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

John Locke Lectures 2018
Dedicated to Derek Parfit (1942-2017)

Lecture 2:
Belief, Affect, and Reliance

Peter Railton
(University of Michigan)
Oxford, May 2018

Seminar discussion

- Tomorrow, Thursday, at 9:00 am

- Ryle Room

- Radcliffe Humanities Building (next door)

- All are welcome!

Some Big Issues

- Large questions in normative and meta-normative theory have hung on the nature of belief vs. desire, and their potential relation to causes and reasons for thought and action:
  - The possibility of truth or knowledge in ethics or practical reason (e.g., cognitivism vs. non-cognitivism)
  - The scope of practical reason ("internal reasons")
  - The nature of agency and free will (higher-order desires or evaluations)
  - The nature of well-being (desire-based theories)
  - The possibility of avoiding regress in theoretical or practical deliberation (non-deliberative dispositions)

The "standard model" of intentional action

- **belief** **+** **desire** **→** **action**

The "standard model" of intentional action

- **belief** + **desire** → **action**

desire_{PR} [*that PR has a cloak Tuesday*]

$\underline{\textbf{belief}_{PR}}$ [*that PR has a cloak Tuesday only if PR purchases one today*]

act_{PR} [*that PR purchases a cloak today*]

The "standard model" of intentional action

- **belief** + **desire** → **action**

desire$_{PR}$ **[**_that PR has a cloak Tuesday_**]**
belief$_{PR}$ **[**_that PR has a cloak Tuesday only if PR purchases one today_**]**
act$_{PR}$ **[**_that PR purchases a cloak today_**]**

But how do the belief and desire come together in this way to create "the beginnings of action"?

Bringing the agent in?

- **belief   +   desire   →   action**

Perhaps what's needed is to bring in the agent:

- Christine Korsgaard (1997): "… to will an end is not just to cause it, or even to allow an impulse in me to operate as its cause, but, so to speak, to **consciously pick up the reins, and make *myself* the cause of the end**."

- David Velleman (1992): "… [T]he agent's interactions with [his mental states] are such as they couldn't have with themselves. His role is to **intervene between reasons and intention, and between intention and bodily movements**."

Bringing the agent in?

- **belief   +   desire   →   action**

- John Searle (2003) similarly locates "**gaps**" in the belief-desire model that the agent must fill:

  - she must bring "deliberation" to a "conclusion" by **forming a "prior intention"**,

  - she must **decide when to enact** a prior intention—to make it into an "intention-in-action",

  - and because actions typically extend over time, she **must continue acting upon the "intention-in-action"** to complete the action sequence.

Such approaches share the problem

- … that they are invoking an agent taking an action or actions in order to explain how an agent takes an action.

  – We would want to know how *those* actions— "taking up the reins", "intervening among his mental states", forming and deciding to act upon an intention—could be performed as apt responses to reasons.

- We need a *non-agential* account of how the structure prerequisite for intentional agency—a structure that brings together what we care about and what we believe to organize possible actions and dispose us to perform them—so that the agent can, e.g., by seeing some new fact or having some new thought, act *intentionally* without deciding upon an intention.

Acting intentionally

- Or even, if I *do* consciously form an intention in a given instance, or in some sense "take up the reins" to intervene in my what I am now about to do,
  - … so that I can do *that* without needing to form an intention to form an intention, or to take up the reins to take up the reins. And so on.
- We can't appeal to prepackaged action patterns or fixed habit:
  - If picking up the reins were a mere mechanical response or habit, triggered by the situation, then *that* would not appear to be an exercise of practical intelligence, or an apt response to reasons, or a form of agency.

A possible solution

- Robert Stalnaker:
  - "Belief and desire … are correlative dispositional states of a potentially rational agent.
  - "To desire that p is to be disposed to act in ways that would tend to bring it about that p in a world in which one's beliefs, whatever they are, were true.
  - "To believe that p is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which p (together with one's other beliefs) were true." [(1984), 15]
- Having beliefs and desires is *already* having a set of organized, representationally-mediated action tendencies.

Last time …

Last time: desire as representationally-mediated, regulative, action-guiding, and adaptive

↑

Renault 14

www.renault-14.net

Une poire c'est confortable.

Renault 14

1418.net

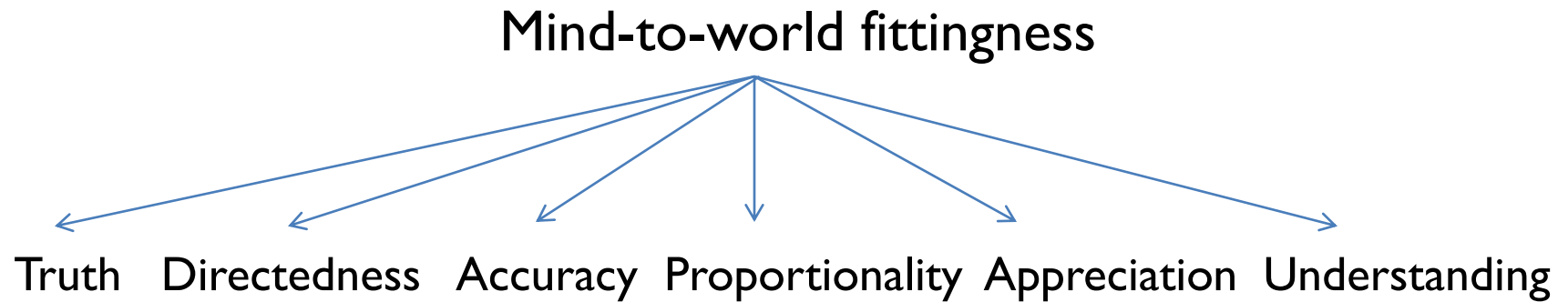**Last time: desire as representationally-mediated, regulative, action-guiding, and adaptive**

↑

"Merely causal"?

- While this is a causal process, it is not "merely causal".
- The affective attitude sustains evaluative expectations about how the world will be,
- … and it elicits and regulates focused motivation on behalf of this realizing these expectations.
  - That is, it organizes an available action-tendency "under an idea" that makes the behavior intelligible as a response to an apparent reason, and holding the idea "in view" to monitor behavior and outcomes.
  - It has mind-to-world, not just world-to-mind direction of fit, is capable of greater or lesser accuracy, and can learn.

In these respects …

- … desire that *p* constitutes an *intelligent* disposition, such that, for example, the agent, upon seeing an opportunity, or having a thought, or hearing another's remark, or … can respond fluently in intelligent, intentional, reasons-sensitive ways.
  - In its regulative structure, it incorporates the model of a hierarchical, evaluation-sensitive motivational structure Frankfurt and Watson identified as constitutive of agency.

A more general view

Mind-to-world fittingness

Truth   Directedness   Accuracy   Proportionality   Appreciation   Understanding

But desire cannot act alone

- It requires a representation of the world's relevant possibilities for action.

- Stalnaker, again:

- "To believe that p is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which p (together with one's other beliefs) were true." [(1984), 15]

- Belief, too, would seem to be an intelligent disposition, capable of organizing behavior under an idea, and providing guidance accordingly.

  – What might this structure look like?

"What I can't build, I don't understand."
(attributed to Richard Feynman)

- As before, our method will be to assemble some field notes—how desire or belief seems to operate, how it is spoken of (including cases where it is called "irrational"), etc.
- We think of these notes, combined with truisms, paradigm cases, causal roles, explanatory desiderata, etc. as a "job description" for an account of desire, belief, etc. to fill.
  - We then ask how we could build a state or process from simpler ingredients that could satisfy that job description, at least to a reasonably high degree,
  - … and might allow us to make sense of how that state might enable apt responsiveness to reasons

A field guide: some obvious features of belief

- **(i) Representational:** Presents the world to us through some form of representation:

    - **attitude [representation]**

    – … and this representation *mediates* the role or effect of belief for thought and action. Lois Lane will not be surprised to see bullets bounce off Superman, but she will in the case of Clark Kent.

- **(ii) Truth or falsehood:** We can believe *propositions*:

    - **belief [that *p*]**

    – … in such cases belief represents *p* as true but "adds no new idea" to it, and so is deemed true iff its content is true.

A field guide: some obvious features of belief

- These two features are, however, insufficient to distinguish **belief [that *p*]** from:
  - **supposition [that *p*]**
  - **thought [that *p*]**
  - **imagine [that *p*]**
- … since all of these attitudes represent their content as true, add no new content, and are spoken of as true iff *p* is true.

A field guide: some obvious features of belief

- So, to distinguish belief:
- **(iii) Mind-to-world direction of fit:**
    - **belief [that _p_]**
- 'Fitting' does not have a universally-agreed significance, but at least one sense is to ask whether there is anything inherently _unfitting_ about, for example, the mental state of someone who _supposes that p_ or _imagines that p,_ even while taking herself to have conclusive evidence that _p_ is false. No.
    - With belief it is otherwise—evidence that _p_ is false tends to undermine belief, and believing that _p_ in the face of what one takes to be conclusive evidence that _p_ is false certain seems to be unfitting in some sense. Assertion.

