

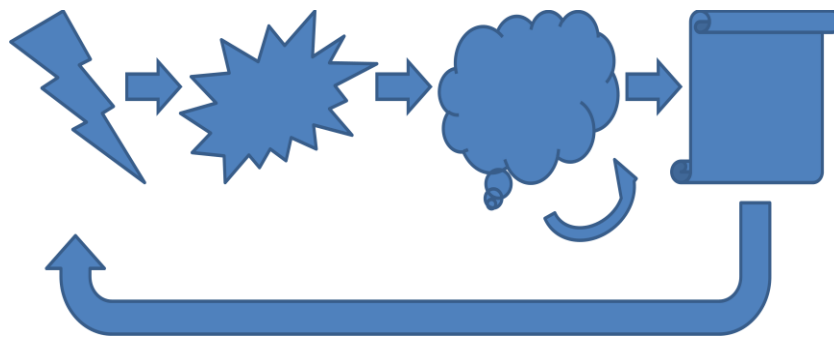
LAUNCHING INTO ACTION

Aim of this segment:

- Introduce the “classical” theory of action
- Flip it inside out

I. The efficient-causal spiral

Recall our little cartoon from day 1:



- The bolt of lightning represents some occurrence in the external world; the flash is a perceptual “experience”; the thought balloon is a total belief-desire intentional state (the round-pointing arrow represents its extensive endogenous dynamics); the proclamation is an action.
- The arrows represent “efficient” causation of the sort studied by classical physics: e occurred because e^* occurred (earlier; thanks to cooperating background circumstances; in a way “governed” by laws that are at least *ceteris paribus*).
- Accordingly, we have a progression in time. The loop-back encodes a deliberate type-token confusion: I don’t mean that the declaration efficient-causes the *very same occurrence* in the external world that set the chain in motion; rather, it causes some distinct external-world occurrence.

A lot of people think this *efficient-causal spiral* is a general form of a person’s psychological history. The argument typically runs:

- i. Someone has a psychological history just to the extent that they are rational
- ii. When someone is perfectly rational, their psychological history is an efficient-causal spiral
- iii. Everyone’s psychological history is an efficient-causal spiral

The argument is not valid but it is clear enough what one would do to patch it up.

A. Rationality

Two motivations for this, Davidson- and Lewis-style. Davidson: (a) psychological predicates only hit the interpretable (b) interpretable to the extent that rational. Lewis: (c) psychological predicates only hit those who meet analytic postulates on said preds (d) said postulates demand some rationality.

Both of these guys are perhaps worried about what fixes the “ultra-rigid rails” determining the extension of psychological predicates beyond our actual experience.

But what if having consciousness suffices for having psychology, and consciousness is a fundamental (or at least natural) kind coming in a range of varieties that can get indefinitely weird (and we are tacitly aware of all this)?

Then against Davidson we deny (a), because the rails are in the nature of the thing itself, rather than in any response we make to it. And against Lewis we deny (d), because this fact about natural kind structure could be analytic.

(By contrast Lewis’s (d) can be made weak enough to be pretty trivial and Davidson’s (b) squares up pretty well with my sense of things: the charge of “irrationality” goes hand in hand with “what the hell is going on here”—though of course that is not to say anything substantive about what “rationality” is!)

B. Bayesian rational choice theory

This is the classical theory of (perfectly) rational psychology. (Origins: Ramsey 1926, von Neuman and Morgenstern 1944; Davidson wrote a book on this stuff with my old grandboss Pat Suppes in 1957.) It comes in three bits:

1. A part specifying synchronic conditions of rationality on an intentional state
2. A part specifying the nature of rational learning
3. A part specifying the nature of rational action

Lots of stuff is contested here and there are a bunch of ways to implement it and these details don’t matter so we’ll just give a very simple one.

1. The structure of a (perfectly) rational intentional state

Understood as a *credence-value complex*.

- One’s *credential state* is understood as a **probability distribution** over sets of possible worlds; [This can be defined more precisely—in many mutually incompatible ways, I believe—but we will leave the notion at an intuitive level]
- One’s *evaluative state* is understood as an assignment of numeric weights to “propositions” or sets of possible worlds [Subject to an **additivity requirement**: if a proposition can come about in a bunch of different ways, its value should be the weighted average of the values of those different outcomes: weights assigned by their relative likelihood.]

