3

A CONCEPTUAL SCHEME FOR ATTENTION

3.1 **Action and the Selection Problem**

The term "action*"* is used in various ways. In everyday contexts, it can refer to *things that we do*, a category we shall make more precise. In contrast, cognitive scientists, especially motor control theorists, use "action" in a narrow sense to refer to movements of the body, "motor actions" such as reaches for an object, pressings of a button to report, or jumps of the eye. In other words, actions are measurable overt responses. This narrow use unduly rules out *mental* actions such as remembering, thinking, reasoning, imagining and so on. These are also things that we do that are of scientific interest, and involve attention, what some call *internal* attention (Chun, Golomb, and Turk-Browne 2011). As James noted, one form of attention is the mind's taking possession of a *train of thought*.

In this book, "action" will be used broadly to cover things that we do with our body *and* in our heads, so movements of body *and* mind. Still, the generic description "the things we do" is too broad since pulling one's hand back from a burning hot stove is something we do, a spinal cord mediated reflex behavior. Such behaviors are not, however, actions in the relevant sense. Actions as we shall discuss are a proper subset of behaviors which include reflexes as a *distinct* kind. As a rough delineation, actions are those things that we can, in principle, *do intentionally* and which are subject to various *normative assessments*. Actions of this sort can be judged rational or not, skillful or not, morally upright or not, done with good reason or not, controlled or not, done freely or not and so on. These assessments do not apply to reflexes. Reflexes are not rational, moral, done for a (person's own) reason, controlled, express freedom of will, and so on.

Reflexes and actions do share a basic psychological structure. Both are composed of the organism's generating a *response,* an output, in respect of *something,* an input. In psychology, the inputs of interest are the stimuli that the experimenter manipulates with the outputs being measurable bodily responses. In action, stimuli make a difference only if they are registered by the agent's mind. For example, an agent's action on an environmental stimulus typically begins with the agent's perception of it. For creatures that have minds, the relevant inputs will be *psychological* states, say a visual experience of the environment or a memory of it. We can then represent the behaviors of minded creatures as a *coupling* of an input to an output response

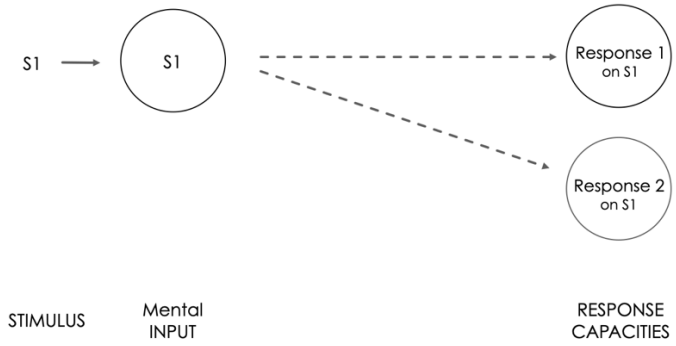
(Allport 1987) where the information or content associated with the psychological input explains the generation and form of the response:

Actions in the normatively relevant sense exemplifies this structure: an agent acts by responding in a particular way, guided by how the agent mentally takes things in. For example, an agent can respond by moving a limb guided by the perceived or remembered location of a stimulus. The reader can fit their favorite actions to this structure.

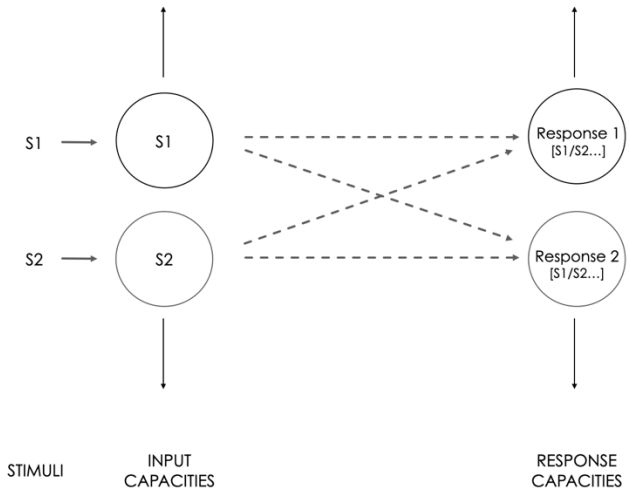
I suggested that reflexes can exemplify this input-output structure. If so, the structure will not separate reflexes from actions. One way to separate them is to note that an agent's action is a solution to the *Selection Problem*. Consider a common instance of the Problem in the fable of the indecisive donkey, inaccurately attributed to Jean Buridan. A donkey stands before two qualitatively identical bales of hay equidistant from it, one to the left, the other to the right. We can diagram the donkey's options, what it can do, in an *action space* that identifies two possible actions, namely eat the left bale and eat the right bale:

These two action possibilities define a *Buridan action space*. Here, the left side depicts two visual inputs, (1) the donkey's seeing the left bale and (2) its seeing the right bale. Each input can guide the donkey's response: translocation and consumption movements. *Doing something* is depicted by "travelling" along one of the two paths (action options), responding to one input rather than the other by eating one of the bales. Otherwise, *nothing* is done. The fable emphasizes that as the donkey has no basis for distinguishing one bale from the other, it has no reason for executing one action rather than the other. The donkey, taking no path, starves to death.