A field guide: some obvious features of belief

- **(iv) Spontaneously action-guiding:** Belief *spontaneously* shapes dispositions to act, even without agential intervention. If one believes that *p* one spontaneously *relies* upon *p* in choosing, and directly immediately shapes one's dispositions to act.

- In contrast, supposition or pretense must be kept in mind if they are to be consistently applied in thought and action. "Oh, what tangled webs we weave …".

A field guide: some obvious features of belief

- **(v) Degrees of strength:** Although we often speak of outright belief and disbelief, we also speak of beliefs as more or less confident, and this degree of confidence tends to vary spontaneously in response to congruent or incongruent experience. Crawling out on a limb.

- Moreover, these varying degrees of confidence spontaneously *modulate* reliance upon belief in action, and shape how beliefs and desires combine to form action-tendencies.

  - "Degree of belief" has **two forms**: one can have a higher or lower *expectation* of some event or outcome of action, and this expectation value can be more or less *confident.* These shape dispositions to act in different ways.

A field guide: some obvious features of belief

- **(vi) Spontaneously thought-guiding:** Belief spontaneously shapes a great deal of our mental economy.
    - *attention*—what we (do or don't) spontaneously notice
    - *perception*—how we spontaneously interpret experience
    - *memory*—what recollections occur spontaneously to us
    - *inference*—what mental transitions we are spontaneously disposed to make and rely upon
    - *expectation*—what we spontaneously anticipate
    - *decision-making*—what we spontaneously take to be given, and how much weight various factors receive
- *That these can occur without intermediation is crucial to avoid regress*

A field guide: some obvious features of belief

- **(vii) Spontaneously feeling-guiding:** Beliefs and changes in belief or belief strength spontaneously affects our "feelings", and this in turn is part of how beliefs contribute spontaneously to the guidance of thought and action:
  - *emotions*—fear, anger, surprise, disappointment, frustration, joy, etc.
  - *sentiments*—resentment, guilt, embarrassment, admiration, envy, interest, conviction, anxiety, suspicion, etc.
  - *moods*—sadness, happiness, anxiety, calm, etc.

A field guide: some obvious features of belief

- **(viii) Implicit or unconscious as well as conscious:** Belief's spontaneous roles do not depend upon whether a given belief is consciously represented or known. When asked, one often seems to "know immediately" whether one believes something or not, but one can be mistaken in a variety of ways about the content or strength of one's beliefs.

- Sometimes one must use inference to explore what one believes, or notices what one believes by what one finds surprising or disturbing. Such experiences can be at odds with what one consciously takes oneself to believe. Being deceived about what one believes is possible.

A field guide: some obvious features of belief

- **(ix) Phenomenologically thin:** Typically, belief has a thin phenomenology—unlike perceptions or aroused emotions. Belief is a "frame" through which we think—a "default" state of confidence about what to expect or what to rely upon in action.

- However, belief's phenomenology becomes more evident during *changes* in belief, e.g., losing confidence or acquiring assurance.

A field guide: some obvious features of belief

- **(x) Belief is spontaneously projective and evidence-responsive:** Given a perception as of p, e.g., we tend "immediately" to believe that p, and to rely upon p as we go forward in time. Belief is typically projective and "inertial" rather than conservative with respect to evidence—and it could hardly play an action-guiding role without this.

- Belief, as opposed to suspension of belief, is the default state. As Carnap argued (1950), and formal learning theory and Bayesian approaches to rationality have since developed, such "priors", projected forward, drive learning, sustaining expectations that shape information-seeking and –reception, and turn experience into experimentation.

Wait—isn't this just wrong?

- Don't people show strong "confirmation bias", ignoring negative evidence, and making regular, elementary errors in relying upon probabilities—unless these are corrected by conscious reasoning (Kahneman & Tversky, 2003)?

Warning:

NOT PHILOSOPHY

Are we terrible at intuitive statistics?

- People *do* deploy the kinds of heuristics, and make the kinds of errors, diagnosed by Kahneman and Tversky when given probability problems as artificial word problems,
  - … but given these same tasks against a background of naturalistic statistics, or in the form of interactive tasks,
  - … they do not make the same errors (Hau *et al.*, 2010; Pleskac & Hertwig, 2014)

Really?

- Human infants monitor frequency distributions in overheard speech even in the first months of life:
  - manifesting selective attention that reflects the conditional probabilities present in natural language, and paying greater attention to anomalous sequences (Aslin *et al.*, 1998; Fiser & Aslin, 2001; Kidd *et al.*, 2012).
- Infants observe *third-person* adult behavior to adjust their attention, learning rate, and behavior, using evidence of adult efficacy or reliability (Koenig & Echols, 2003; Wellman 2014).

Really?

- Infants pay attention to base rates in samples in expressions of surprise (Schulz, 2014)

- Bayesian models predict a range of features of the development of causal cognition and theory of mind (Gopnik & Wellman, 2012; Wellman, 2014; Hamlin *et al.*, 2013).

- Adult perception of changes in underlying frequencies in a stream of events approximates ideal probabilistic inference (Gallistel *et al.*, 2014).

- Given complex simulated foraging tasks involving multiple forms of value and risk, humans develop near-optimal foraging strategies through experience (Kolling *et al.*, 2012).

Recall: the neural substrate of reward-based learning

(Schultz *et al.*, 1997)
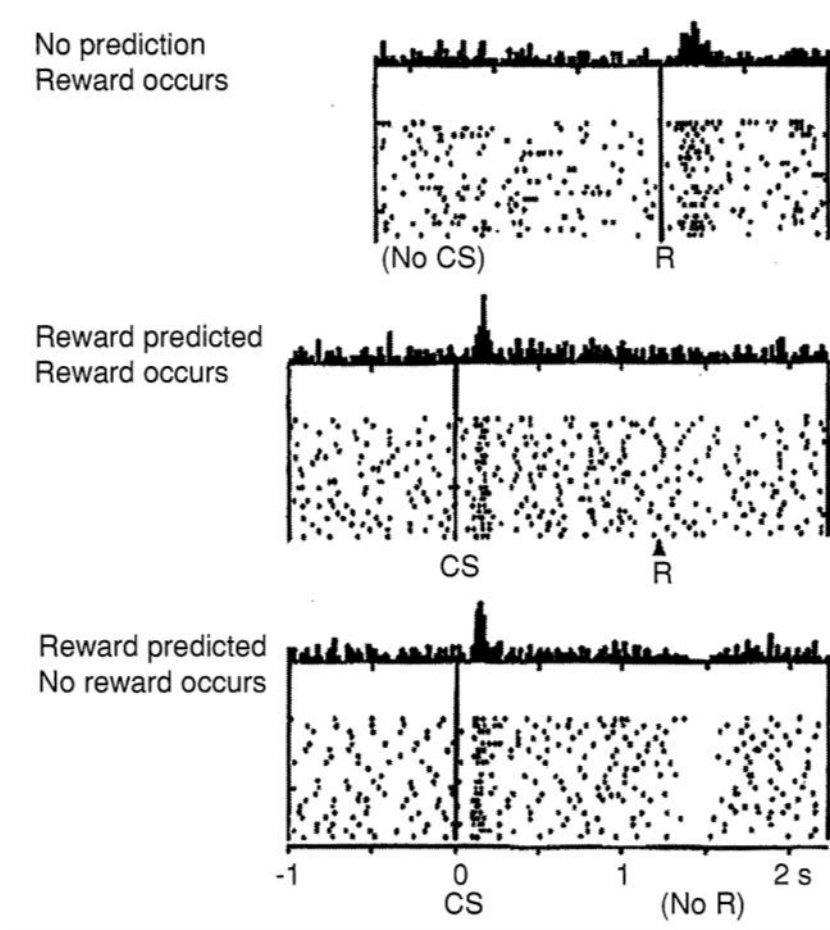
Recall: representation of expected value vs. risk