This imposes certain restrictions:

- If I think there are exactly two possibilities for this coin flip, it’s not OK for me to assign one of them credence .3 and the other .9, or one .3 and the other .2: they have to sum to exactly 1;
- If I think I can get an ice cream in exactly one of two ways, and I *just love* one of them *absolutely adore* the other, it’s not OK for me to find the general possibility of getting an ice cream *utterly hideous*: rather, my attitude toward the general possibility has to be somewhere between just-loving and absolutely-adoring; also not OK for “preferences” to be cyclic (pops out of the assignment of numeric weights).

But that's about it:

- No problem with assigning a higher valuation to *the entire world is destroyed* than to *my little finger is scratched*;
- No problem with:
 - Assigning a higher credence to *all the emeralds I've seen so far are green and the rest are blue* than to *all the emeralds I've seen so far are green and so are the rest*;
 - Assigning whatever credence you like to *time is discrete and finite* or *every event has a cause* or *wherever there are simples arranged table-wise there is a table* or *God exists* or *das Nichts nichtet*.
 - Assigning credence of 1 to *the chance of heads is .5* and assigning credence of .3 to *heads will come up*).

We might debate about whether to impose restrictions on what counts as perfect rationality that would eliminate various allegedly wacked-out views, how to formulate them, etc. Feel free to bundle in your favorite set of such restrictions as desired.

2. The structure of learning

Learning is “conditionalizing on evidence”. Picture the following: at t , I am uncertain whether E : I have some credence spread over E -worlds, and some credence (.4kg, let's say) spread over $\neg E$ -worlds. Then, by some later time t^* , I become certain that E (and nothing “additional” relevant to my credential state has changed). What should I do?

- Well first, I should get rid of all the credence that used to be spread over the $\neg E$ -worlds. There should now be 0kg of credence remaining outside of the E -worlds. This leaves .4kg remaining credence piled up on the workbench. Since my credential state has to be a probability distribution, I need to have 1kg credence piled around at all times. How to redistribute the remaining .4kg?
- Idea: the mere fact that I decided in favor of E shouldn't effect how I think it to be more likely for E to come true: if I earlier thought E could come true in either the red way or the blue way, and gave 50-50 to each, now that I've decided that E *will* come true (and not having learned anything else) I shouldn't crank my credence in red up to .8. So what I should do is *renormalize*: pile the remaining .4kg of credence in around that doesn't change the ratios of credence piled up inside the E -worlds (if 20% of it was in the H -worlds, they should have .8kg of credence spread around on them, and so forth).

Math for this: my credence that $H \neg C(H)$ —should now go to my *old* credence that H and E (the amount of H inside of the E -worlds), *divided by* my *old* credence that E (that's the effect of renormalizing). If we write credence that H and E as $C(HE)$ [short for $C(H \wedge E)$], we can use the following abbreviation:

$$\circ C(H/E) = C(HE)/C(E).$$

We pronounce ‘ $C(H/E)$ ’ as “credence of H on E ” or “credence of H given E ” or “conditional credence of H , assuming E ”. This is the so-called “definition of conditional probability”.

Next, if we write my old credences as C_{old} , and my new credences as C_{new} , we get the following:

$$* C_{new}(H) = C_{old}(HE)/C_{old}(E) = C_{old}(H/E).$$

(More cryptically and efficiently, we could write $C_E(H) = C(H/E)$, where C_E is the distribution of credence after learning E and C is the distribution of credence immediately prior to learning E .) The rule of modifying one's credence state in accord with this rule is known as "conditionalizing on E ".

Then, the claim is that **perfectly rational learning is conditionalizing on evidence**. Alternatively, one is rational to the extent that one updates one's credence state in response to confrontation with evidence by conditionalizing on that evidence. That is the characteristic claim of "Bayesianism".

What is this "evidence"? It has to be a proposition, and it has to serve as a sort of a posteriori impingement: the conditions on credence in (1), whatever they may be, already stipulate perfect rationality or perfect knowledge of that which is knowable a priori. It is natural to look to perception in this context to play the needed role. So the evidence is a proposition that is somehow related to perception. There is nowhere else in the picture to locate consciousness so that theorists of this stripe tend to regard "experiences" as generating the needed propositions. More on this below.

