**Rules and Rule Acquisition: From S-R to TNT**

Lee Roy Beach

(with help from Jim Wise and Larry Phillips)

In another essay on this site, I claimed that Edward Thorndike (1874-1947) is ‘psychology’s Darwin’. This is because of his ‘discovery’ of the Law of Effect, which states that the effect of an action (either annoying or satisfying) in particular circumstances determines whether it will be repeated in the same circumstances. He also ‘discovered’ the Law of Exercise, which states that the more frequently the action occurs in the annoying circumstances, the more strongly the circumstances and action are connected.[[1]](#footnote-1) Of the two, the Law of Effect was the more fundamental because it postulates a connection and how that connection comes to be established; the Law of Exercise is just a technical detail about what determines the connection’s strength. Together, these two laws formed the foundation for decades of research on learning, primarily rote learning. True, most of this research rather missed the point about what was going on, but it nonetheless it kept lots of psychologists employed and, in the long run, provided a solid empirical foundation for something very basic—rule acquisition.

Missing the point was a side-effect of the behaviorist experiment that preoccupied psychology in the Anglo-centric world from 1913 until the middle of the 1960’s. This experiment was an attempt to account for all behavior without recourse to anything that smacked of ‘mind’, which included both emotions and thought as causes of behavior. The thesis, reflecting the prevailing Positivist view of science, was that psychology was solely about observable behavior; emotions and thoughts were merely responses to specific stimuli like any other behavior. The basic units of analysis, the behavioral molecules, if you will, were simple input-output (stimulus-response, S-R) connections. Collectively called *reflexes*, these S-R connections included innate behaviors such as startle responses to loud sounds (which was the original meaning of ‘reflex’) as well as S-R connections acquired through experience and strengthened by outcomes (called ‘reinforcement’). Complex behavior was the result of sequentially linked reflexes.

I doubt that the founders of the behaviorist experiment wholly believed this extremely reductionist view or were inflexibly doctrinaire about it. I suspect that for them it was all an experiment, an attempt to see how much mileage could be got out of reducing everything to S-R reflexes. But, as so often happens in the world of ideas, those who followed took a more rigid and narrow view. The behaviorist movement began to resemble a cult in which its members were quick to identify deviations from the party line. As a result, it rapidly became *de rigor* for all psychologists, even psychotherapists, to avoid any wording that, however vaguely, implied that emotion or cognition played a determinant role in behavior. Everything had to be boiled down to S-R reflexes—to connections between stimuli and responses established by reinforcement (rewards) and strengthened by repetition (exercise). So called, conditioned reflexes.[[2]](#footnote-2)

As with most movements, the shine eventually wore off of behaviorism. As psychologists’ interests expanded beyond basic perception and learning, the ponderous S-R explanations required to account for complex behavior simply collapsed under their own weight. The futility was clear to anyone who bothered to look. And, in the 1960’s, almost as quickly as it rose to dominance, behaviorism began to fade (aided by the loosening grip of Positivism on science in general). Of course, there were stragglers and pockets of research in which the S-R lingo and research paradigms were still used, but it wasn’t clear what more could be added to the immense body of research that already had been done. Even more, it wasn’t clear how the old lingo and paradigms could address the problems that younger psychologists found interesting. Nature may not abhor a void, but science does. This one quickly was filled by a new, more sophisticated, cognitivism. Emotion and thought were readmitted, but this time there was little reliance on introspection, the fallible research technique that had, in part, spurred the behaviorist revolution in 1913.

This isn’t to say that the immense body of data that behaviorists left behind disappeared or that what was learned was rejected or forgotten. When everyone moved on, they carried that knowledge forward but were less slavish in how they talked about it. The word ‘reflex’ faded away, along with the reductionist and mechanistic view of behavior, and cognitive research took center stage. Now, it was acceptable to talk of mental states influencing behavior and to conduct research on judgment and decision making, on the nature of language and language acquisition, and all the rest. But, the lingering influence of behaviorism was clear in the avoidance of anything suggesting a mystical mind—even the word ‘mind’ was avoided. In this sense, behaviorism left its imprint, inclining psychology toward empiricism and objectivity and away from ‘squishy’ things, like Freudian psychology. Subsequently, cognitive research split from the rest of psychology, divorced itself from the American Psychological Association, and eventually merged with linguistics, neurology, and similar disciplines to form the new discipline of cognitive science.