(Fiorillo *et al.*, 2003)
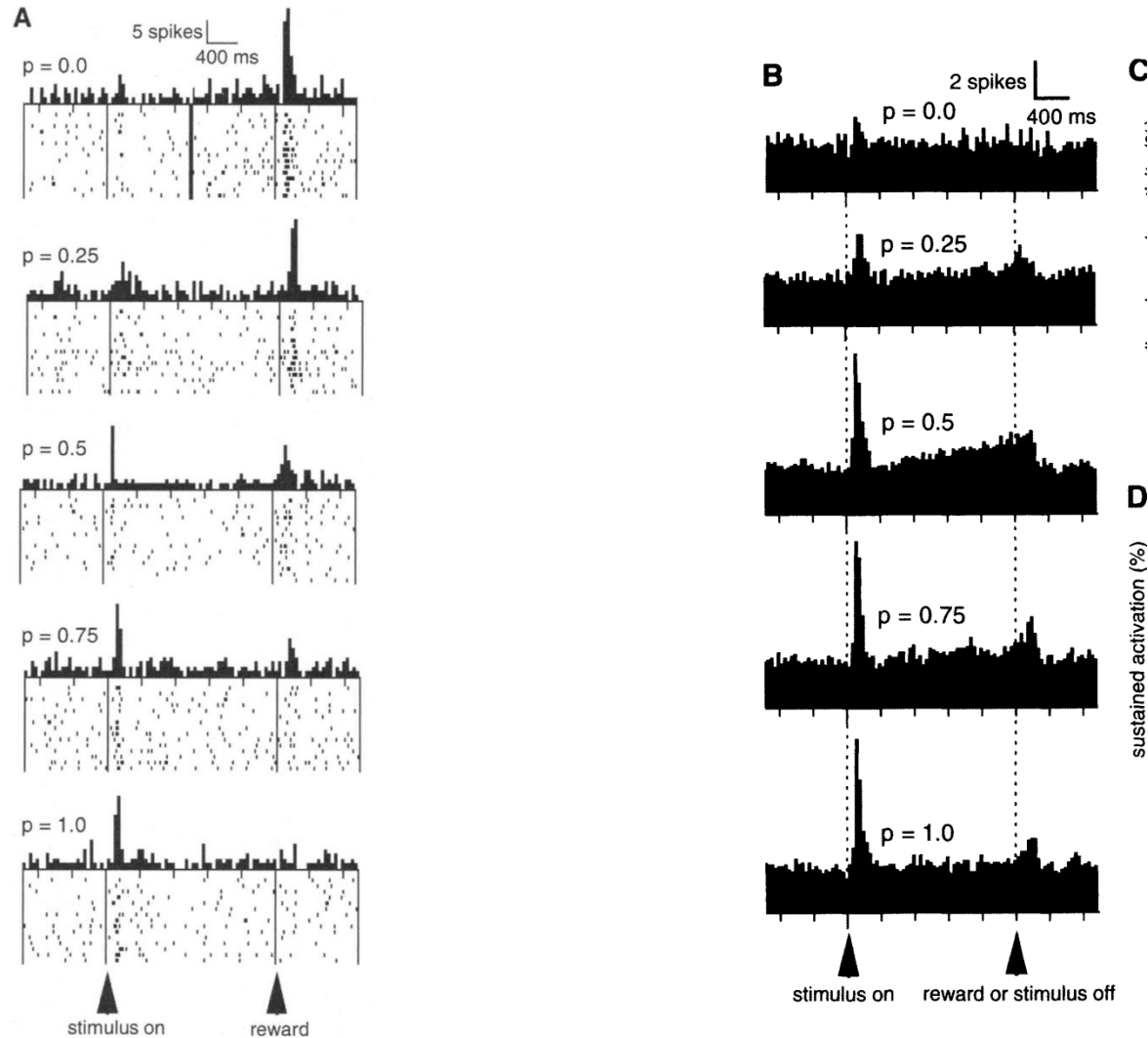
Representation of expected value vs. risk—activation in the human ventral striatum (Quartz, 2009)
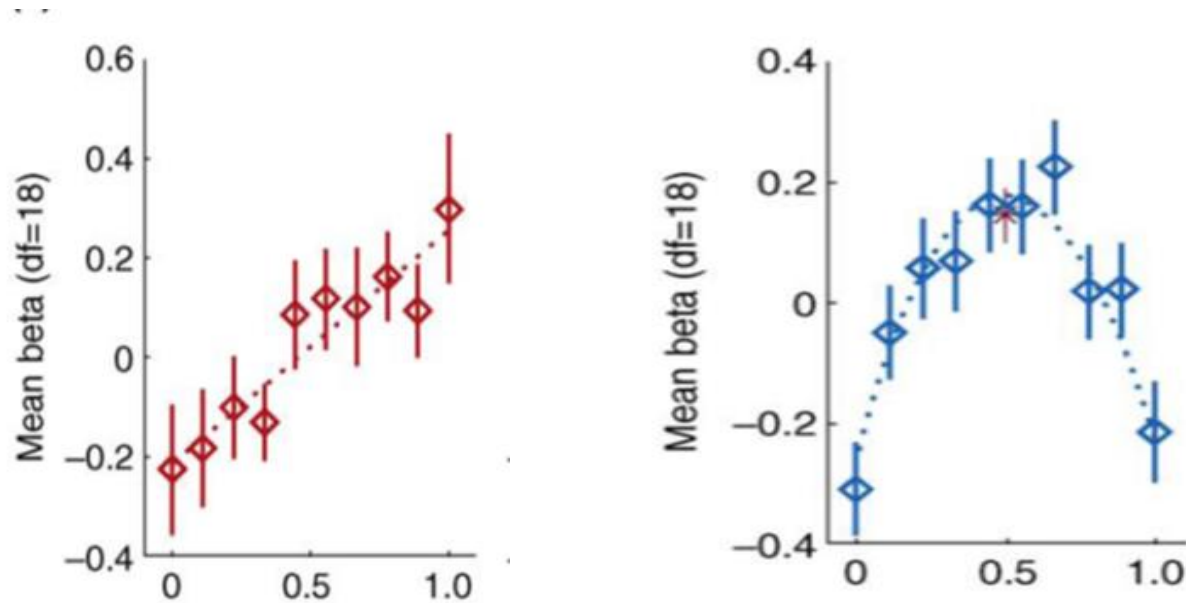
Back to philosophy

- … with a few parentheses

A field guide: some obvious features of belief

- **(xii) Relational and intensional:** Belief can take multiple objects—one can believe *that* a certain proposition is true, but also "believe one's eyes", "believe a person", "believe in oneself", "believe in the scientific method", etc. Belief is able to *shift* modality—e.g., from believing one's eyes to believing the content of one's perceptual experience.

- Belief can relate one to its object *de re* or *de se* as well as *de dicto*. Intensionality directly influences the spontaneous action-, thought-, and feeling-guidance of belief. A lost camper who trusts an air-dropped map and believes that the lost camper is just north of a stream might fail to believe this *of himself*, and so set out to "find" the lost camper.

A field guide: some obvious features of belief

- **(xiii) Non-volitional:** We can exert voluntary influence on our thinking in many ways—by shaping what experiences we have, what we attend to, what inferences we carry out, what we try to recall, etc. But we cannot simply *decide* to believe or believe at will. One can *accept* a given proposition for certain purposes, even accept it as best supported by the evidence, and use it deliberately in thinking about how to act, etc., but this does not amount to *belief* or yield belief's *spontaneous* roles. E.g., accepting my son's story; nervous flyers.

A field guide: some obvious features of belief

- **(xiv) Spontaneously resistant to instrumentalization:** Relatedly, we cannot decide to hold beliefs for *good* reasons of a purely practical kind—e.g., in order to improve our immune system, to lighten our mood, or fit in with our peers. To be sure, pragmatic considerations do exert considerable influence on what we believe—often we will *end up* believing what benefits us in various ways, but believing typically cannot be done *for* such purposes, in contrast to, e.g., *supposing, imagining, or accepting.*

- *Something* about belief seems to prevent direct instrumentalization *without* our policing this—but what?

I hope this is enough …

- … to convince you that belief is a complicated animal, with a diverse yet structured and integrated array of functions. But our field notes might also draw a bit more from the popular culture of belief.

Do features (i)-(xiv) sound familiar?

- They parallel closely the features we found in our field notes on desire.

- We might try:

Belief?

Confidence or trust

- These are also terms we use for *affective* attitudes.

- Might they be as they figure in belief? And might that help explain the obvious features of belief in a unified way—the way the role of affect in desire helped explain the obvious features of desire in a unified way?

- *Well, what's in a name? Why call it—or not call it—affect?*

- First, the neural evidence we looked at was of the affect and reward system, with direct projections to action-orienting behaviors. We're looking for a *psychological kind that might have been evolved*, and these are evolutionarily conserved.

- Second, what does psychology takes affective attitudes to be?