3. The structure of rational action

What is rational for one to do in a particular case? Perhaps, to act for the best. One might be crucially ignorant of or confused about certain facts, or one's system of values might be internally flawed in ways that escape one's ability to detect or improve. These are issues that confront all of us equally, so we might want a system of rules that "divide through" by such intrinsic flaws. If we had that then we could say that one's act was at least *instrumentally* rational or *subjectively* rational: it was the best thing to do given what I had to work with, or would have been the best thing had my credences and values been ideal, or something.

Maybe what this involves is *maximizing expected value*: choosing the act that is most likely (given your credences) to bring about the best outcome (given your values). Suppose one has a range of acts available at t : a_1, a_2, \dots . Let's say that $A_i = I \text{ perform } a_i \text{ at } t$. Then the "weighted risk" of a_3 -ing myself into w is my value of w , multiplied by my credence I wind up in w by a_3 -ing. (This will be zero if there is no chance of winding up in w by a_3 -ing.) The *expected value* of performing a_3 , then, is the sum of weighted risks of a_3 -ing myself into w , for every w ; or else the weighted average value of all the w , with weights being how likely I think the w is to result from my a_3 -ing:

$$* \quad EV(a_i) = \sum_w C(w/A_i)V(w)$$

Picking the available act with the highest EV is the rule of "value-maximization". The characteristic claim of (Bayesian) rational choice theory or decision theory is that **perfectly rational action is value-maximization**. Alternatively, one is rational to the extent that one's act at a time is that available act with the highest EV.

- (An internal dispute concerns whether "Newcomb problems" show that weights should be assigned not by conditioning but by a rule known as "imaging", which more closely tracks my opinions about the ability of a_3 -ing to *bring w about*, rather than my opinions about the likely correlation of being in w after a_3 -ing.)

C. Worries

- A. The theory is principally synchronic: requirements under (1) and, surprisingly, (3) provide *synchronic* constraints; in effect "consistency" requirements on momentary intentional states. The diachronic component of the theory, by contrast, is etiolated: the only respect in which

mentality distinctively “unfolds” over time is by the updating of credence in response to evidence, as specified in (2). For that matter, updating is only weakly diachronic: all that is necessary is for the contents of the prior and posterior states to stand in a certain formal relation of consistency. Nothing about the specific way in which one gets from the prior to the posterior credential state is distinctively mental. Nor is it distinctively progressive: updates are snaps from one state to another, achievements perhaps. Distinctively mental activities and accomplishments are hard to find.

- B. The theory is highly “rationalist” in the sense of being principally concerned with what happens “between” the experience and the action. Component (1) deals exclusively with formal requirements of consistency on C and V. Component (2) concerns experience; still, the focus is on consistency of the posterior state with the prior state and the proposition somehow associated with experience. Similarly, while component (3) concerns action, the focus is exhausted by the consistency of the belief about the action with credence and value at that time.
- C. Expanding on this point a bit, the theory leaves the nature of experience and action opaque.
- i. The sole constraint on the nature of experience is that it have somehow can be captured in a proposition, the entertaining of which serves as the input to future credence in accord with (2). Prima facie, it is not obvious that experiences are propositional, and it is not obvious that we have the sort of certainty in our experiences that (2) requires. If the proposition in question describes one's experience, how is it related to one's experience, in such a way that would motivate the claim that one has “learned” that proposition?
So if (2) concerns anything, it seems more like *immediate a posteriori recognition* than experience. But if we take this route, the nature of experience is left opaque: what exactly are my experiences like, and how are they related to the judgements that feed into the credence system? Perhaps we could find this out by examining the *contents* of these judgements in ordinary cases, but the theory itself tells us nothing about the contents of these judgements.
 - ii. The sole constraint on the nature of action is that it can be captured in a proposition known with certainty upon performing the action, and thereby be the momentary output of the collision between the credence and value systems as in (3). Prima facie, it is not obvious that actions are momentary, and it is not obvious that there are any actions that we can perform with certainty: none less than John Pollock noticed this, regarding the outputs as “conditional policies” rather than actions. People sometimes speak of “basic actions” to solve this problem: finger twitches, perhaps, or “trying as the pineal gland” as O’Shaughnessey once put it. But finger twitches and tryings can fail: “you’re not trying!” “sorry, I was *trying* to try”. If so this is no stable resting place; more worries below from Thompson. Moreover, the doctrine that trying to A is the only sort of *real* action, while actually Aing is only an action derivatively concerned insofar as the Aing has the right causal relationship to the trying to A is perversely ballistic: through basic-active flashes, we *launch into action*.
So if (3) concerns anything, it is more like *decision* than action. Still, decisions seem to be momentary events, and rather plausibly I can with certainty decide to A. But, once again, if we take this route, the link to action is left opaque: what are actions like, and how are they related to decisions to perform them? Perhaps we could find this out by examining the contents of these decisions—but, of course, the theory is silent about these contents.
- D. The theory ignores the pervasiveness in mental life of experience and action.