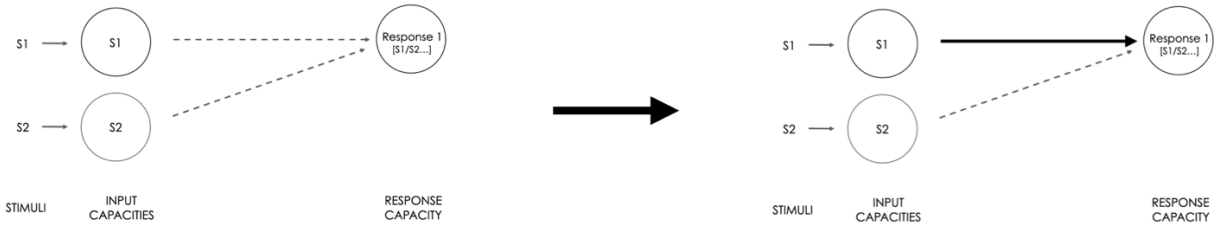
We can pose another instance of the same problem by taking a mirror image of the Buridan space yielding a structure with one input and two responses. For example, perhaps the input is the perception of a predator and the responses are to flee to the left or to the right. If there is no greater advantage to one versus the other, the donkey again has no basis for selection, does not flee and alas, again perishes.



This "mirror images" helps us see that typically, an agent faces a *many-many version* of the Selection Problem (Allport 1987; Wu 2011). That is, for every input there are typically many responses and for every response there are typically many possible inputs:



The Selection Problem raises the question: What is to be done? Here it is important to note that "selection" does not have its ordinary connotations, implying that the solution requires *choice* or the agent doing something (e.g. "selection" when we talk about selecting socks). Rather, "selection" is a theoretical construct. In the theory, it refers to cases where the Selection Problem is solved. A simple explication of this idea begins with (a) inaction, with no path taken in an action space full of possibilities to (b) action, with some path taken, some subset among available possibilities. So, an unfree action, habitual action, and other actions done without decision, choice or intention on the agent's part, are no less solutions to the Selection Problem. Put in *nonpsychological* terms, talk of selection (a solution) is just talk of a specific mapping where an agent moves from an action space where all actions are merely possible (each path represented by dotted lines) to some action being executed (at least one path taken, so an action done). In the Buridan case, the donkey's survives if it instantiates the following transition. Its action is the solution. Solving the Selection Problem is depicted in this mapping, one of two possible mappings.



Note that the transition can be between two actions, say when one switches what one is doing, voluntarily or not. What effects the transition? An explicit decision can do so, but as we shall see, there are other psychological sources that can as well. For example, a salient stimulus such as a bright flash that drives automatic responses like an eye movement to it can compel selection in the sense diagrammed above (e.g. through a salience or priority map, Chp. 3.XX). Buridan's fable vividly brings out the consequences of failed selection, hence failing to solve the Selection Problem.

We can briefly return to what makes a reflex different from an action. For present purposes, a reflex is a behavior which does not result from confronting a Selection Problem and where a solution is in some sense "guaranteed". That is, if there is a reflex response to a given stimulus, then that response will (ideally) always happen when the stimulus is registered. In effect, a well-functioning reflexes allows for no other options and it guarantees a transition to an actual behavior. This is an idealization, but from the engineering perspective, this *pure* reflex eliminates other possibilities to ensure that the requisite behavior happens (Wu 2018). In contrast, actions arise from a Selection Problem where there are no guarantees that anything will be done. Just ask the donkey (or better, bring one of the bales to it). The Selection Problem thus shows that action comes about only if the Problem is solved. Yet when we do act, why do we do *that* specific thing?

3.2 **Biased Solutions to the Selection Problem**

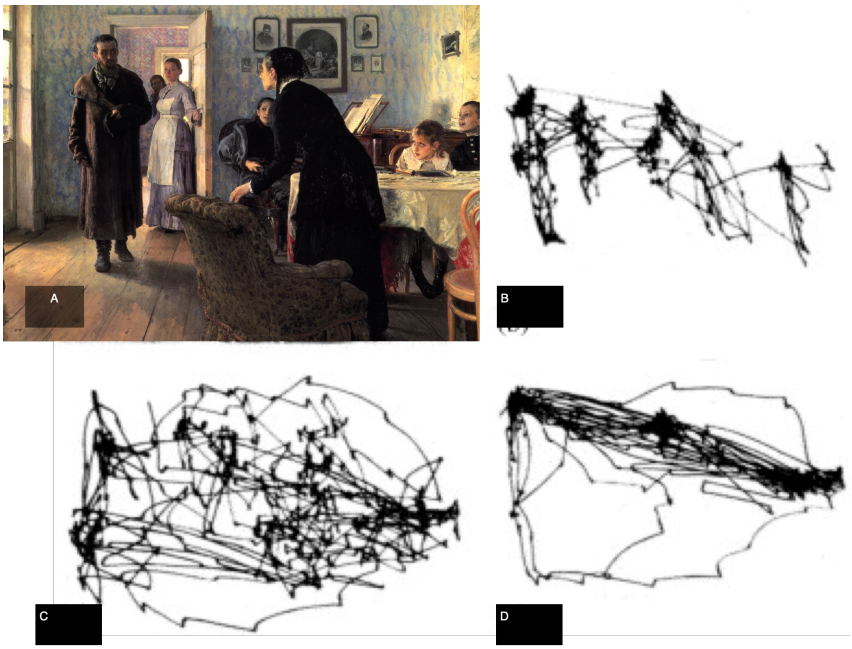
The donkey would have lived had it only decided to eat one of the bales of hay. Consider as well your saying with exasperation, "Just choose something!", to a friend sitting paralyzed between the options of homemade chocolate ice cream or chocolate mousse at a restaurant. One point we extract from the fable is that there must be a basis for selection. Yet recall that the left bale is not qualitatively different from the right, so there is no reason in the perceptible environment to favor one over the other.