Affect or emotion
(with modification, from Frijda, 2007)
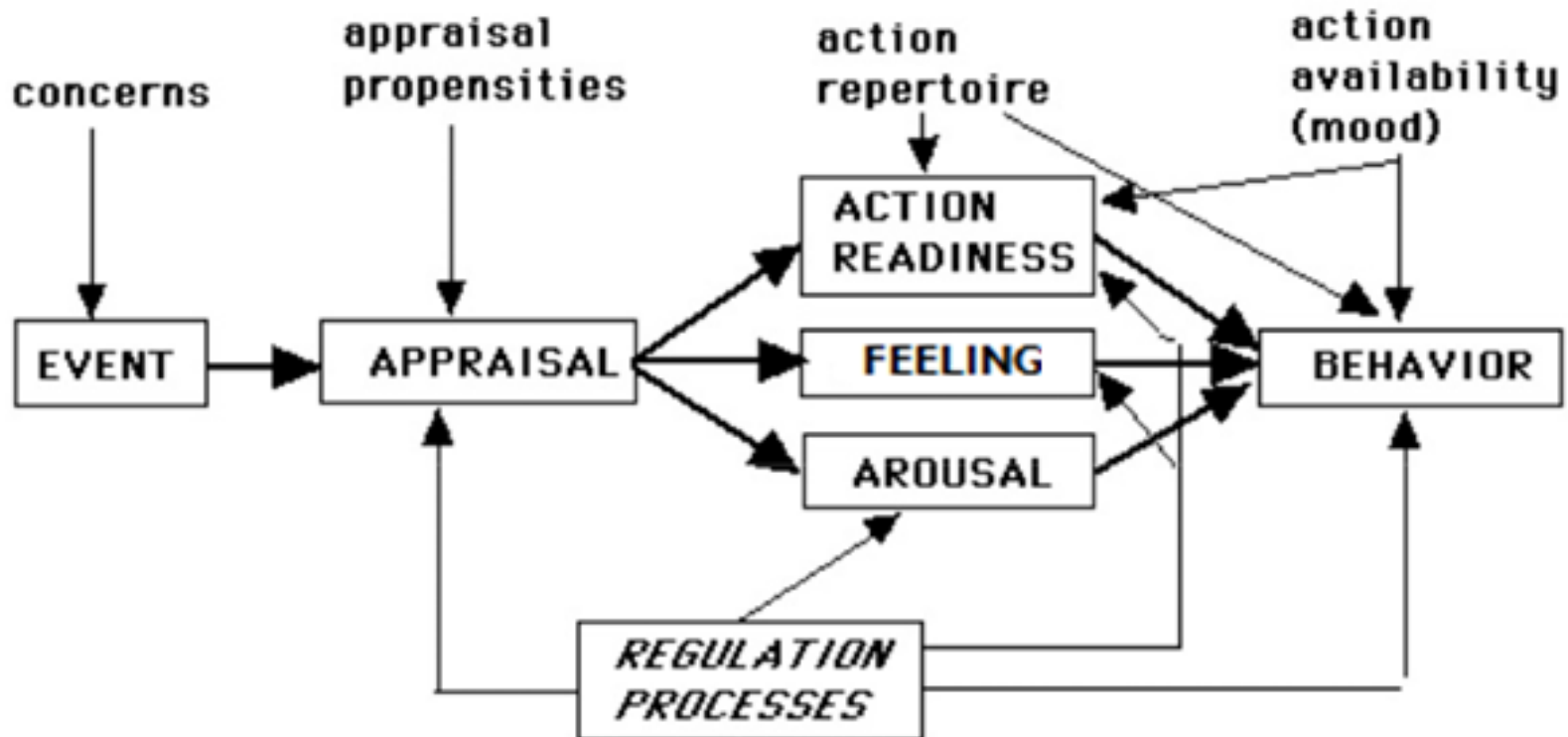
The marginality of affect—old school
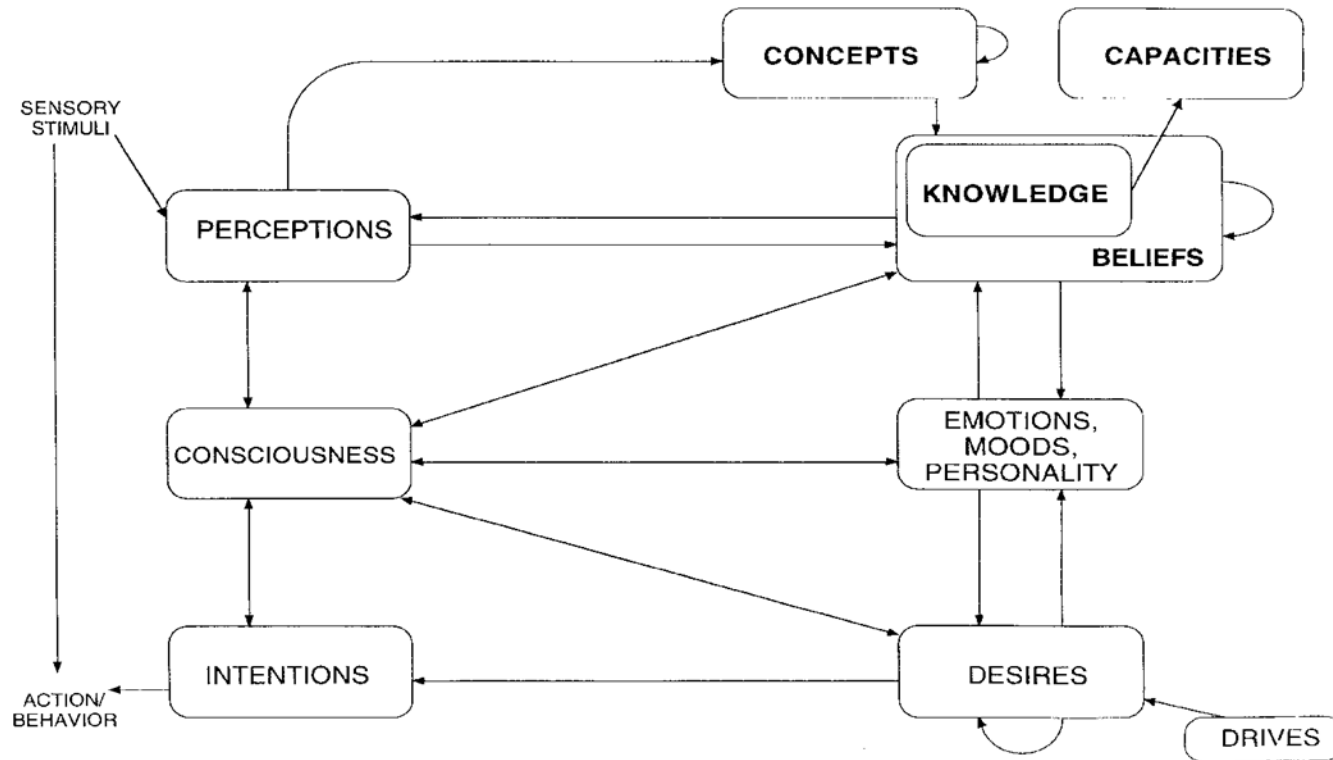
The centrality of affect– new school
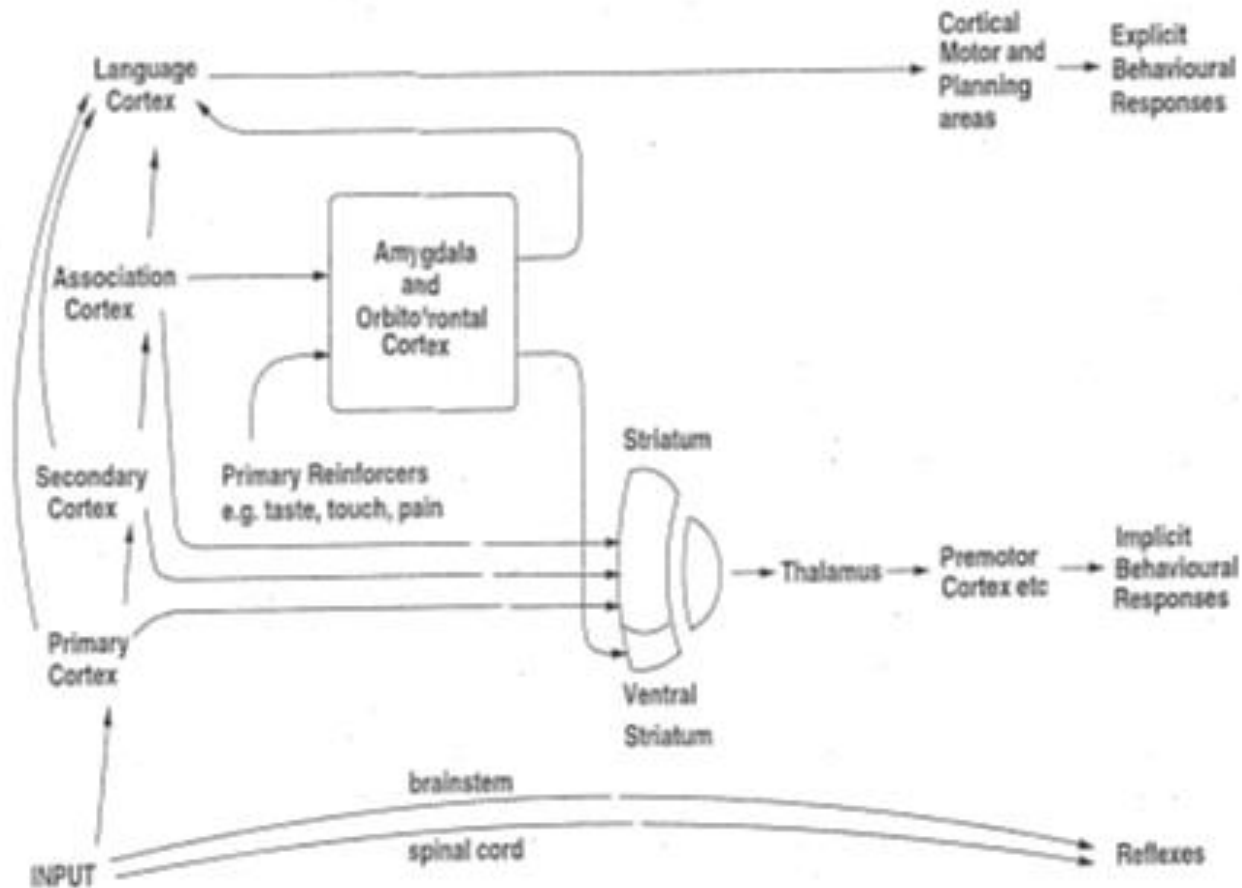
Neural processing for valuation and decision
(Grabenhorst & Rolls, 2011)

