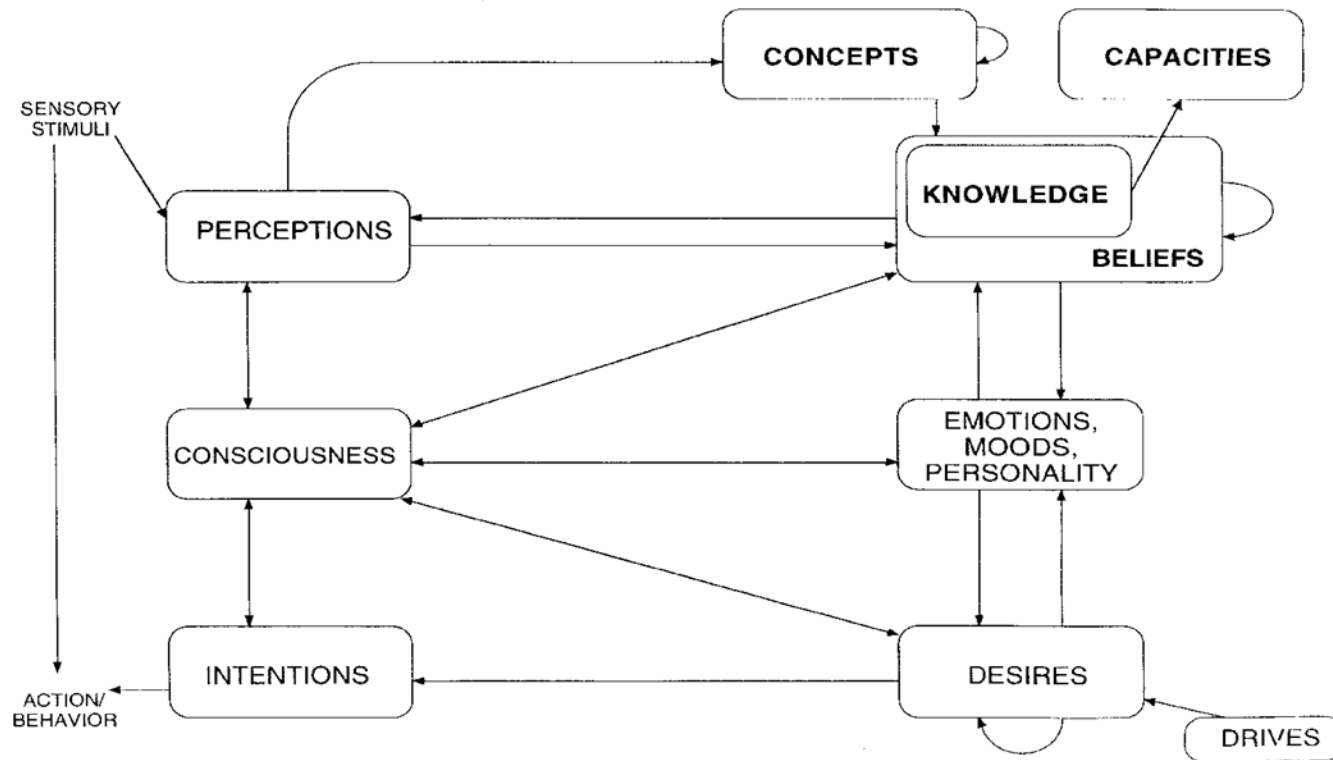
Image

Reality

Desire: job description

- (vii) **Disorders of the affective system manifest as disorders of desire.** Depression (deficit of positive affect) and mania (surplus) result in dysregulation of motivation, and failure of motivation to adapt to actual outcomes.
