

Attribution and Behavior

The Problem of Attribution

The Photophilic Fly - Flies will fly toward light (a “trophism”) - It “likes” light, it “wants to go” to light

- Should we attribute “desires”, “goals” to flies? To mammals? To primates? To humans? To robots?

Braitenberg’s Vehicle

- Mechanism: Activation of right light sensor drives left wheel (& vice versa) >> vehicle turns toward light
 - Vs. Activation of right light sensor drives right wheel >> vehicle turns from light
 - OR Activation of right sensor inhibits left wheel that is driven by internal battery >> turns from light
 - SO, behaves as if it “likes”, “wants”, has the “goal to go” to (or from) light
- Some implications:
 - 1) Behavior cannot tell you unequivocally what internal mechanism is involved!
 - So, not justified to attribute internal cognitive mechanisms based on observations of behavior
 - 2) An attributional term (“want”, “like”) is a convenient summary of regularities of behavior
 - NOTE: In that case, these terms *would* just as well apply to vehicle, and fly, and human

Traditionally, researchers DO make such attribution of mechanisms

- e.g. Children *have* a “ToM module” that “turns on” between 2 & 3 yrs of age (based on behavior!)
- Let’s examine the presumptions at work here...

“Theory of Mind (ToM)” = I have a theory that you (and I) have a mind

- **False Belief Task**, AKA “Sally-Ann Test”, Considered the definitive test for “Theory of Mind”
 - Subject sees Sally & Ann, as Sally puts an object in opaque container A.
 - Subject sees Sally leave scene, Ann moves object to opaque container B, and Ann leaves.
 - Experimenter asks Subject “Where will Sally look for the object when she comes back?”
 - Results: 2 yr olds “fail” (pick B) while 3 yr olds “succeed” (pick A)
- Attribution Interpretation:
 - 2 yr old “believes others believe the same as he believes”
 - 3 yr old “believes that other has a ‘false belief’, which is different from his own”
 - i.e. Results explained by the presence/absence of a “mental ability” to rep the beliefs of others
- We would say that 3 yr old has come into coordination with behavioral regularities in social activities involving placing, seeking, finding, and not finding objects.
 - Plus, has developed relevant linguistic competencies - e.g. Use of “will”, “where”, “look for” etc.
 - i.e. Adapted to the multi-party coord of naming, looking, not looking, searching, procuring/not, etc.
- Whatever ToM is, it is not monolithic; Emerges in stages, typically characterized as...
 - 6 mos attribute animacy - *Behavior*: Treat obj diff if initiates own movement vs. only moved by other
 - 1 yr olds attribute intentions (goals) - *Behavior*: If adult reaches for object, infant will get it and give it
 - (Some argue attributing intention is required for Imitation, but see de Barbaro, Johnson & Deák!)
 - 2 year olds attribute different preferences - *Behavior*: Infant says “I like!” and “Mommy not like!”
 - 3 year olds attribute different beliefs (per classic False Belief Task) = “Mindreading”

Nominal Fallacy = Confusing naming a thing with explaining it

- e.g. Saying kids develop a “ToM ability” is used to *explain* why they succeed at False Belief
 - We would say: Sufficient scaffolded experience with situations in which players had differential access to an object’s shifting location and thereafter predictably engaged in different search routines
- e.g. Saying monkeys have “ability to deceive” used to *explain* why subordinates have sex in the bushes
 - We would say: Sufficient experience with, and observation of, conditions under which dominant male does/does not show aggression toward subordinates (e.g. not when his back is to them) leads subordinates to recognize & manipulate affordances (move so no line of sight)
- So, learned regularities in social & material conditions promote predictable behavior
 - Naming the achievement of such behavioral regularity as “having an ability” explains nothing
 - Can be misleading, actually obscure cognitive processes involved

“Mindreading” is Behavior Reading Just what behaviors do we read?

To attribute “Motivation” – Read . . .