Brain connectivity graph, amygdala
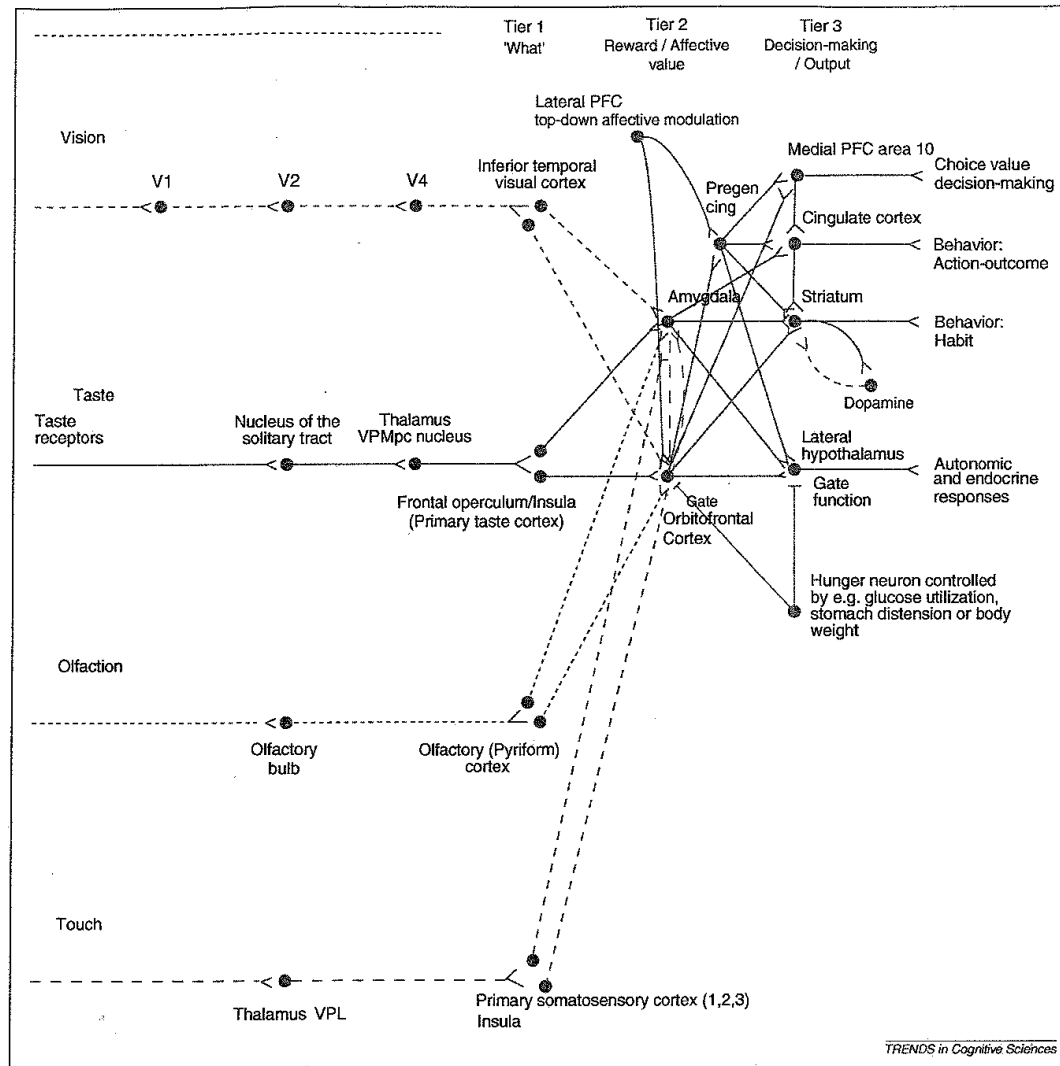
(Pessoa, 2008)
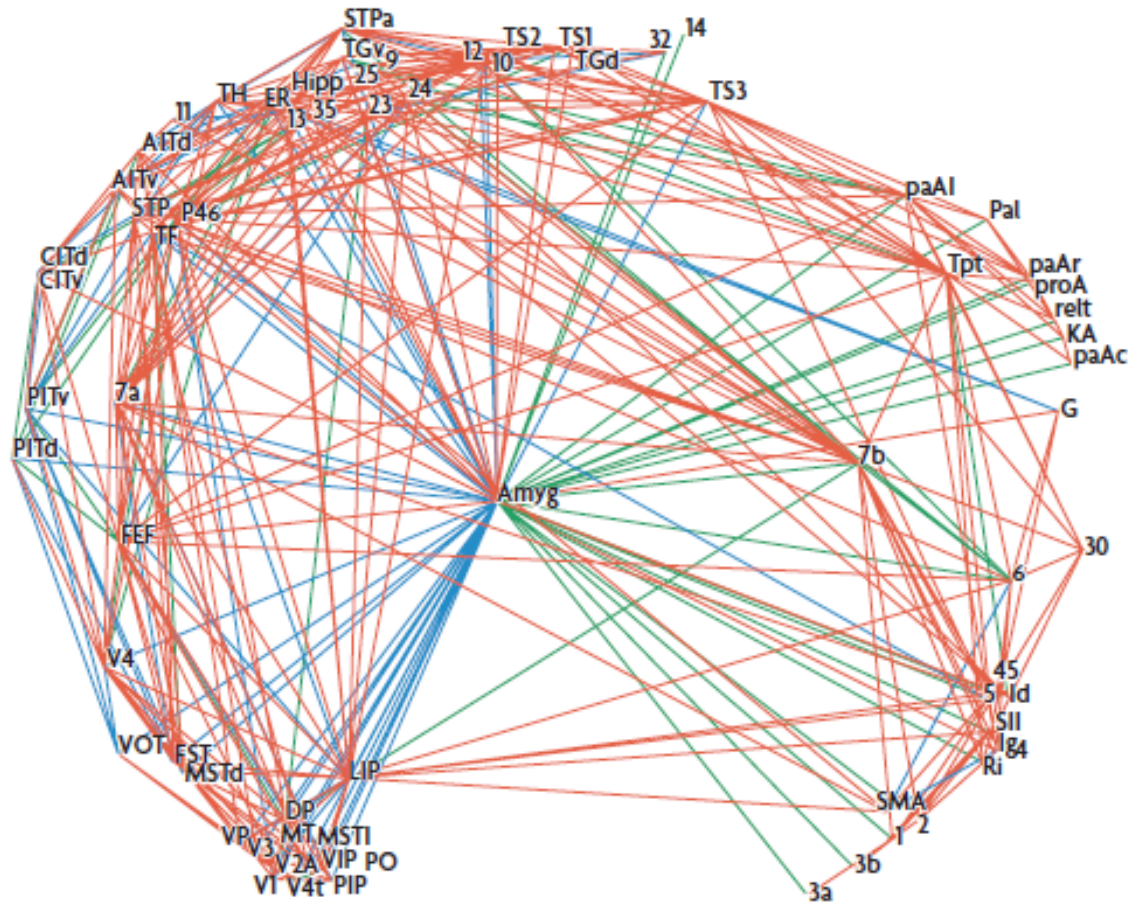
Affect and evaluation

- Affect is central because *evaluation—or risk or reward—is central.* Indeed the affect/cognition distinction has been widely rejected in contemporary psychology and neuroscience.
- Like other forms of affect, e.g., fear brought about or removed *at will* or *merely instrumentalized.*
  - Confidence or trust vs. acceptance.
- Like other forms of affect, confidence and trust respond to changes in experience spontaneously.
  - Like fear, they vary in strength and spontaneously tend to *calibrate,* and they differ from *judgment,* e.g., that *p* is dangerous, or reliable, or trustworthy.

Relations and intensionality

- Confidence and trust can take a variety of *objects*, and *transfer* reliance from one to another

  – E.g., confidence in a *process* can translate directly into confident reliance upon its propositional *conclusions*.

- Confidence and trust can also be *intensional* attitudes.

  – Just as the lost camper can believe the lost camper is near a stream, but not believe this *of himself*, he can be said to be confident of the map and its indication of the lost camper's location, but not confident of his own location.