As I’ve said so many times before, the notion of the human mind investigating the human mind, of the investigator being the investigated, is unique in science. There was no precedent to follow, no other science provided a paradigm. To a large degree, behaviorism was a response to this void, so when it disappeared in the 1960’s, the new cognitive psychology faced the same problem. It now embraced empiricism and objectivity, so it could not call upon the pre-behaviorism models, like Freud, which were too mystical and short on proof; indeed, it wasn’t at all clear if you could ever prove or disprove them. But, there weren’t any new models to rely upon. So, some psychologists used their new cognitive viewpoint and the growing interest in ecological and developmental research to create a new model in which animals, including, humans, are viewed as purposeful gathers and users of environmental information to formulate a concept of their individual worlds, a concept upon which they base their interactions with that world. This suggested a new metaphor, creatures as intuitive scientists.

The intuitive scientist metaphor turned out to be quite valuable for a while, particularly in thinking about how people maneuver their physical worlds and deal with uncertainty. The latter is what gave rise to the idea of humans as intuitive statisticians and the resulting controversy about the degree to which human judgment and decision making is rational. Whether the metaphor still has value is debatable, but it makes little difference because the computer metaphor soon became the lodestar for cognitive research; the brain as a computer-like information processing device.

In all of this, Thorndike and the behaviorists and S-R psychology became rather lost. They were seen as an early, but now closed, chapter in the history of cognitive science. But, I think that this perhaps a bit too dismissive. I think that they were on to something fundamentally important, something that got overlooked in the whole ‘reflex’ and behaviorism-as-a-cult misadventure. That something is rule acquisition and rule use as basic to human behavior. Not reflexes but rules. And these rules cover a range from the simple to the complex, from reflexive muscle twitches to thinking about the universe.

**Rules**

For me, rules and their acquisition became an important topic in 1966, when I became an assistant professor of psychology at the University of Washington. As the new guy, I was drafted to teach introductory psychology to large auditoriums of late-adolescents with short attention spans. The behaviorism experiment was waning and it was increasingly acceptable to speak other than S-R lingo. So, I tried to teach about learning without using that lingo, if only because using it quickly put the students to sleep (and they thought it was stilted and silly). The result of my efforts to make things more accessible to my students was published as the learning chapter in my first book, an introductory textbook (Beach, 1972). My take then is still my take nearly 50 years later (Beach, 2010, 2019; Beach, Bissell, & Wise, 2016). Only now it is in the context of a wholly cognitive theory about the structure of conscious experience called the Theory of Narrative Thought (TNT), an almost complete reversal of what I had been taught as an undergraduate in an S-R world. In what follows, I will briefly outline what I said in that long-ago textbook and tie it to TNT. In this, I will start with the S-R interpretations and then translate them into TNT terms as acquired rules rather than conditioned reflexes.

**Kinds of Conditioned Reflexes (Kinds of Rules)**

The two kinds of conditioned reflexes are *operant*, for which Skinner’s lever-pressing rat is the prototype, and *classical*, for which Pavlov’s salivating dog is the prototype. Historically, classical predates operant, but it is convenient to discuss operant first.

TNT Translation: There are two kinds of rules, *what-to-do* and *what-to-expect*, roughly corresponding to operant and classical conditioning respectively. Further, there are two kinds of what-to-do rules, the first of which is of the form “In this situation, do such-and-such”; written S1→A1, where the S1 stands for a specific situation and the A1 is action, the such-and-such that is to be done.[[3]](#footnote-3) This rule merely tells you what to do with no reference to the results of your action. This is the prototypical physiological reflex; someone yawns and you yawn, you touch a hot surface and you pull your hand away, or a nearby car honks and you flinch.

The second is of the form, “In this situation, do such-and-such and then so-and-so is likely to happen”; written, S1→A1→S2, where S1 is the specific situation, A1 is what is to be done in this situation to produce S2, which is the so-and-so that is likely to happen. This is the rule Skinner’s thirsty rat was learning in the ‘Skinner Box’. Thirst and the box with a lever constitute S1, pressing the lever on the box’s wall constitutes A1, and the few drops of water that appears below the lever constitutes S2. Unlike the first kind of what-to-do rule, this second kind includes the expected results of executing the correct behavior and, in doing so, constitutes a *prediction* of those results.

There is just one kind of what-to-expect rule (aka classically conditioned reflex); “If this occurs, that follows”; written, S1→S2, where S1 is the situation and S2 is the new situation. That is, the rule does not involve any action on your part. So, for example, Pavlov’s dog learned that when a bell rang, food would follow. Consumption of food requires the dogs’ mouths to be moist, so dogs typically begin to salivate before eating (which is a S1→A1 rule, a physiological reflect). Pavlov found that, after a few pairings of the bell and food, the bell itself (S1) elicited salivation in anticipation of the food (S2). In effect, the bell told the dog to expect food and to be prepared for it by salivating. That is, a S1→S2 rule constitutes a *prediction* of events predicated upon past and current events.