The donkey's life depends on a "nudge" to move from inaction to action. Solving the Selection Problem requires what I shall call a *bias,* a selective weighting on the options in the action space. This weighting favors the performance of one option rather than another, and it has many sources. For example, if the donkey intends to eat the left bale, we have one familiar type of bias

embodied in “the will”. Intentions as psychological phenomena have the role of representing an action to be done in order to bring about such actions. For the donkey, an intention to eat the left bale of hay (versus the right) would, barring any impediment, lead to its eating that bale. It would thereby solve the Selection Problem…and secure survival! The general notion of a bias in this context is understood functionally as a factor that influences the agent in yielding solutions to the Selection Problem (see figure above). Intentions (plans, decisions, choices) are a familiar type of bias. As noted earlier (also Chp. 3.XX), the salience of an object is another type (a bottom-up bias; we return to the top-down versus bottom-up distinction below).

A theorist who is inclined to be skeptical about attention is also likely to be skeptical about intention. This seems to me to be misplaced for two reasons. First, we can give a clear characterization of intention as a psychological kind in terms of what it represents and its causal function. An intention represents an action as to be done by the subject, often as a way of retaining (remembering) a decision. The commitments need only be to some shift in the agent’s intentional orientation to the world. Second, experimenters manipulate intention so understood in the lab by giving task instructions. After all, if a subject does not intend to follow the instructions, then there’s no point in running the experiment. Experimental psychology presupposes intentions as that state of mind set by task instructions and which organize subsequent action. Intentions so understood can be categorized, provisionally, as a top-down, cognitive influence. Intentions provide a *cognitive* bias to solve the Selection Problem.

Alred Yarbus provides a beautiful experimental demonstration of bias by intention where what the subject intends affects how she moves in an action space. Yarbus presented his subjects with a painting of a homecoming scene and asked them to perform various tasks including: (i) remember what the people in the picture are wearing; (ii) remember the location of people and objects and (iii) estimate how long the visitor has been away. He then tracked their eye-movements (overt visual attention) while they carried out his instructions and noted the following patterns:



Legend: Instructions are as follows: (b) memorize the clothes worn by the figures; (c) memorize the location of objects in the room; (d) Estimate how long the father has been away.

Interpret Yarbus's experiment as manipulating his subjects' intentions through changing task instructions. Given a constant input, the painting, Yarbus measured visual attention through patterns of overt visual attention. In effect, modulating intention as a causal factor on responding to a constant stimulus, Yarbus examined effects on overt response in the eye. What is striking is that the patterns of eye movements makes sense given the subjects' more abstract intentions to carry out Yarbus's various tasks. For example, asked to remember the clothes, the subject intentionally looks at each figure. Asked to estimate how long the father has been away, the eye fixates the faces of the family members, presumably to assess emotional response. The subject's intention need not be to move the eyes in any specific way, but the resulting pattern of eye movements in each panel is intelligible given the intention in question. Perhaps not surprisingly, modulating intention through task instructions leads to different solutions to the Selection Problem, different actions in response to the same target. The sign of this is that the movements make sense given the intention set by task instructions. This suggests that intention also influences attention, canonically top-down attention. Yarbus' experiment provides an example of biased behavior in light of intention.

3.3 **Attention in Action**

If we view action through the lens of solutions to the Selection Problem, then an action is composed of an output response that is guided by a psychological input. Wu (2018) argues that *all* actions necessarily have this structure by drawing on the contrast between actions and reflexes. For our purposes, it is enough that all the actions that a psychologist, neurobiologist, or ethicist cares about are solutions to a Selection Problem. Crucially, it then follows by the Jamesian condition that *every such action involves attention*. After all, James's condition states that attention is the mind's taking possession of an object to deal with (act in response to). We can illustrate this with the Buridan space, reimagined so the donkey lives. For whatever reason, the donkey eats the bale to the left which we represent by showing that one of the two options is taken (solid versus dotted lines):

Notice that the structure satisfies the antecedent of James' condition. We can read this off from the diagram. Take the idea of the mind's taking possession of one of many possible objects to deal with. In the diagram, this is instantiated in one of the psychological inputs *rather than others* playing a role "in further processing". The mind, via perception, takes possession of the target left bale and withdraws from the ignored right bale. It does so to deal with the left bale, i.e. move to it and chew. In the diagram, we capture the Jamesian antecedent (taking possession of to deal with) in linking the seeing of the left bale to the response (solid line). By the Jamesian condition, the donkey thereby visually attends to the left bale. This is the right result, what we would say in everyday descriptions of such behavior. Importantly, we derived the attribution of attention from a condition that we argued is central to experimental cognitive science along with the structure of action identified in the Selection Problem. We have thus explicated the general psychological structure at issue in the knowledge that we share with James.