Two kinds of strength of belief

- In the compound model:
  - … we see two different kinds of *strength* of belief in an outcome—degree of confidence and magnitude of expectant reliance—which can vary independently,
  - … and which have the same effects upon behavior (e.g., betting odds vs. willingness to bet) as the two different kinds of strength of belief.
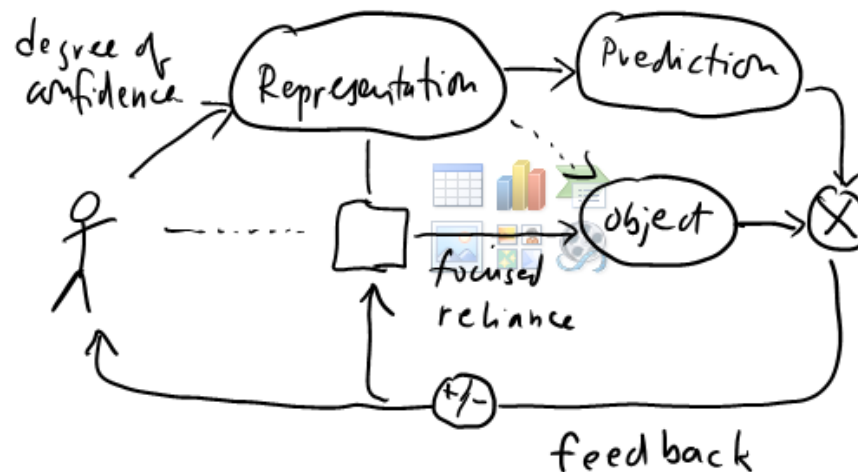
Projection, learning, and phenomenology

- Like belief (or credence in formal models of learning and decision-making), confidence and trust have a *default, inertial, projective* character.

  - In developmental psychology, confidence and trust play foundational roles in learning akin to the role of "priors" in formal learning theory.

- While aroused affective attitudes like fear or surprise have a distinctive phenomenology, calling our attention away from "business as usual", default affective attitudes like confidence and trust have a *non-distracting* phenomenology—except perhaps for when they undergo important changes.

Implicit as well as conscious

- While affect is typically associated with "feelings", most affective attitudes like fear, confidence, or trust can be implicit and non-conscious. They can respond to experience, gain or lose strength, shape attention and perception, shape or color thought, elicit motivation, and adjust behavior adaptively, without obtruding into conscious awareness.

Some dysfunctions or irrationalities in belief

- (a) Phobias and **epistemic** *akrasia*
- (b) "Affective transfer" and **wishful thinking**
- (c) **Affective disorders** disrupt the usual regulative structre and lead to dramatic and characteristic changes in belief.
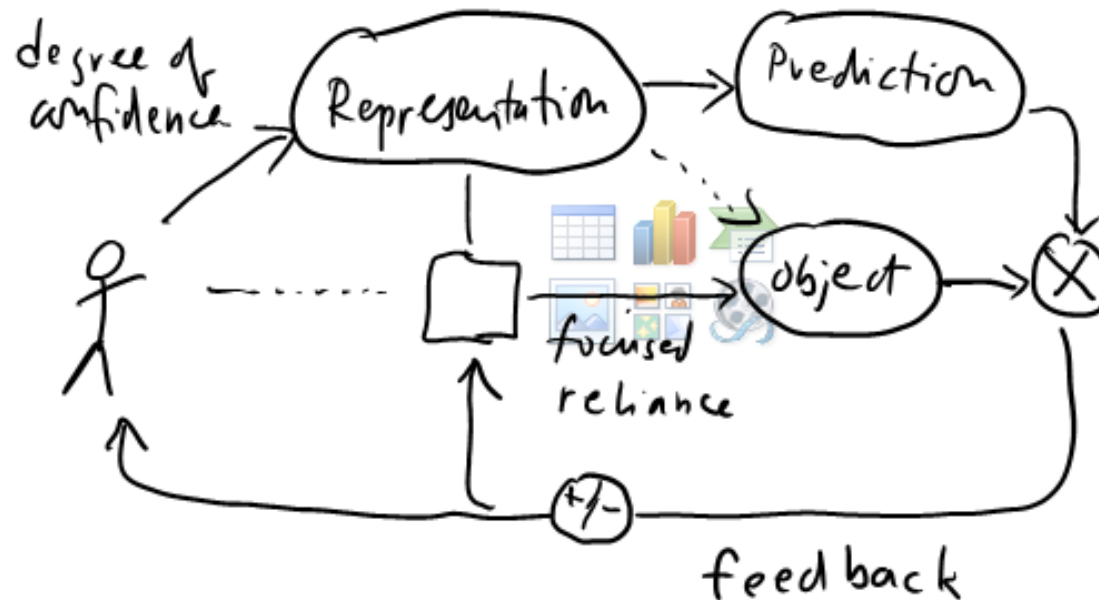
Working together with desire to structure capacities for acting intentionally

- Owing to their parallel structure, and "common pathways" through affect, expectation, and the eliciting and focusing of action tendencies, belief and desire (on the compound model) can work together to satisfy Stalnaker's characterization, and to underwrite intentional agency.

  - These structures tend to update spontaneously and can be deployed by agents without deliberation, e.g., upon receipt of some new information,