So, the upshot is that both the second kind of what-to-do rule (S1→A1→S2) and the sole kind of what-to-expect rule (S1→S2) constitute predictions about the future, one conditional upon your doing something and one just because that’s usually what happens.

**Conditions that Promote Conditioning (that Promote Rule Acquisition)**

1. Motivation. The point of a S1→A1→S2 rule is to do A1 to produce S2. This implies that you must value S2 to some degree, and the more you value it, the more prone you are to learning the rule for obtaining it. This means that motivation has two aspects: First, it prompts you to learn the rule and, second, it prompts you to use the rule when the conditions (S1) are right.

2. Transfer. When in a situation that is similar too, but not exactly the same as one for which you have a rule, the old rule is a good beginning point for a new rule for the new situation. Indeed, the old rule may be sufficient, called positive transfer. But, sometimes similarly appearing situations actually aren’t, or the differences are somehow crucial, so the old rule does not work or even may cause damage, called negative transfer.

3. Shaping. When an old rule does not work well in a new situation, it is modified in terms of how closely its results approximate the valued results. Iterative changes in the old rule transforms it into a new rule that is suitable to the new situation.[[4]](#footnote-4)

4. Feedback (Reinforcement). For either an S1→A1→S2 rule or an S1→S2 to be established or modified, the results (S2) must occur quickly enough to make it clear that the behavior (A1) or prior event (S1) caused it, rather than some outside influence. Timing is particularly important for S1→S2 rules. Think, for example, of trying to discover what food might have caused an allergic reaction—unless the reaction occurs pretty quickly after eating the food, or the food is particularly distinctive, it is difficult to figure out what caused it.

**Conditions that Promote Change or Abandonment**

1. Generalization and Discrimination. The application of rules is not a rigid thing. For example, each time you use an S1→A1→S2 rule, the S1 is not exactly the same as in the past, there usually are alternative ways to do A1, and S2 seldom is precisely the same as it was last time. So, if you are thirsty, S1, it may be from exercise, or because you haven’t drunk enough water that day, etc. And, you can deal with it (A1) by drinking water, soda, coffee, etc. or, maybe, even sucking on a lemon. And the quenched thirst (S2) can be complete or partial, satisfying or not. This allowable variability is called stimulus and response generalization by the S-R psychologists and simple (rule) adaptability by the rest of us.

By the same token, if there are meaningful differences between the current situation (S1) and previous situations in which the rule worked, it is necessary to learn to discriminate between them and adapt one’s rules accordingly. Learning the difference is called stimulus discrimination by S-R psychologists and adapting behavior to fit the situation is called response discrimination.

2, Extinction. Sometimes a what-to-do rule fails to work because things have changed, usually because the A1→S2 part of it is no longer applicable—the action no longer produces the valued result. This can be because the change is permanent (somebody changed the lock on your office door, so using you key won’t work anymore) or because it is temporarily suspended (road repairs necessitate you to take a different route to work for a while). The usual response to a rule not working is to try again, and maybe even again. When it becomes clear that something has changed, you give up the rule and look for another way forward or forget obtaining S2 altogether. Temporary suspension of the rule, like the traffic diversion, prompts temporary suspension of the rule, but you’ll probably try every now and then to see if the rule works again. And, (it is difficult to imagine when this might happen), if the R2 is only available now and then, either randomly or at fixed intervals, you’ll adopt a strategy to suit. Randomness may eventually lead to abandonment of the rule, but frequently the occasional triumph of being right may keep it in place long after its long term usefulness is overwith—this is what keeps gamblers going. And even an abandoned rule may pop up again from time to time just to see if it is valid again.

**Origins of Conditioned Reflexes (Rules)**

In that ancient introductory textbook, I described a continuum of ways in which rules are acquired (p. 80). It was:

**Reflexes Instincts Predispositions Experience Imitation Intellectual**

S1→A1 S1→A1 ? ? ? S1→A1→S2 S1→A1→S2 S1→A1→S2

S1→S2 S1→S2 S1→S2

*Reflexes* are exactly what you know them to be, innate automatic physiological responses to specific classes of events—like Pavlov’s dog’s salivation in preparation for eating. Every animal has such rules and some have only them; usually the lower on the phylogenetic scale the animal is, the more of its behavior is the result of S1→A1, what-to-do reflexes.