- i. Experience seems to pervade the remainder of the phenomena the theory addresses: it is not easily shunted into the leftmost edge. Looking here and there, I notice such and such, assess what it is, and come to judge it to be a snail. Reasoning about the theoretical impact of these or those considerations, I decide to follow certain trains of thought, mulling them over, trying out various techniques. Engaging in practical reasoning, I envisage various strategies, consider their outcomes, brainstorm alternatives, muster up the nerve to do the one I have proclaimed the best. Practicing an instrument, I try out this or that fingering, this or that rhythm. Biking to school, I guide myself in accord with my mental map of the area and my sense of what would be most interesting.
- ii. Action seems to pervade the remainder of the phenomena the theory addresses: it is not comfortably shunted off in the right-most edge of the mind. These examples of experiences are also examples of actions.

D. Historical remarks

The perversity of this picture should not be especially surprising.

The rationalist aspect of the doctrine is due to its stemming from investigations that are at bottom *logical*: that are concerned with the synchronic consistency of classes of propositions (and/or their representations), and building on this, with valid rules of implication among propositions (and/or their representations). Understood along these lines, logic does not care where the propositions come from or what will be done with them.

The background influence of logical theory as a paradigm on Ramsey, Davidson, and Lewis should be relatively clear. I'm a bit less certain what VN/M's background was, but the index of names in their bibliography is instructive, including a range of figures working on the axiomatic foundations of math (Dedekind, Brouwer, Tarski, Hilbert, Fraenkel, Zermelo), and both of them published papers in math journals.

More specifically, Bayesian decision theory was developed to serve the purposes of *economic* reasoning understood as reasoning about complex sorts of "games" (Ramsey, von Neuman and Morgenstern, Davidson and Suppes). Many games -- poker, chess, monopoly, clue -- have a number of distinctive features that are not present in the real world, stemming from the fact that they are rule-governed:

1. There is a class of meaningful actions that can be regarded for practical purposes as certain to succeed, namely the well-defined "moves" in the game. (Certain overall strategies might also be regarded as "actions", despite their uncertainty of success: eg, buying up the railroads or playing the Nimzo Indian Defense. A case of carrying out such a strategy can be plausibly regarded as composed out of its component actions.) The existence of such moves of course depends on their isolation within the structure given by the rules.
2. Time can be sensibly regarded as discrete for games that proceed in turns; accordingly, moves can be regarded as instantaneous (I raise \$200 on my turn).
3. There is a well-circumscribed class of moves available to each player at his or her turn.
4. The manner by which one reasons about what to do at each move isn't itself part of the game: to count as playing the game one must merely make moves that are permitted by the rules with the aim of achieving the official goal of the game. Accordingly, it is possible to simplify a description of what to do by collapsing the time at which the move takes place and the attendant reasoning that leads up to that move into a single instant.

5. The rules typically specify (at least implicitly) a class of facts which should be obvious to all players under conditions of normal play: Joan's shoe is on Atlantic Avenue, Bill's Electric Company has been mortgaged, and the like. Alterations in these can be regarded as experiences: one can treat oneself as certain that they obtain.
6. Patterns of mutual compatibility among states at turns and among courses of history composed of turns are well-defined and "trivial" to reason out a priori (in the sense that we can teach a computer to play chess).

Accordingly, the focus on games can enable us to avoid the sort of probing questions we have been addressing.

This is not a defect for the intended domain of the theory. Well-run financial and commodity markets backed up by legal systems can also be regarded as having these features, for instance bidding at auction or purchasing a certain quantity of steel from a supplier to be delivered at a certain date. But the extension to ordinary psychology is absurd: most of these conditions do not apply to most actions in the real world. For instance, when gardening, what counts as a move? What must I conditionalize on if I am to garden with perfect rationality?