The presence of attention holds for any solution to the Selection Problem, that is any action. For when an action happens, there will be a prioritization of an input in the sense that it is that input that guides behavior. Hence, the point also applies to the mirror image Buridan case where there is one input, seeing the predator, and two outputs. Although there is only one object, the donkey in fleeing by running to the left away from the predator takes possession of the danger to deal with it. This still exemplifies the taking possession of a target to deal with, but with the absence of withdrawing from potential distractors. The key lesson is that the fundamental expression of

attention is when a psychological state informs a response. On the Jamesian view, attention is expressed *in action.*

Interestingly, on the resulting picture, attention is found in the basic components of action. Notice that if the Selection Problem identifies the basic components of action sufficient for attention, there isn't the need for an "attentional spotlight" in the sense of a distinct mechanism that *is* attention (recall the problem of polysemy, Chapter 2.XX). There are no additional nodes needed in our causal structure to secure attention beyond the input (vision), the output (motor response), and a bias, such as the intention to eat that bale. Nothing corresponds to a spotlight or other type of mechanism.

To clarify this point, consider building an agent. We'll assume that the nitty-gritty technical problems are solved and that we can build a system that can see and can move. This agentive system has a mind, and its movements are guided by what it sees, so the visual inputs inform the production of appropriate output movements. An engineer wants to build an android rover to act on an alien planet, namely to collect rocks of certain kinds and to avoid certain threats. So, the engineer builds in a visual capacity that allows the rover to locate rocks and recognize threats. Movement capabilities are also added that allow the rover to translocate to and from relevant targets. Since the engineer does not want a Buridan agent, she builds in a "control" system that encodes two basic goals: find rocks and avoid specific threats. She can also build into the resulting weightings a greater bias towards avoiding threats, but wiring this is the engineer's problem. The point is that given the information encoded, the control system functions like the rover's intention to look for rocks and avoid threats.

We let the agent loose, and it fulfills its goals and survives. When the agent finds a rock and moves towards it, it is acting in a way that is guided by visual attention to its target and when it flees from threats, it uses visual attention in the same way, visual selection to guide a fleeing response. The engineer is entitled to this claim because her system satisfies the antecedent of the Jamesian condition. Notice that attention is present without having to build a separate attention module distinct from the visual, movement, and control modules already in place. Attention is secured when agency is secured. Of course, the engineer *could have* built such a module, but this would be redundant, an unnecessary extra node between intention and visuomotor processing. Securing attention in the agentive system does not require further engineering. This perspective is consistent with those who deny that attention is a distinctive mechanism (Anderson 2011). The absence of agreement about the existence of such a mechanism/module was forcefully emphasized by Alan Allport (1993) (see also the first objection in (Hommel et al. 2019)).

Attention on this *deflationary* or *economical* view is not an epiphenomenon. It is a causal factor in the following sense: the visual state that guides the response when the agent acts just is the agent's visual attention to the target. That is, the agent's seeing, when it contributes to behavior, is the agent's visually attending (taking possession of the target to deal with). Accordingly, the rover's response is causally guided by visual attention in being guided by the prioritized visual input. We can talk about a "spotlight" if we wish, but we don't thereby refer to something distinctive in the machinery that isn't already there. The rover case is a thought experiment to show that attention in agency does not require a special attentional mechanism. It is, however, an *empirical* question whether human or other actual biological agents have a distinct attention module. That question is not prejudged by the present point which is merely the economical one that we get attention "for free" when we have agency. Interestingly, this economical perspective on attention was expressed in one of the central theories of attention in cognitive science: *biased competition.*

3.4 **Biased Competition**

John Duncan and Robert Desimone (1995) presented a *biased competition* theory of attention. Competition can be connected to resource limitations. For example, the Selection Problem can be taken to be a type of competition, namely for the control of the agent's behavior. The solution to the Selection Problem is effectively a resolution of action competition that involves each action path trying to take control of what the agent does. The failure to resolve competition is the failure to solve the Selection Problem (recall Buridan's donkey). Accordingly, when competition among actions is resolved, an input is prioritized in being that which guides the resulting behavior. Such prioritization constitutes the agent's attention in action. In biological agents, the resolution of competition is realized in the agent's brain.

Desimone and Duncan focus on this issue through examining the shift in neural response in the visual system during visually guided behaviors. They noted that neural representational capacity is a limited resource, a point they characterized in terms of a visual neuron's *receptive field.* A visual neuron's receptive field is that area of visible space wherein an appropriate stimulus will drive a neuron to respond by generating action potentials or spikes, a fundamental signal in the nervous system. Given various biophysical constraints, there is a limit to the number of spikes per unit time that a neuron can generate.