- **AFFECT** – Esp in mammals, “emotional displays” correspond to satisfaction, fear, aggression, etc
 - These play a role in negotiating alliances, power struggles, collaboration, parenting, etc.
 - Aspects of facial expressions are mimicked at birth; Smiles shared early in mother-infant interactions
 - Usually associated with own evaluation of situation, but potentially “deceptive” for impact on other
- **TO/FROM**: Observe subjects’ basic tendencies to move TO (approach) vs. FROM (avoid) stimuli

- Ethology: Animals approach some stim (food, mates, friends), avoid others (predators, enemies)
- Psychology: Stim that provoke approach = “positive reinforcers”; avoidance = “negative reinforcers”
- To/From practices “carve up” co-inhabited space, establish “boundaries” in inter-individual distance
 - e.g. “Personal space”, “Approach/avoidance conflict”
 - Can observe “boundaries” functioning as trigger points for change
- **LONG TERM PATTERNS**: W/repeated experience, can also use long-term behavioral patterns
 - Predict will tend to act, in context, as have in the past; This can become very complex in humans...
 - Who tends to attend, turn to/from whom? - Who displays what affect toward whom?
 - How effected are these by who else present/absent? - Who sides with whom in conflicts?
 - Who prioritizes gaining which rewards? - Who tends to adopt which cultural conventions?
- **Biases** in social inference - 3 Parameters interact: Distinctiveness, Consistency, Consensus
 - Distinctiveness: Action/event (by anyone) familiar or novel to B?
 - Consistency: Has A tended to do this to B in the past?
 - Consensus: Does A tend to do this to others?
- If high distinctiveness, high consistency, high consensus
 - e.g. B hurt, A has hurt B in past, A seen hurting others - Attribution: A is at fault, A is a “hurter”
- If high distinctiveness, low consistency, low consensus
 - e.g. B hurt, A never hurt B before, A never hurt others
 - Attrib: No one’s fault, circumstances/accident responsible
- If high distinctiveness, high consistency, low consensus
 - e.g. B hurt, A hurt B in past, A never seen to hurt others- Attrib: A at fault, deliberately target B
- NOTE! Above reminiscent of “Population, Evaporation, Dispersion” in ANTS
 - That is, emergent attribution depends on configuration of multi-dimensions
 - Above require tracking historical “long-term” patterns as well as
 - A+B (you & me) relations as well as A+Others (you & them) relations
 - i.e. Access to full ecology is required, typical of attributors

To attribute “**Knowledge**” - **Epistemic Status** (who knows what) read ...

- e.g. Expert knows more than Novice, acts more Knowledgeable (see Lab 4)
 - e.g. If see other searching, we attribute epistemic status: “He does not know where it is”
- **Surprise** = Typically involves open mouth (gasp) and wide-eyed fixation
 - If person shows surprise, we tend to attribute that they did NOT know
- **Attentional Behavior** = Sensors directed to a target, especially when it involves effortful change
 - e.g. Turn head, reach to touch, re-position for better access, co-ord action w/others, etc.
- Especially **Social Attention** = Target of attention is, or is influenced by, a social other
 - e.g. Solicit attention, Gaze follow, Direct attention of other, Synchronize, Imitate, etc.
- **Monitor Attention**: Tracking other’s attention, including their attention to the attention of others
 - As in **False Belief** task, above, Subject sees that Sally did not see that object moved
 - Competition for attention (often a valued social resource) >> attribute interest
- **Change in Saliency** (increased likelihood of noticing) can be used to assess learning
 - Since affordances change w/learning, expert can observe saliency of affordances to novice
 - How quick is novice to detect, adapt to critical features, opportunities?
- **Direct Other’s Attention** - Show-er is knowledgeable, Show-ee less so
 - Use indexical gestures (see next lecture) to point out, show others objects, events
 - A key function of **Language** is to direct interlocutor’s attention by naming object, place, topic
- **Hearsay** = info to which you have no direct (perceptual) access
 - Have only “the word” of (presumably knowledgeable) speaker
 - e.g. If I say “I ride bikes”, you now attribute knowledge of bike-riding to me, w/o seeing me ride
- **Epistemic Stance**- Any behavior that displays epistemic status (e.g. familiarity vs. uncertainty)
 - e.g. Heritage 2012: In human conversation, information differential is an “**Epistemic Engine**”
 - Ignorant asks, informed replies, stabilizes when both informed
 - Request Information – All languages provide grammatical ways to pose a question
 - Who? What? With whom? Where? When? Why?
- **Exploiting Epistemic Differences**
- Deception (acting “as if” know/don’t know) often involves manipulation of attention
 - e.g. **Gaze aversion** (since looking draws other’s attention), **Mis-direction** (actively direct away)
 - e.g. “**Alibi**” by fixating on object/event conveys attention “occupied”, as alternative to engaging