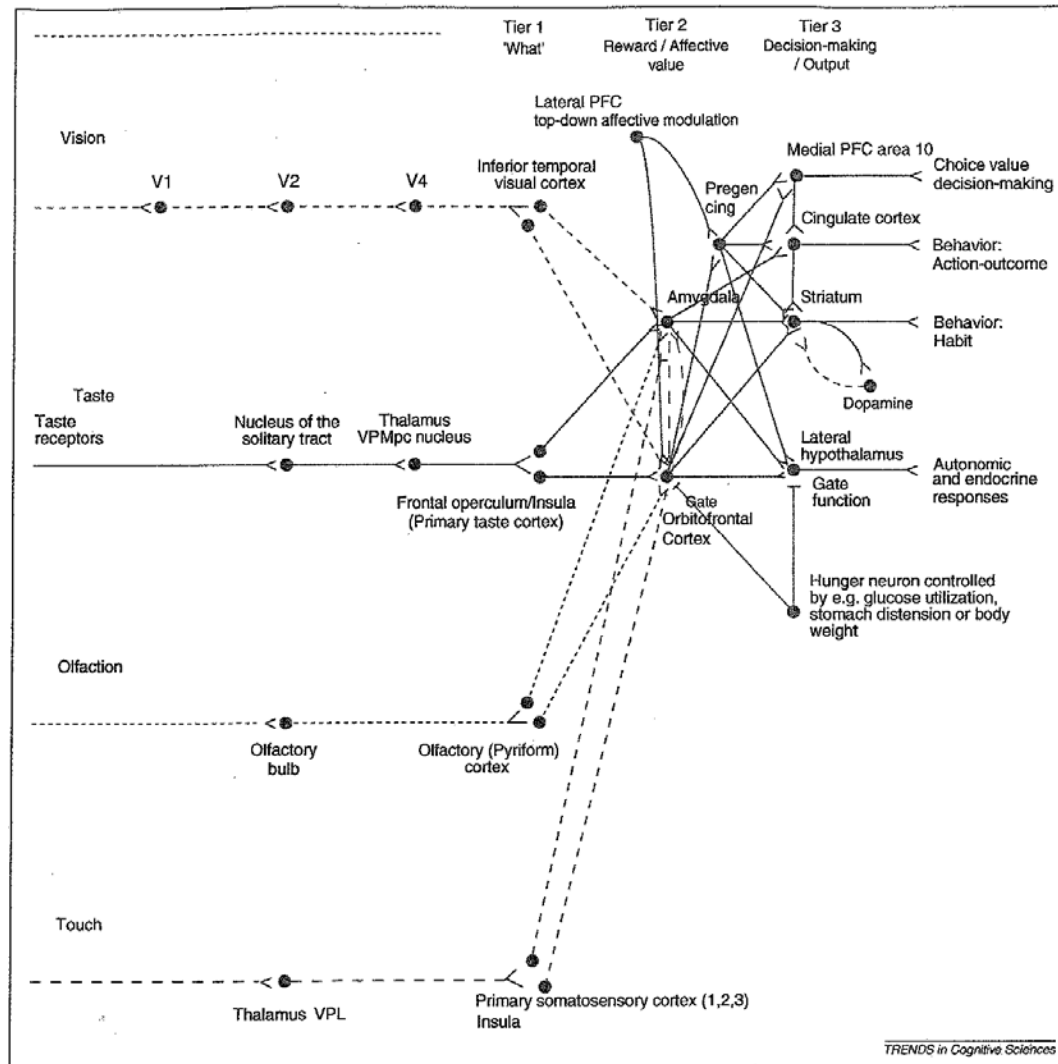
Warning:

NOT PHILOSOPHY

Could affect be so central?

Brain boxology—old school

Affective primacy—new school

(Grabenhorst & Rolls, 2011)

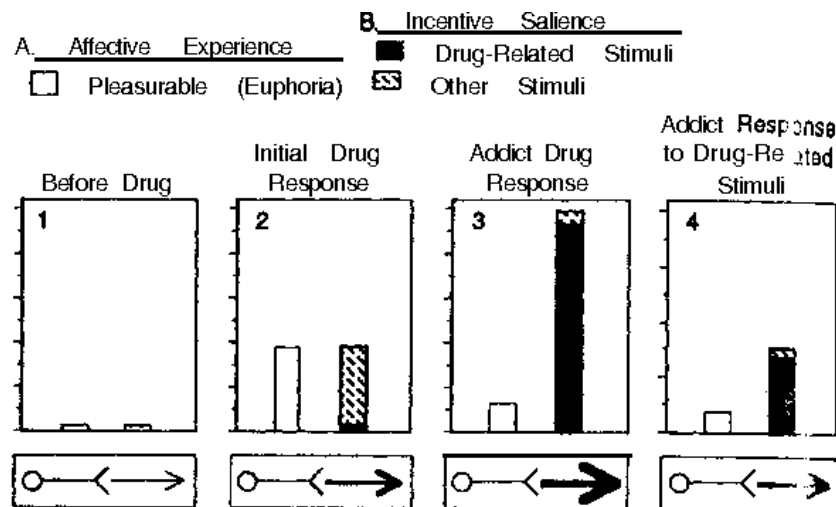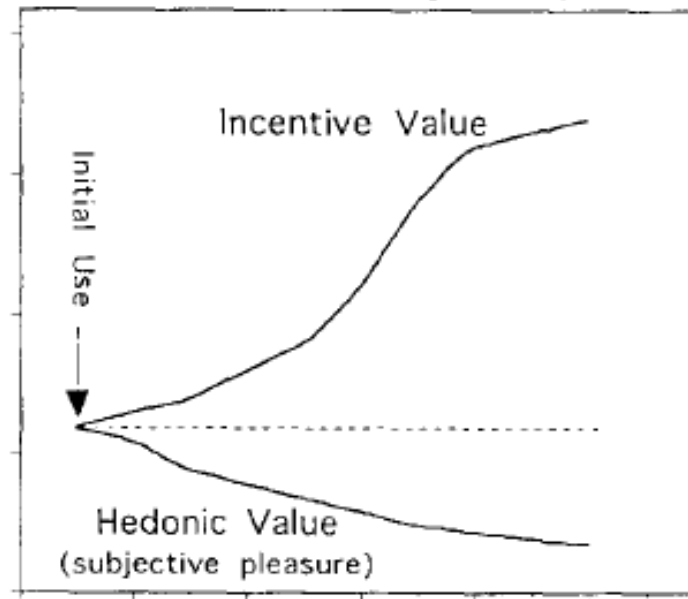# Are "liking" and "wanting" genuinely distinct?