We shall discuss the neuroscience of attention in more detail in Chapter 5 (here, we give a schematic overview of work done on nonhuman primate vision, (Chelazzi et al. 1998; 2001)). For present purposes, we need only note that the visual neurons at issue have varied responses to different stimuli. Take two visual stimuli X and Y and a neuron N that responds to each (in principle, X and Y could be two bales of hay, but these were not the stimuli used in the experiment!). When we present each stimulus singly in N's receptive field, N generates different levels of activity, say *n* spikes per second to X and *m* spikes per second to Y. Interestingly, when X and Y are presented *together* in the receptive field, N's neural response is the weighted average of the individual responses rather than, as you might have expected, the sum ($n + m$). We treat this averaging as reflecting competition between the stimuli for the neuron's limited capacity for generating spikes. A resolution of competition, a sign of winning, is that one stimulus, say X, ultimately gets the neuron to respond *as if only X is in the receptive field.* Accordingly, the loser, Y, gets ignored from the perspective of neuronal response. Thus, if N initially responds to X and Y with the weighted average of n and m (competition), X wins the competition when N's response comes to be at n even though Y remains in the receptive field. It is as if Y's presence no longer contributes to neural response. N's receptive field has seemingly shrunk its boundaries around X thereby excluding Y. Notice that we have a neural correlate of the Buridan space.

In the experiments under discussion, competition is resolved when one stimulus is favored over another in respect of neural representation, which one the neuron effectively responds to. Notably, these experiments were carried out while the animal performed a task given an instruction that identified a specific object as *task relevant,* say X. That is, the experimenter indicates to the animal which of several stimuli is the task relevant target. Once the animal (here a macaque monkey) is informed which target is the one to respond to, and hence the animal makes precise its intention to act ("*That* one!"), it then acts on that target. Assuming that the changes in neural activity described are part of the neural basis of implementing a visually guided action, we can hypothesize that the intention, representing this X as the one to acted on, then biases the animal's response to the Selection Problem. It is when the animal acts (solves the Selection Problem between X and Y) that the observed *remapping* of N's receptive field occurs, narrowing on X rather than Y. That is, the behavioral selection of the task relevant target that is part of solving the Selection Problem is reflected in corresponding neural selection of the task relevant target. Putting this in the context of the Marrian approach, at the computational level, solving the Selection Problem that yields attention in the Jamesian sense is based on resolving competition that yields neural selectivity.

Given our earlier point about economical engineering in the rover, it is worth noting that Desimone and Duncan cast their model of biased competition in relation to attention as follows:

> The approach we take differs from the standard view of attention, in which attention functions as a mental spotlight enhancing the processing (and perhaps binding together the features) of the illuminated item. Instead. the model we develop is that attention is an emergent property of many neural mechanisms working to resolve competition for visual processing and control of behavior (194).

We can read their denial of the spotlight as a denial of the need for a separate module for attention that is the source of the neural modulation observed in neural recording. Once biased competition is resolved, the resulting selectivity in the visual system *yields* attention (attention "emerges").

We can speculatively extend the basic picture from a single neuron to the relevant neural systems. Consider the visual system's being presented with stimuli X and Y, say two bales of hay. If the subject attends to X, then at appropriate stages in the visual processing hierarchy, receptive fields might remap in favor of X over Y due to biased competition, so that it is as if only X is present in the visual field. X has won the competition over Y. Let this process play out throughout the relevant parts of the visual system such that at the end, for sake of argument, all task relevant neural representational resources has been apportioned to the winner X. So, the system is in a state of having selected X rather than Y at the level of neural response where the visual system's selective processing of X can then inform response. This links a schematic mechanism, resolution of competition for spikes and consequent remapping of receptive fields, to the idea of selecting a stimulus for further processing such as performing a task.

The expectation, then, is that large-scale neural activity will be apportioned to the winning stimulus and away from the losing stimulus in the competition for neural activity. Experiments using imaging such as functional magnetic resonance imaging (fMRI) provides a snapshot of brain activity that is consistent with the predictions of biased competition. fMRI measures metabolic changes in the brain in changes in blood flow through the BOLD (blood oxygen-level dependent) signal. We assume that such metabolic changes reflect changes in neural activity as since neural activity consumes energy supplied by blood flow (we should note that the BOLD

signal is not a direct reflection of the generation of action potentials, XX). This requires that the relevant brain areas receive additional oxygenation. What is observed is that attention to X rather than Y activates regions devoted to processing X at the expense of those regions devoted to processing Y. For example, the observed BOLD signal can increase relative to baseline in regions of the brain that respond to X while it decreases in regions that respond to Y (see (Kastner and Ungerleider 2001) for discussion of some relevant results).

Recall skepticism about attention. In the current skeptical climate, a recent paper by Hommel et al. (2019), "No One Knows What Attention Is" is often cited. Let us close this section by making a few comments about that paper. Hommel et al. raise three worries: (a) the concept of attention might mislead theorists into affirming that attention is a single mechanism, (b) "attention" is confusedly used to refer both to an explanandum (that to be explained) and an explanans (that which explains), with this leading to problematic circular statements (attention explains attention; see Chapter 2.XX); and (c) that the attentional operations are wrongly taken to be distinct from other psychological operations including "action-planning" (2289). These worries are well taken, but note that they are compatible with our discussion.