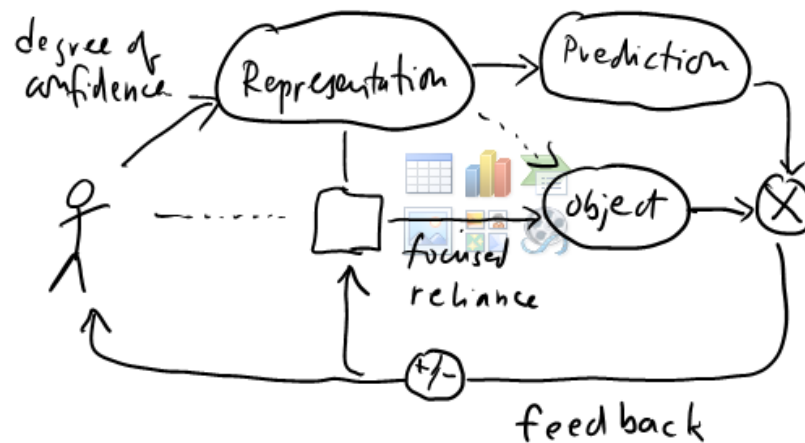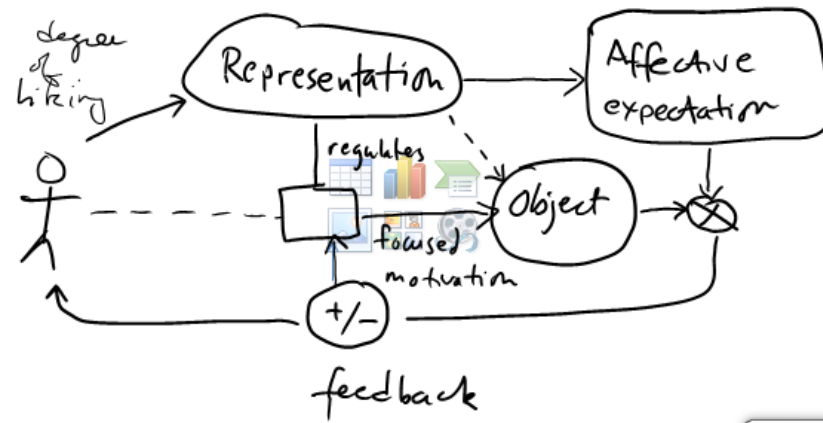 - … but also used as the foundation for more self-conscious, deliberative action—without regress.
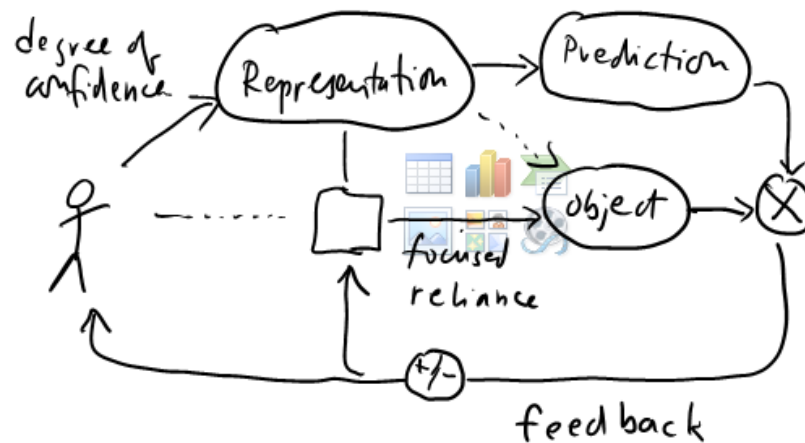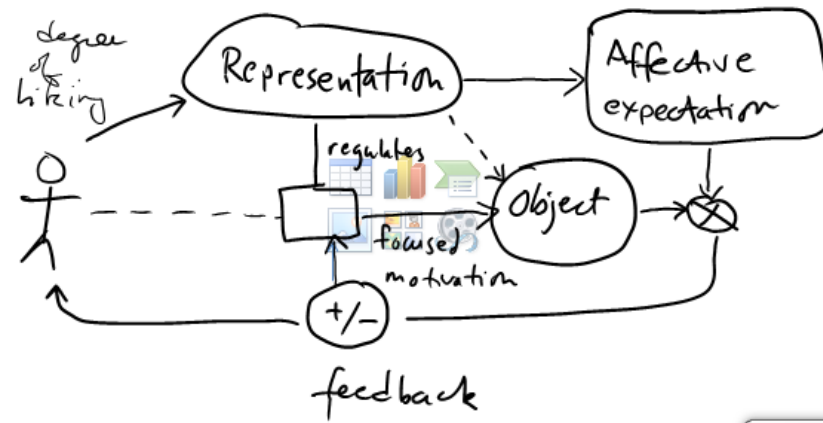
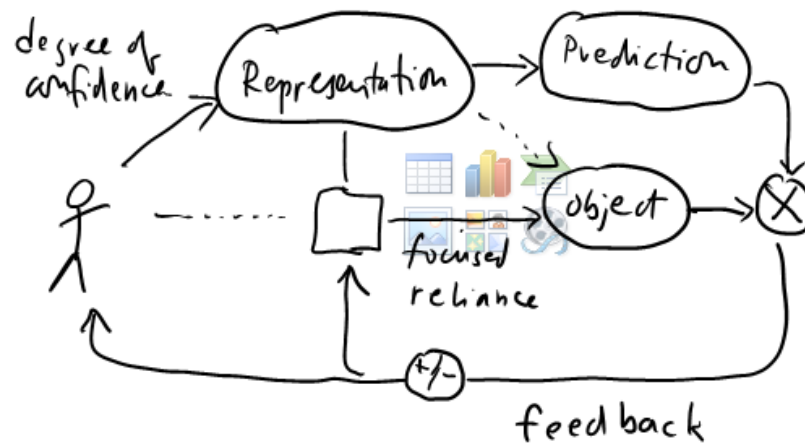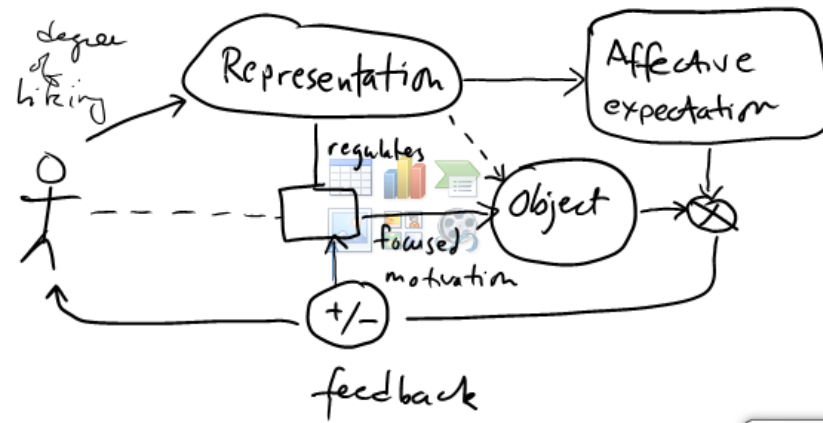# Desire and belief
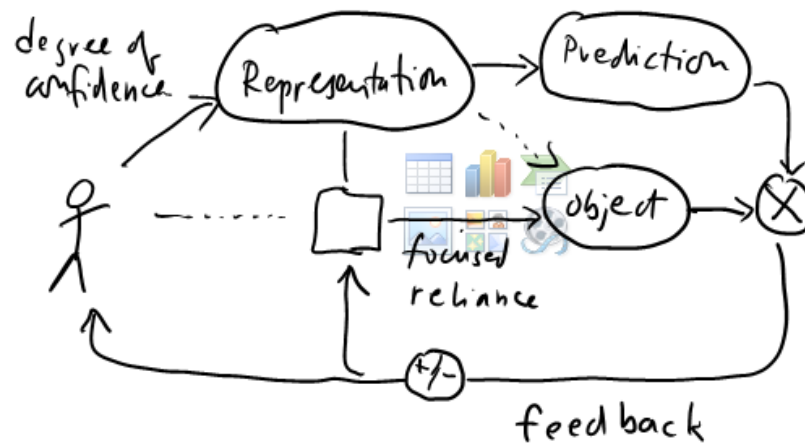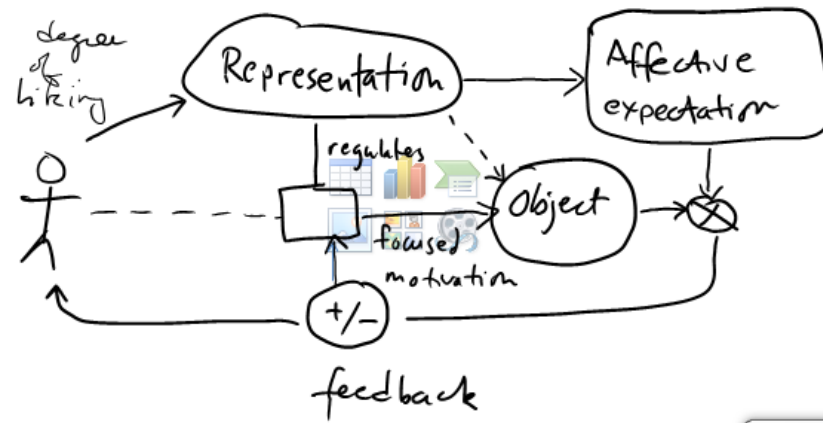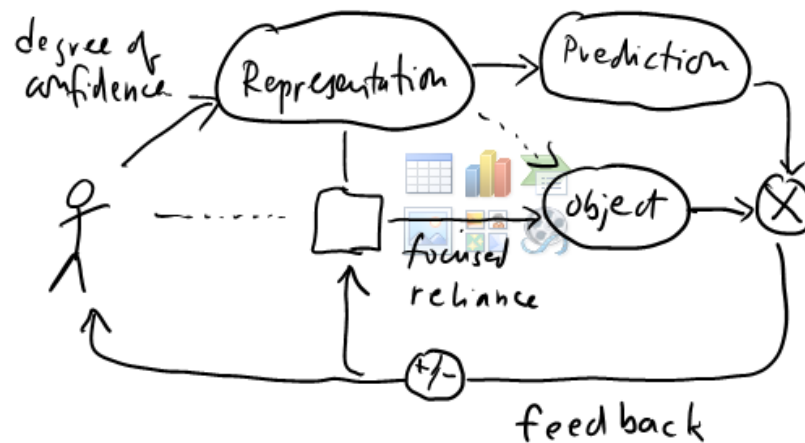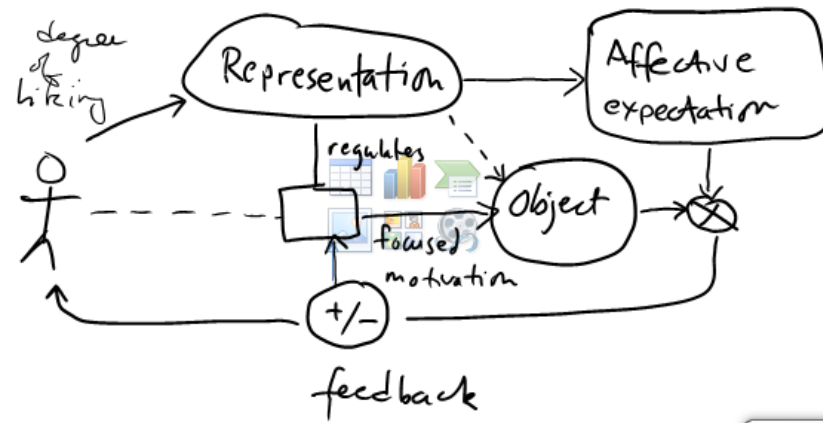
Desire and belief

Desire and belief

Desire and belief

Desire and belief

As we will see more fully

- … when we turn to action in the next lecture, such underlying, intelligent dispositions give us ways of modeling the nature of skilled or fluent action, and fit well with what is known about adaptive control in animals and humans,
  - … and even in artificial, but intelligent, systems.
  - Literally, we might be able to understand action by building it.
- But this can also leave you with a thought: intelligent systems can be built to form and rely upon expectations, to estimate uncertainty, and to pursue goals effectively in light of these.
  - Do they desire? Do they believe? What might be missing?

A possible answer:

- Affect.

- They would have neither confidence nor doubt, neither desire nor frustration, neither fear nor trust.

A possible answer:

- Affect.

- They would have neither confidence nor doubt, neither desire nor frustration, neither fear nor trust.

- Yet.

Back to the Big Issues?