Liking and wanting

Liking – a species of positive affective interest, attraction to a prospect; hedonic expectation

Wanting – a species of "incentive salience" that galvanizes attention, motivation, and action; not inherently hedonic

(Berridge, 2008)

Are these attitudes coupled in a regulative structure?

Sensitization of "wanting" but not "liking"
(Robinson & Berridge model)

Incentive Value

Initial Use

Hedonic Value
(subjective pleasure)

A. Affective Experience
☐ Pleasurable (Euphoria)

B. Incentive Salience
■ Drug-Related Stimuli
▨ Other Stimuli

Before Drug

Initial Drug Response

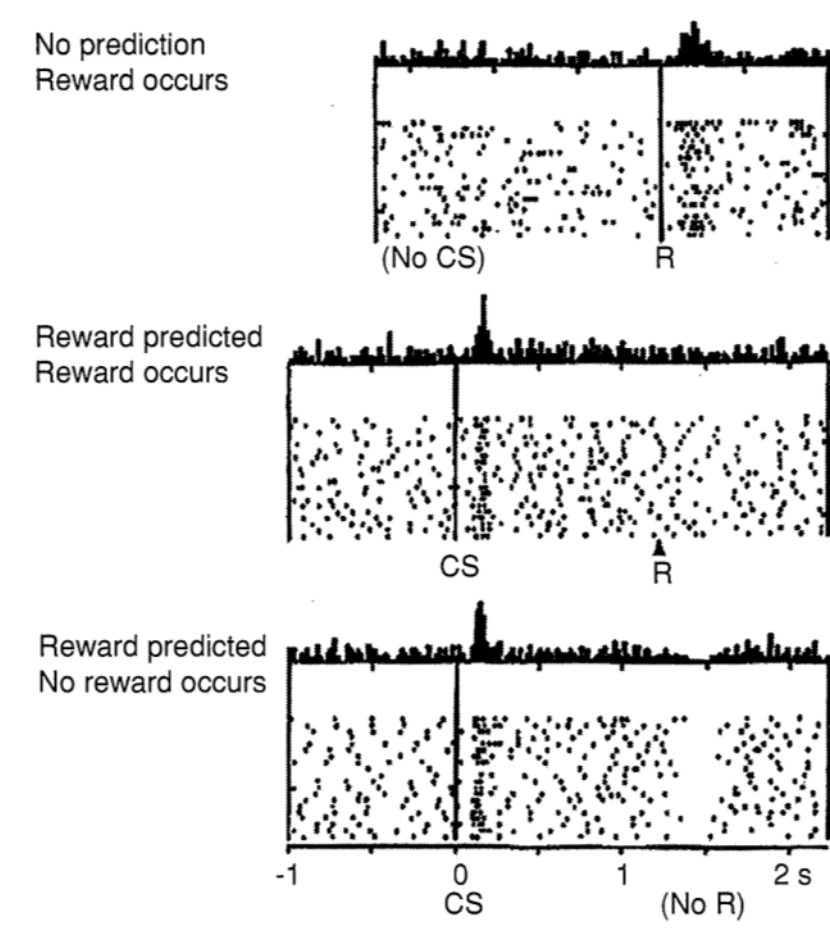Addict Drug Response

Addict Response to Drug-Related Stimuli

Does the affective system really represent, and learn, *value*?