In light of their worries, Hommel et al. advocate an approach "synthesizes elements that actually correspond to real biological processes at both neural and functional levels" (2294). In particular, they consider the evolution of biological systems wherein attention is found, and notably highlight organisms that involve the need to "*select* between actions, such that one completely suppresses the other. This kind of selection could be accomplished through lateral inhibition that produces "winner-take-all" dynamics." This characterizes the Selection Problem in competition terms. Hommel et al. at this point invoke James's observation that attention "implies a withdrawal from some things in order to deal effectively with others." The authors eventually conclude:

> selectivity emerged through evolution as a design feature to enable efficient goal-directed action. Such selectivity became necessary as the action repertoire of the given line of organisms that led to humans increased. This means that selectivity is an emerging property arising from a myriad underlying processes (2298).

Arguably, their conclusion is Jamesian in the sense that we have been explicating with attention as explained by the demands on agency. As argued in Chapter 2, that is what everyone knows.

3.5 **Dichotomies**

3.5.1 Top-Down versus Bottom-Up

There are two common ways to characterize forms of attention. One is to distinguish *top-down* from *bottom-up* attention. Bottom-up attention is sometimes described as, or includes, *stimulus-driven attention, attentional capture, pop-out*, even *automatic* attention. These ideas are connected to a technical notion of *salience* (next section). In contrast, top-down attention is often explained as *goal-directed* attention where the goal is part of the intended task, so top-down attention involves "cognitive" influences and is sometimes spoken of as *voluntary* attention. Top-down and bottom-up attention can act at cross purposes, for bottom-up attention might disrupt top-down attention as when our attention is captured and pulled away from an intended task. The two forms are also thought to be subserved by different neural systems (Corbetta and Shulman 2002).

The contrast between goal-directed and stimulus-driven attention is *not* exhaustive since it leaves out other forms of attention that we shall discuss. Hence, if top-down and bottom-up aligns with goal-directed versus stimulus driven, as many seem to think, then the former is also not exhaustive. Neither dichotomy completely partitions attentional phenomena. For example, what of attention shaped by the emotions, as how fear or love can direct attention? As characterized above, emotion attention is neither top-down nor bottom-up. Some have advocated rejecting the top-down versus bottom-up *dichotomy* (Awh, Belopolsky, and Theeuwes 2012). What should be rejected is the two ideas capture all the attentional phenomena. Noting this, one can continue to talk of top-down versus bottom-up without denying that there are other interesting forms. For example, Awh et al. go emphasize various historical effects (sometimes referred to as *hysteresis*) which we shall discuss later.

It is worth noting that the introduction of hysteresis introduces a *diachronic* versus *synchronic* conception to attention, that categorizes attention *temporally*. Such categorizations will be relative to how we decide to divide time in analysis in specifying the boundaries of the relevant epochs. Consider explicating synchronic versus diachronic in relation to an *experimental trial*, one round of performing an instructed task. For example, an experimental trial might involve

first identifying a target among other possible targets to respond to, and in that way set a goal for attention to a target. In this case, we can think of top-down modulation as synchronic relative to a trial, for the cue's influence is within the trial by making the task, hence operative intention, more specific. Relative to a given trial, a diachronic influence can be understand as rooted in events prior to the trial in question that affects attention during performance such as the effects of past reward or priming in previous trials. Naturally, we can set other temporal boundaries, depending on our interests, but the point is to make clear what the relevant boundaries are. Hysterisis focuses on diachronic attention.

The negative response to the top-down versus bottom-up dichotomy is that it leaves out many interesting forms of attention. While I have noted that the pair is harmless so long as one recognizes that it is not exhaustive, a clearer way to proceed is to use the generic notion of a *bias* and explicitly enumerate different sources of bias of relevance to a discussion, thus biases due to goals, stimulus properties (salience), emotions, historical effects like reward history, and so on. The advantage of talking about an *X*-bias (an emotion bias, a reward bias, a priming bias, etc.) is that the terminology explicitly identifies the relevant causal factor, and thus, conveys more information about the causal structure of attention.

3.5.2 Control versus Automaticity

XX

3.6 **Salience**

Task instructions that set intentions lead to prioritizing certain targets as the ones the subject should deal with. These targets are *task-relevant* in light of what we intend to do. Yet good intentions are an imperfect shield from *distractions,* those things in the environment which, as one might say, are salient and grab our attention. We shall return in more detail to the folk-psychological sense of salience when considering ethical aspects of attention. Here, we shall consider the notions as technical terms that draw their meaning from the theoretical constructs of *salience maps* and *priority maps.*

"Salience" is a tricky term in the theory of attention because its ordinary use covers things that are noticed or things that *should* be noticed. As such, saliency is a property ascribed to an item in

light of the subject's attention or ability to attend to it. The everyday use of the idea is something like this: An item is perceptually salient to a subject precisely because it *is* noticed by said subject (or perhaps even *can* or *should* be noticed). This relativizes salience, for in a given context, what is salient to you might not be salient to me since you, but not I, notice the item. The item *could be* rendered salient to me if I paid attention to it. Used in this *broad* (everyday) sense, salience can characterize anything that we can attend to: the color of a hat or of someone's skin, the sadness apparent on a face, the face itself, a disturbing thought or the subtle negative meaning of a passing gesture.