The impact of this sort of economic thinking on Davidson is relatively clear: his earliest work in philosophical theory was on the foundations of Bayesian decision theory, and this formal approach serves as an acknowledged paradigm for his understanding of rationality. The impact of Davidson's approach on Lewis is also evident: Lewis in essence grafted the assumption that Bayesian decision theory is (part of) folk psychology into his pre-existing commitment to functionalism in the course of preparing a response to Davidson's 'Radical interpretation'. Lewis was also of course influenced by Schelling, who was at Harvard in Lewis's graduate days, and who provided the background for Lewis's initial investigations into game theory as applied to the analysis of convention.

II. Action as the fundamental practical psychological category

Now let's take a look at Michael Thompson's stuff on action. Central doctrines articulated concern the *metaphysics of action* and the *semantics of "practical psychological discourse"*—'wanna'/'tryna'/'finna'. The view is that actions are processes which ontically ground their parts and are mereologically gunky, while practical psych discourse concerns actions, but does not register a natural joint. As a result, the serious metaphysics around here sees actions as written into the world but psych-phenomena as shadows of language.

A. Metaphysics of action

The picture here is that the actions are among the processes; they have distinctive parts, and are prior to their parts; they are also "gunky" in the sense that every action has parts. In virtue of this "holism" they are thus distinguished from other processes, and cut nature at its joints.

A restriction: the doctrines are only intended to concern the accomplishments, actions which can be in a meaningful sense "incomplete". By contrast, the achievements are "over as soon as they have begun" (noticing Sandra), and the activities are "complete as soon as they have begun"—if Joan is skating, she has skated. [Question, what changes if we bring in the activities? Can the achievements be helpfully brought into the picture?]

1. The actions are among the processes

Argument, this is obvious prima facie:

[A certain advocated thesis seems odd due to] the tendency of students of practical philosophy to view individual human actions as discrete or atomic or pointlike or eye-blink-like units that might as well be instantaneous for all that it matters to the theory. Part of the present effort, then, is to break up such concepts. A person might, after all, spend a few years building a house, a few months raising an acre of cantaloupe, a few hours, baking a loaf of bread a few minutes playing a hand of poker—or a few seconds assassinating a political opponent. Any of these will make an apt illustration of the concept of intentional action, none more apt than any other. [fn: Notice also that the periods mentioned might be superimposed in a description of the activity of a single person. Having set a few more bricks this morning, and irrigated the melons this afternoon, I might pick off a passing peasant organizer as I sit on the veranda, waiting for the bread to rise and for my friend to place his bet.] If we reach for the last and shortest of these as our preferred illustration, as the one that makes that makes everything especially clear – and proceed to dwell, for example, on its supposed identity with an apparently unanalyzable moving of a finger, rather than its equally attractive and likely resolution into reaching for, raising, aiming and firing a gun, to say nothing of checking to see if the victim is done for and repeating as necessary -- it is, I will suggest, because we are moved by considerations alien to the philosophy of action, however legitimate they may be from the point of view of, say, a physiologist investigating "voluntary" as opposed to "reflex" movement. The nature of intentional action, or of the kind of being-subject-of-an-event that characterizes a rational agent and a person, resides in the peculiar "synthesis" that unites the various parts and phases of something like housebuilding, e.g., mixing mortar, laying bricks, hammering nails, etc. This synthesis is rendered explicit in naive rationalization. It can be exhibited, I will suggest, even in the moving of a finger. (91)

And so it is.

2. Action holism

The doctrine is that there is a certain respect of ontic priority or grounding such that actions are prior in this respect to their parts. (When S is part of B, B is ontologically prior in this respect to S.)

Note up front that often, one is/was S-ing because one is/was B-ing

- “small” and “big” actions;
- examples: crossing the street because one is getting lunch, attending seminars because one is studying for a PhD, and so forth;
- the “because” is the because of “final-clausal or purposive or ‘instrumental’ or ‘teleological’ formulation ... of *straightforward rationalization*”.