- With magnitude, valence, ordering, cardinality, marginality, and discrepancy-detection, and choice-guidance?
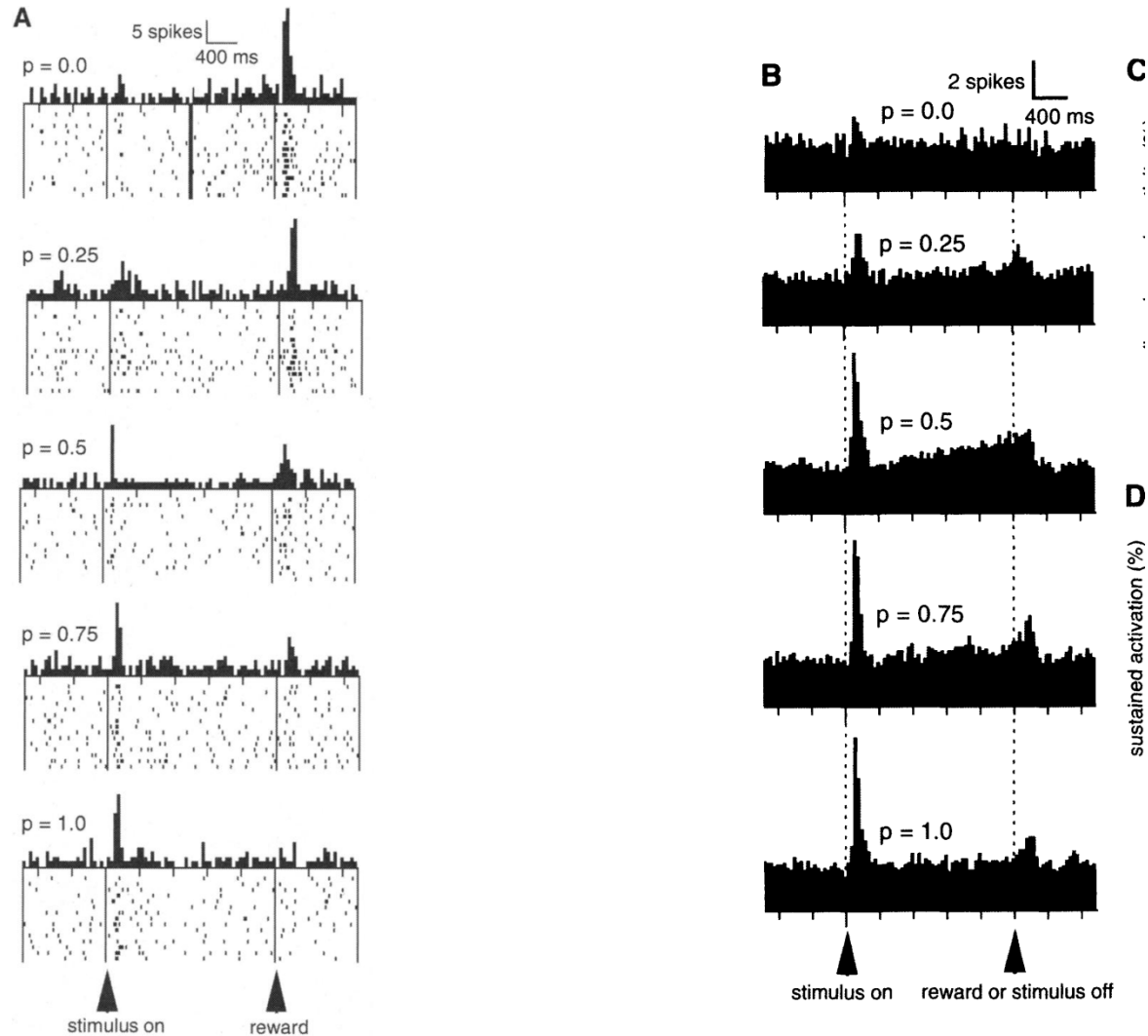
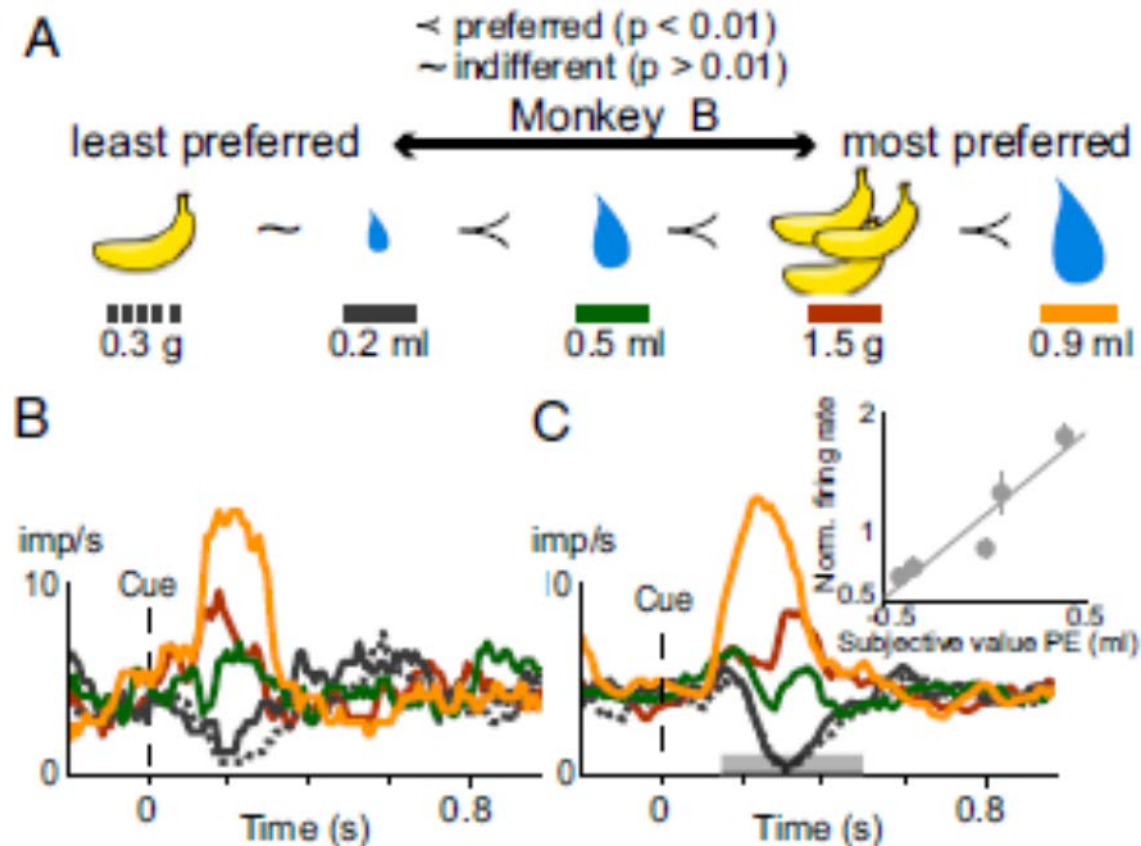# The neural substrate of reward-based learning