In cognitive science, however, "salience" has a *narrow* meaning most notably tied to visual spatial attention and the concept of a *salience map.* In this narrow sense, only *locations* are salient. An algorithm for computing visual salience was initially presented in a foundational paper by Christoph Koch and Shimon Ullman (Koch and Ullman 1985) and implemented more fully by Koch and Laurent Itti (Itti and Koch 2000a) (for an accessible review, see (Itti 2007); for a deeper dive, see (Itti and Koch 2001b)). A salience map assigns a scalar value to a location, what we will refer to here as its *salience value* (it is not clear if salience map theorists would say that the map *represents* locations as having the property of being salient to X magnitude).

On many models, the salience map can be represented as a spatial map with peaks at each location where the heights indicate the salience value at that location. The Itti and Koch algorithm takes individual *visual feature maps*, say for color or intensity, as input and computes the *conspicuity* of each location L with respect to that feature by comparing the feature value at L with a set of neighboring locations along the relevant feature dimension. This is done at various spatial scales. For example, the algorithm takes a map of intensity and identifies the brightest location as the most conspicuous relative to its contrast with the intensity value at nearby locations. The *conspicuity maps* generated by each feature map are then normalized and combined to establish a salience map which assigns a value for each location. Other salience maps differ in their explanation of how features maps are transformed to salience maps. The central idea is that salience maps are the outputs of a computation over feature maps.

The visual salience map has been used to describe and predict saccadic eye movements that are putatively driven by visible features of the environment. That is, it accounts for the generation of bottom-up overt visual attention. This is because what matters to generating the eye movement are basic visual features and not, say, task. Programming where the eye moves to next is explained by a *winner-takes-all* mechanism that identifies the location with the highest peak on the map as the location to which the eye is to move next. To ensure that the mechanism does not

perseverate, constantly oscillating between the two most salient locations, the algorithm builds in an *inhibition of return* mechanism to ensure that the next highest peak will be the target of the subsequent eye movement. In Itti and Koch's algorithm, if the inhibition of return period was reduced from 900 milliseconds to 50 ms, the algorithm would endlessly cycle between the two most salient items in a display (p. 1503). At each fixation, the information at the salient location is transferred to a *central representation* for further processing such as to determine the identity of an object at that location.[1]

FIGURE XX

Strictly speaking, only locations are salient in the sense tied to a salience map. A conceptual challenge is to avoid equivocating between "salience" in this narrow sense with "salience" in the broad (everyday) sense where any target of attention, what one notices, can be said to be salient. After all, the broad claim that a face is salient, while natural in ordinary speech, has no meaning in a salience map because that map, exempting spatial information, is *featureless* and, presumably, *objectless:* the "saliency map only computes bottom-up information and has no notion of what an object is" (Elazary and Itti 2008, 12). Let a signal be feature- or object-*silent* when it provides no information about features or objects ((Veale, Hafed, and Yoshida 2017) speak of salience maps as feature-*agnostic*). Given the peaks of the salience map, one has information about where the eye will move next, but one can only guess what features or objects are at that location.

Jeremy Wolfe (1994, 238), fn. 6) suggested that while the salience map does not represent objects, "in an array of items [objects], activation will be higher at the items than in intervening blank space. Objects and locations will be, in effect [practice], the same thing" (ibid). A natural hypothesis is that a location is typically salient in the narrow sense because the object at that location is salient in the broad sense. This correlation is not perfect. Lior Elazary and Laurent Itti

[1] While Treisman invoked an attentional spotlight that binds features for object representation, Koch and Ullman construed selective attention as the mapping itself: "Selective attention is a mapping of the properties of a given location, the "selected" location, into a higher, non-topographic [central] representation" (220). One could read this as saying that selective attention is realized by the resolution of competition when the maximum peak transfers its contents (features, objects) to the central representation (cf. biased competition). The winner-take-all mechanism might play part of the role Treisman attributed to her spotlight.

(2008) computed the salience of locations on photographs where notable objects had been labeled by human observers (salience in the broad sense).[2] They reported that the when constructing a salience map of these images, "the saliency map showed a 43% chance of finding a labeled object within the first predicted location and over 76% chance within the third predicted location." Itti and Koch (2001a) trained their algorithm to more closely align the output of a salience map to highlight locations of (broadly) salient objects such as pedestrians on a city street (op. cit.) or military vehicles in an isolated landscape (Itti and Koch 2000b). This involved using supervised learning to increase the weight on relevant feature maps that were correlated with the broadly salient objects. Notice that this begins to shift the salience map from a purely stimulus-driven, bottom-up construct to one that is influenced by additional information such as knowledge, prior experience, and goals.