Then:

1. One’s S-ing is a part of one’s B-ing just when one is S-ing because one is B-ing
 - a. seems fairly obvious, I suppose; lacking any developed theory of parthood that would have interesting applications here we are not in a position to go much further than this
2. When X straightforwardly rationalizes Y, Y owes its being to X
 - a. Suppose that some process were unrationalizable: the S-ing occurs but for no B-ing is the S-ing *because* of the B-ing in the relevant sense. Then plausibly the S-ing would be arational

- b. Arationality is incompatible with status as an action, plausibly; so the S-ing is not an action.
 - c. If not, then when an occurrence is an action, this is plausibly grounded in its being straightforwardly rationalized by some other action
 - d. Since action-status is essential to an occurrence, a rationalized action is ontically grounded in the rationalizing action
3. So when one's S-ing is part of one's B-ing, one's B-ing ontically grounds one's S-ing.

3. Action gunkiness

Every action has parts/straightforwardly rationalizes some action.

Argue that: *moving my arm (pushing this rock) from A to Z* rationalizes *moving my arm (pushing this rock) from A to M/M to Z/H to R* etc.

1. Grant that if I *were* moving my arm from H to R this would be rationalized by my moving my arm from A to Z
2. I am moving my arm from H to R:
 - a. My arm does move so denying this would require some floor of smallness beneath which actions aren't found
 - b. The only reason to say this would be the assumption that A-ing requires the formation of an explicit intention to A;
 - c. But "skill or craft or techne often drives out deliberation"; moreover in the present case the characterizations of action are homogeneous, so if accept any are present to mind why deny any are? (111)

Disappointment: distinguish organic parts from trivial parts, where the latter exist solely in virtue of the mereology of space and time imposing an image on its inhabitants, while the former carve the whole at its joints better. The argument gets us trivial gunky parts but not organic gunky parts, which would be more interesting.

B. Semantics of practical psychological discourse

The picture here is that practical psychological discourse concerns to actions; but we use practical psychological discourse to satisfy a spate of messy pragmatic requirements, so this discourse doesn't track a natural joint in the actions.

1. Practical psychological discourse concerns actions

Actions are only ever explained by things in progress, namely unfolding actions (132):

For example, while we say:

- I am/was S-ing because I am/was B-ing
- I S'ed because I am/was B-ing;

we do not say:

- * I am/was S-ing/S'ed because I B'ed

unless we are not giving a *straightforward* rationalization, such as “I’m driving 15 miles out of the way because I missed the exit”—namely, we’re citing some fact which provides a context within which we can fill out a picture of the world as you are responding to it (fn7, 89).

Practical psych verbs are no exception. Wanna-finna-tryna are “imperfect markers”, like the progressive marker ‘-ing’. This explains their common tendency to take only non-finite verb phrases as syntactic arguments. Good stuff in fn11, 128 trashing language abuse to get propositional desire.

He seems to be positing the following similarity in their derivation:

For the progressive:

- to cross the street
- $\text{t}\emptyset$ cross the street + PROG
- be + [$\text{t}\emptyset$ cross the street + PROG]
- [be + TENSE + NUMBER + PERSON] + [$\text{t}\emptyset$ cross the street + PROG]
- Mary + [[be + TENSE + NUMBER + PERSON] + [$\text{t}\emptyset$ cross the street + PROG]]
- aka ‘Mary is crossing the street’

For ‘want’:

- to cross the street
- to cross the street + WANT
- to cross the street + [be + TENSE + NUMBER + PERSON]
- Mary + [to cross the street + [be + TENSE + NUMBER + PERSON]]
- aka ‘Mary wants to cross the street’

For the perfective:

- to cross the street
- $\text{t}\emptyset$ cross the street + PERF
- [TENSE + NUMBER + PERSON] + [$\text{t}\emptyset$ cross the street + PERF]
- Mary + [[TENSE + NUMBER + PERSON] + [$\text{t}\emptyset$ cross the street + PERF]]
- aka ‘Mary crossed the street’

The thought is that the way this works semantically is that the non-finite VP specifies a sort of “process type” which is represented by the remainder of the material as complete or incomplete, perhaps in a certain way.

The claim then is that since the psych-verb sentence clearly doesn’t represent the action as complete, it is representing it as incomplete in a certain way: a way somehow characteristic only of incomplete actions.

2. It doesn’t cut the actions at the joints

“Is there anything to choose between ‘she’s making tea’, and ‘she’s putting on the kettle in order to make tea’, namely ‘she wants to make tea’? Of course not” —Anscombe. (132)

We use this sort of language to mark how far away from completion the process is, but that’s all interest-relative.