(Schultz *et al.*, 1997)

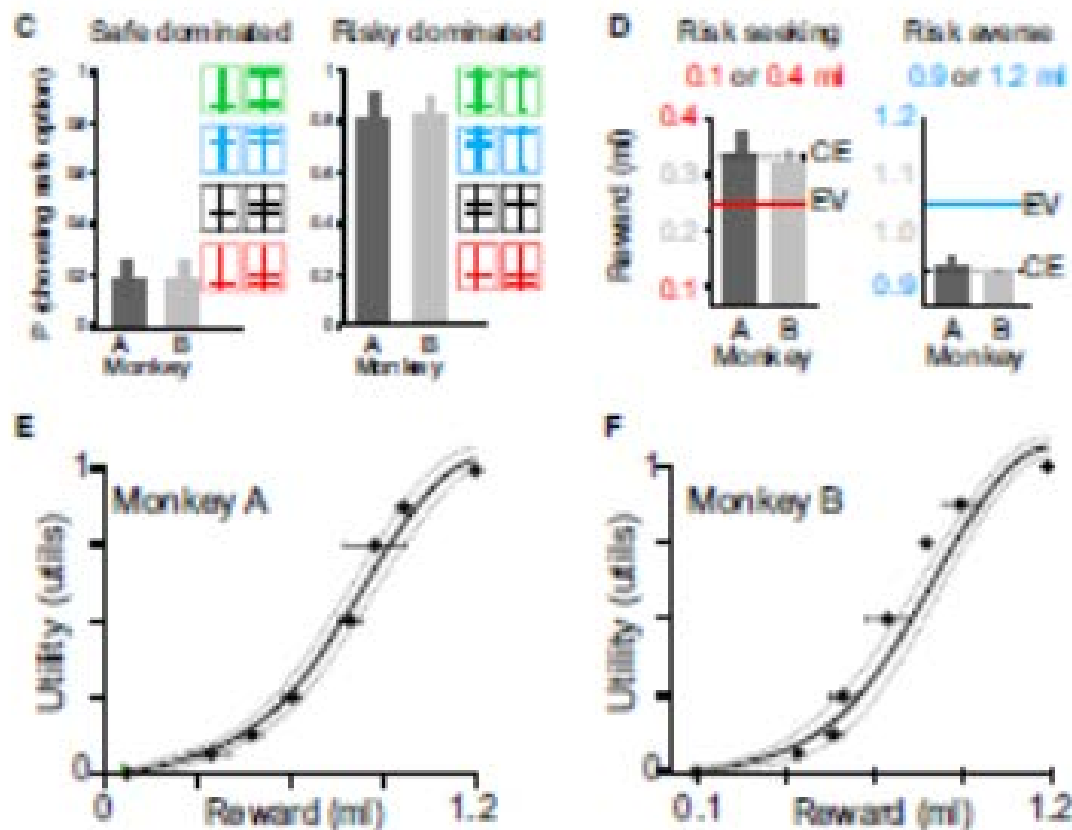# Representing expected value vs. risk
(Fiorillo *et al.*, 2003)