It should be noted that even if the correlation between narrow and broad salience were perfect, the salience map remains feature- and object-*silent*. This means that after the eye moves, there is still work to do by action systems since task-relevant information must be selectively extracted. The agent's goal is often to respond to features or objects, not merely to move the eye to the right location. Spatial attention leaves work to be done for feature and object attention (this is revealed in certain classic experiments on feature-based attention, Chp. XX). As Koch and Ullman conceived of it, the salience system identifies a relevant area for further analysis. The non-spatial targets at the salient location are passed on to a *central representation* (draw figure 1, Itti and Koch 2000, p. 1491). Further selection mechanisms focus on features or objects, say to aid task processing.

The salience map provides a biased view of the visual world in that saliency is computed in a way that emphasizes featural differences. In a sense, each location in a feature map is competing with others to be picked as most conspicuous. As an engineer, one could define a notion of salience in some other way, say the inverse of conspicuity. There are natural situations, say teaching in the class room, where one does not want to pick out the most salient individual, the person whose hand is always up, but rather the individual who is most silent and has not participated yet. A good teacher is one whose notion of salience, in that context, does not track conspicuity but is informed by a broader range of factors. Said teacher must fight a type of visual bias in order to further her pedagogical goals. This points to the broader notion of salience that we deploy in ethical and evaluative contexts.

[2] Russell, B. C., Torralba, A.,Murphy, K. P.,&Freeman,W. T. (2005). LabelMe: A database and Web-based tool for image annotation. MIT AI Lab Memo AIM-2005-025

The salience map is one of visual space. Does salience apply to other modalities, sensory or otherwise? There is work on auditory salience that draws on similar ideas (e.g. (Kayser et al. 2005)) and to the extent that a sensory modality is subject to attentional capture, it is likely that salience maps can also be applied to those systems. Still, extending the idea of a salience map is not as easy as extending the idea of attention as selection to guide behavior. We are also apt to speak of thoughts and memories as salient in that they can trigger responses, but a cognitive salience map is not obviously spatial, so the extension of salience maps to certain cognitive domains is not obvious. The same issue arises with respect to sensory modalities that are not strongly spatial such as olfaction. Certain smells, after all, can be quite salient (broad sense).

We have noted that there is a general notion of a bias needed to solve the Selection Problem and the salience map identifies a type of bias that resolves a Selection Problem in respect of potential eye movements given the visual world. That bias, in the pure bottom-up case, is driven by properties of the stimulus as processed in the relevant perceptual system. Yet there are many types of bias that also influence behavior. As we saw in Yarbus' experiment, eye movements are sensitive to task that goes beyond salience in that the eye can move in different patterns relative to the same painting. Accordingly, to explain the movements that Yarbus observed, a salience map provides an insufficient explanation. In this context, many speak of a *priority map* which takes in other biases including task relevant information.

3.7 **Priority**

A prominent conception of priority maps is that they are transformations of salience maps that factor in all the other biases that influence attention to produce a map of visual space that assigns a scalar value for priority. Among these additional influences, Rebecaa Todd and Maria Manaligod (2018) list statistical learning, semantic associations, reward, and affective salience. We have already mentioned task, but no doubt the list is incomplete. Bodily states like hunger or thirst, representational states like memories of various sorts, beliefs or desires, and historical effects like priming also have an effect. No doubt there are others. To better describe observed visual behavior, Itti and Koch had to include appropriate training into their algorithm to alter weights that sculpt the construction of their salience map towards objects that are salient in the broad sense.

At this point, to maintain a clear theoretical apparatus, it is best to restrict the technical notion of salience to the original conception of the salience map and to use "priority" to pick up the content of the priority map. As the priority map takes in the salience map as an input, but produces a similar map of scalar values for priority tied to locations, the resulting representation remains feature and object silent. This means that the argument for the insufficiency of the salience map to explain all forms of attentional behavior remains for the priority map. Even when the eye lands on a location of highest priority, there is still further selection needed in order to serve the subject's precise goals, say the extraction of specific features or objects at that location. One might wonder whether the salience map becomes superfluous given that it is factored into the priority map, but the former might still provide an account of purely stimulus-driven or bottom-up attention.

One potential biological challenge is that the computational models seems to commit itself to a *single* salience map. The biological data has identified various potential neural substrates of salience in visual areas including the frontal eye field, that lateral intraparietal area and the superior colliculus, all three known to be involved in the generation of eye movements and of visual spatial attention. XX. The formal model, however, need not be committed to a single neural localization for salience or priority computations. For example, there might be different salience maps for different behaviors: salience as driving visual attention tied to the eye movement system and visual attention tied to other motor actions, each effector system using its own dedicated salience map.

This leads to the question whether one needs to speak of a map in the more localized sense that comes out of the computational literature but with an eye towards implementation. We have already noted that a map of salience or priority that is limited to spatial location leaves many selective behaviors targeting non-spatial targets unexplained. It is open for theorists to extend priority maps to represent features and objects. At this point, one might wonder whether the idea of a map is specifically necessary rather than simply the idea of a bias to task-relevant representations such that prioritization of such representations is part of those representations coming to serve the guidance of behavior. The signature of such prioritization can be the relevant strength of the signal relative to other competing signals that represent "distractors". That captures a core idea of of the salience and priority map literature while deemphasizing the idea of a map, something that can itself be located in the brain.