- Recall what differentiated belief from supposition or imagining
    - *not* representing their content "as true"
    - *not* the possibility of being called 'true' or 'false'.
- Rather, two clusters of features are most prominent:
    - (a) belief's spontaneous responsiveness to evidence, and spontaneous guidance of thought, feeling, and action,
    - (b) belief's non-voluntary character and resistance to instrumentalization.
- These features make belief, unlike supposition or imagining, a fitting attitude toward *evidence that p* or *the fact that p*.

Desire and belief as fitting attitudes
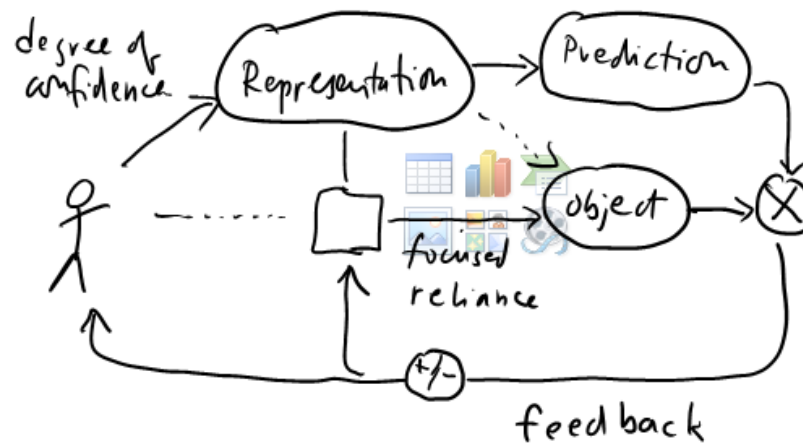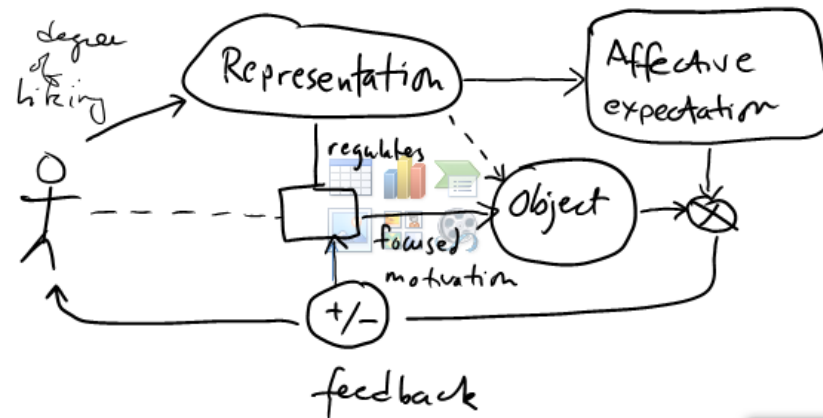
Belief and appreciation of one's epistemic situation

– These are not features we *bring* to belief via the exercise of some judgment or the following of some norm—they are inherent in it, as they are in inherent in desire.

– And just as desire is distinct from "judging oneself to have a reason to act", belief is distinct from "judging oneself to have a reason to believe".

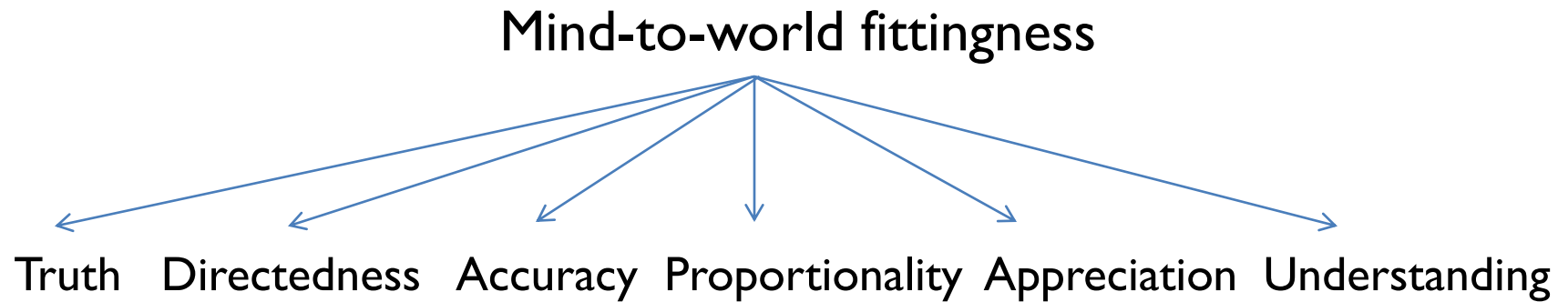 • Nervous flyers, Hume on skepticism.

Back to the Big Issues?

- The distinction between "purely cognitive" judgments concerning the existence of reasons or requirements vs. "genuinely committal" action-guiding attitudes is thus *common* to ethics and epistemology.

  - In neither case does it suggest that *judgments* concerning the existence of normative reasons or requirements are "non-cognitive" or incapable of truth or falsity.

  - Neither does it suggest that belief and desire are "non-cognitive" states since they *are* genuinely committal.

- Rather, it suggests that normative practice needs representational *and* committal states, directly responsive to evidence and independent of will, e.g., desires and beliefs.

Desires and beliefs

- … are the *subjective face* of potentially *objective reasons*.
  - They are not *themselves* the reasons, in the normal case, any more that *fear* is the reason to exercise caution, rather than the risk it represents to us.
- Desires and beliefs *point us* to reasons, help us be *alive* to them, and enable us to respond aptly to them without a regress of judgment or deliberation.
  - They enable us to *understand* and *appreciate* reasons, and to mobilize our resources to translate them into thought and action, just as *fear* enables us to understand and appreciate risk, and to mobilize our resources to meet it.

A more general view

Mind-to-world fittingness

Truth   Directedness   Accuracy   Proportionality   Appreciation   Understanding

The alternative belief-desire model: Hume 2.0?

- **belief   +   desire   →   action**
- *representational   representational*
- *affective + active   affective + active*
- *mind-to-world   m-t-w as well as w-t-m*
- *accuracy of exp.   accuracy of evaluative expectation*
- *representation   representation regulating motivation*
  regulating exp.   and learning from discrepancy
  and learning
  from discrep.
- *potentially rational + potentially rational   →   potentially*
  rational

… or is it Hume 1.0?

Another shameless appeal to authority

Another shameless appeal to authority

- … in the *Treatise*, Hume writes:
  - "*belief is more properly an act of the sensitive, than of the cogitative part of our natures*" (SBN 183),
- Indeed he considers this one of his most important discoveries,
  - … and essential to his critique of rationalism,

Shameless appeal to authority: Hume on belief

- "… 'tis evident, that belief consists not in the nature or order of our ideas, but in the manner of their conception, and in their feeling to the mind."

- "… in philosophy we can go no further, than assert, that **[belief] is something *felt* by the mind**, which distinguishes the ideas of the judgment from the fictions of the imagination. It gives them more force and influence; makes them appear of greater importance; infixes them in the mind; and renders them the governing principles of all our actions." [*Treatise*, Appendix]

Belief and learning

- "… when an object produces any passion in us, [this] **varies according to the different *quantity* of the object** … ." [*T* I.iii.12; SBN 141; emphasis added]
  - So that, when belief encounters varied relative frequencies in the environment, belief is "**broke into pieces**", and "**diffuses itself**" over the various patterns observed, in proportion to the frequency of those patterns in experience (*T* I.iii.12; SBN 134).
- "Probability is of two kinds, either when the object is really in itself uncertain, and to be determin'd by chance; or when, tho' the object be already certain, yet 'tis uncertain to our judgment, which finds [evidence] on each side … " (*T* II.iii.9).

… and the normative status of such feelings?

- Experience-based probabilities of this kind are "receiv'd by philosophers, and **allow'd to be reasonable foundations of belief and opinion**" (*T* 1.iii.13; SBN 143).
  - " 'Tis certain a man of solid sense and long experience *ought to have*, and usually has, greater assurance in his opinions, than one that is foolish and ignorant, and that **our sentiments have different degrees of authority, even with ourselves, in proportion to the degrees of our reason and experience**." [*T* 1.iv.1; SBN 182; emphasis added]

Abstract and appendix

- Hume devoted considerable space to this novel account of belief in the *Appendix* to the *Treatise*, and in the *Abstract.*
  - He admits that this view of belief "seems a little surprizing", and expects it will not be widely understood or accepted, or its fundamental importance grasped [1740: 24].
- Indeed he anonymously published the *Abstract* to promote better understanding and recognition of the *Treatise.*

Even though …

- "we are led" to the affective account of belief "by a chain of propositions, which admit of no doubt" [1740: 22], …

Even though …

- "we are led" to the affective account of belief "by a chain of propositions, which admit of no doubt" [1740: 22],

- … he didn't trust us to get it …

Even though …

- "we are led" to the affective account of belief "by a chain of propositions, which admit of no doubt" [1740: 22],

- … he didn't trust us to get it …

- … and he wasn't wrong about *that*, either.