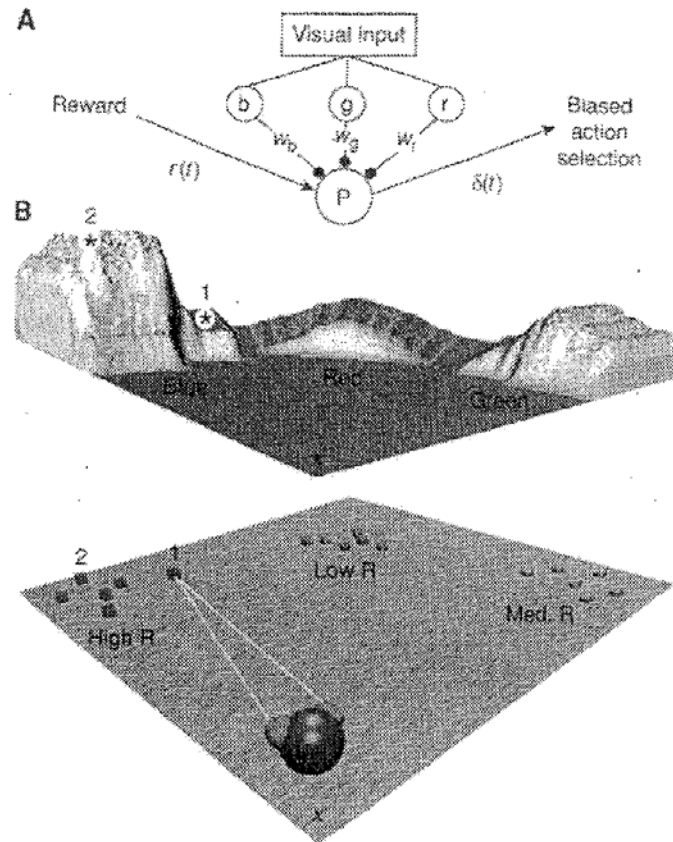# Encoded *preferences* among different rewards
(Lak *et al.*, 2014)

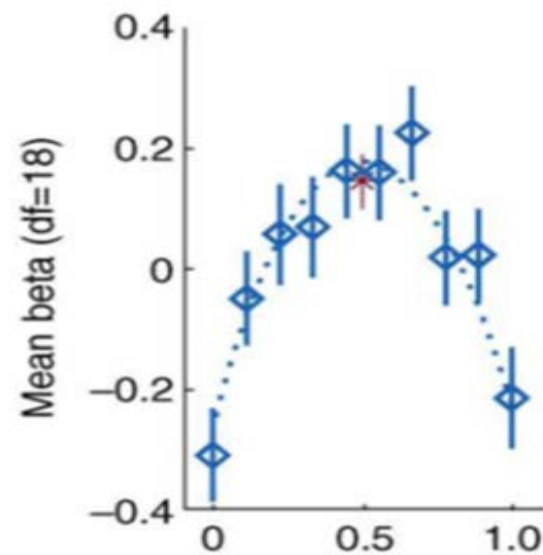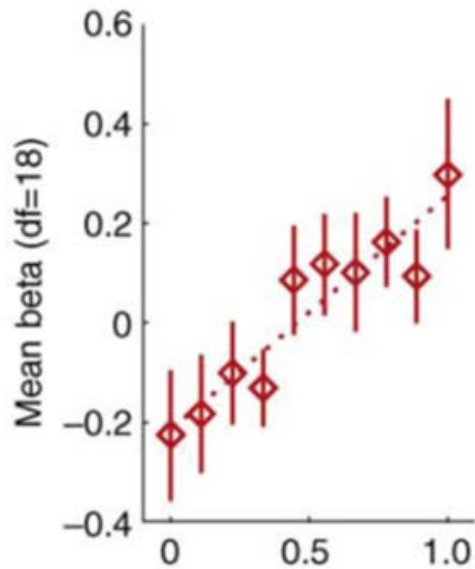# Constructing predictive utility functions from gambles
(Stauffer *et al.*, 2014)

"Value-coded" cognitive mapping – foraging for value
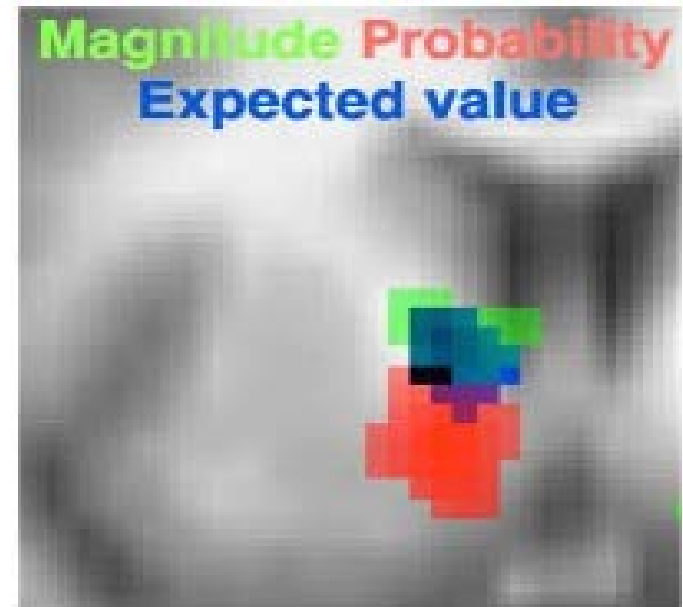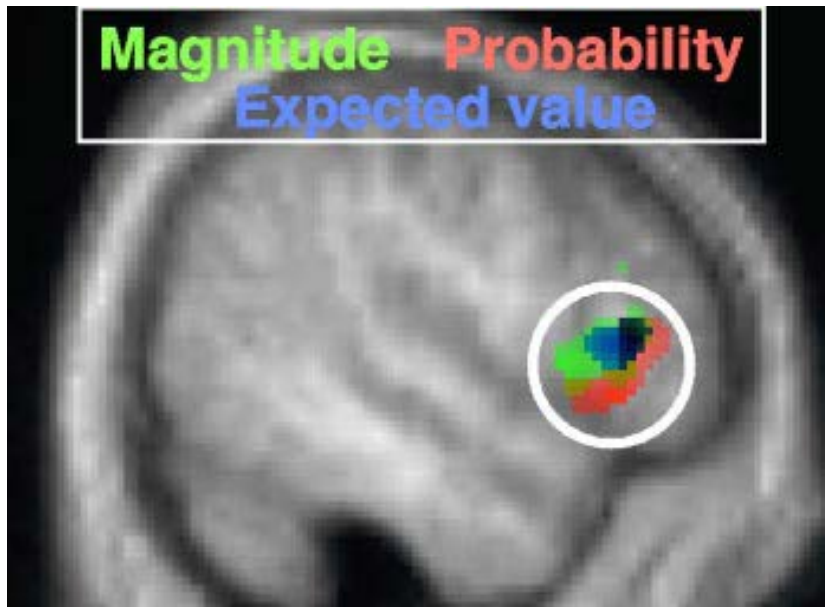(Schultz *et al.*, 1997)

Representation of expected value vs. risk—activation in the human ventral striatum (Quartz, 2009)

Coding for value, probability, and expected value: lateral prefrontal cortex and striatum
(Tobler *et al.*, 2007)

Back to philosophy

Desire as a fitting attitude with respect to value

- A positive affective representation alone could capture the category, valence, and magnitude, but would be no more than **wish** without active motivational force. **Desire** involves the allocation of attention and motivation, befitting a crucial part of *what matters* in value—*that it actually be sought and realized.*

- A directed **motivational force** alone would capture this idea of committing resources in response to as a recognition of value, but, without a positive affective representation of the object, would be an opaque **urge** or **drive** without rationalizing potential.

- By **coupling** these components of responsiveness to the reasons afforded by values, **desire** is a fitting attitude

Contrast *valuing attitudes* with *judgments of value*

- Similarly, desire can be a more fitting response to practical value than belief, since belief incorporates neither the affective evaluation nor the directed motivation.

- Desire thus constitutes a kind of *practical appreciation* of value while belief constitutes a *theoretical recognition* of value.

  – This helps us understand the debate over internalism.

  – The non-cognitivist is right that *mere belief* is not enough to capture genuinely practical appreciation of value.

  – The cognitivist is right, however, that the attitude that *does* capture such appreciation is *not* non-representational.

A more general view
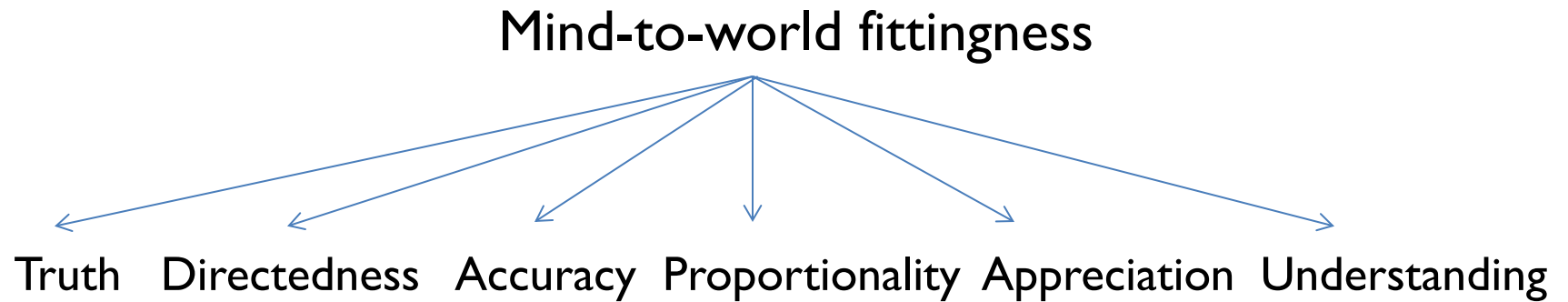
Mind-to-world fittingness

Truth   Directedness   Accuracy   Proportionality   Appreciation   Understanding

A shameless appeal to authority: Kant on the fitting attitude toward the moral law and others as ends

- In the *Metaphysics of Morals*, Kant distinguished a "merely theoretical" cognition of what our duty objectively requires— for example, through application of the categorical imperative and a *practical appreciation* of the moral law or others' as ends.

- Practical appreciation involves the "moral feeling", and distinguishes action from duty from merely legalistic action.

- This feeling involves "respect (*reverentia*)" for the moral law or for others as ends, "a feeling of a special kind, not a judgment" (*MM* 6:402).

- Without this feeling, we would be "morally dead" (6:400).

A shameless appeal to authority: Kant on the fitting attitude toward the moral law and others as ends

- This feeling is necessary to get duty underway and motivate it in action, without regress: "A duty to have respect would … amount to being put under an obligation to duties" (*MM* 6:402).

  - But this moral feeling must be a *desire*—it must elicit and motivate action in accord with the moral law or from respect for others, "not, to be sure, a mere wish, but the summoning of all means in our power" (*MM* 6:403; *G* 4:394).

What about the will?

- The will, for Kant, including the good will, belongs to the "faculty of desire", and:

- The will, for Kant, including the good will belongs to the "faculty of desire" (6:211).

- And:
  - "The *faculty of desire* is the faculty to be, by means of one's representations, the cause of the objects of these representations. The faculty of a being to act in accordance with its representations is called *life*." (6:211).

- For Kant as well, then, desire is a fitting attitude toward value—even the highest and unconditional value.

Actual Radioman, Craig Castaldo/Schwartz

- Not a tale of bare, blind urges, but of desire and learning.

Radioman's responsiveness to reasons

- Homeless and alcoholic in NYC, he kept a radio around his neck as a source of information and strength.

- Hung around movie sets from an interest in movies, became a regular "extra" homeless person in city scenes—gaining acceptance and recognition.

- Used this recognition to advocate for the homeless.

Radioman

- Argued that the homeless made an ecological contribution through their small carbon footprint.

- In photo, had bicycled 8 hours to attend a fund-raiser on homelessness.

- Desire kept him going.

- He was drawn to an idea, and pursued it committedly. He makes sense. His story could make a movie—and did.

Remember Louis B. Mayer

Cet obscur object du désir

But this is a Hollywood movie

~The End~