*Instincts* are a step beyond; more complex and the rule’s application is less blindly automatic. Mating rituals are good examples; they’re fairly formulaic but are adaptable to circumstances.

I’m not too sure about *predispositions*, but if they are things like humans’ predilection for hostility toward strangers, they are a step above instincts in that they often can be unlearned or at least tempered when the context requires.

*Experience* is the origin about which the most is known, thanks to the S-R psychologists’ research. All of their ‘conditioning’ experiments were exercises in acquisition of experience-based rules. This body of research was expanded by neo-S-R experimenters who studied the conditions contributing to learning by *imitation*. So, we know that many creatures can learn by watching what some kindred creature does and humans can learn from verbal instruction (although in both cases, learning is surer if the learner actually practices the skill after watching or being instructed).

And, finally, the least understood and least researched origin, *intellectual* rules. The paucity of understanding is the result, I think, of the problem of the mind studying the mind; it has been difficult to get a clear grip on this kind of learning. Just look at how we appraise intellectual learning in the classroom, by examinations that focus on recall or application of specific sets of rules, like algebra or geometry, but seldom is the focus on use of broader knowledge in various contexts (in this sense, plumbers get a better education than undergraduates). It is very difficult to study people’s ability to reason without reference to some kind of artificial system, such as logic or statistical inference—both of which are tools devised in the first place to help us do intellectual tasks at which we know we aren’t particularly adept. The real intellectual feat here is with the folks who devised the tools, not college students trying to use them. Anyway, lest I digress, less is known about the formulation of intellectual rules than one might wish.

**TNT**

Recall that TNT is based on the idea that experience is structured as an ongoing narrative and that the elements of the narrative are events and temporally linked causal rules. That is, the organizing principle that provides meaning and direction to our thoughts and actions is temporally structured causal links between experienced events. The part of this narrative that is about past events is studied as memory. The part that is about the brief and transient present, is studied as perception. And the part that is about the as-yet-nonexistent future, is studied as inference and imagination. It always has been assumed that by studying the parts, cognitive science would come to understand the whole. I doubt that this will happen spontaneously. When everyone is specializing in studying a part, or even a part of a part, there isn’t anyone to study the whole. After all, given the nature of research, the various literatures are unlikely to fit together seamlessly, so seeing the ‘big picture’ is difficult, especially for those most closely involved. (Recall the old story about the blind men and the elephant.) As a result, it seems to me that the big picture has not emerged and is unlikely to emerge from simply generating more and more specialized data.

Rather, I think the time has arrived for theorizing. That is, instead of remaining in its adolescent data-driven state, cognitive research should grow up, should mature into a theory-driven enterprise. And those theories should be data-based but at different conceptual levels above data, drawing together different parts of the parts and then drawing those larger parts into even larger parts until we have a broad, data-based, unified theory of cognition that meshes with the larger body of science. This will take time and it will require imagination, something not greatly prized by an exclusively data-driven approach.[[5]](#footnote-5)

TNT is, from my viewpoint, a step in this direction. It is based upon my understanding of cognition, reflecting years of trying to see the big picture as I taught introductory psychology classes and conducted research in judgment and decision making. I have no illusions that it is the last word, or anything close to it. But it is a start.

**So, Let’s Start**

The rules in TNT include, but go beyond, the rules that, starting with Thorndike, the S-R psychologists explored so thoroughly. They go beyond because they are more flexible, more contextual, and many of them are the result of thought and discovery and can be modified or changed by more thought and discovery. This is not to say that the actor, you or me, has a mind of the mysterious sort (which has no place in a science, except perhaps as the object of study in its own right). Rather, it is to say that the process by which simple rules are learned—the kind of rote rules studied by S-R psychologists—also allows for richer rules that reflect back upon themselves and in doing so provide thought, discovery, modification, and change.

There are two keys to this. The first we share with other animals, your cat, for example. Almost all animals, but particularly those with more complex brains, are capable of acquiring rules and of modifying or replacing the rules in light of feedback about success or failure. This was well explored by the S-R psychologists in the form of schedules of reinforcement, extinction, and stimulus and response discrimination. It is part of the wiring, as it were, and it was Thorndike’s genius that led us to understand it in tiresome detail.

The second key is language. In fact, for humans (and other animals insofar as they have some sort of abstract communication system) language is the game changer. Evolution of the ability to invent, learn, and use complex language for communicating with ourselves and others made humans different from other animals. And, as we learn to internalize communications with ourselves, it becomes thinking—the little voice in your head is one part of what happens, but language-enabled thinking is a good deal broader.

The flexibility afforded by language is abetted by something even more basic, *imagination*. As I’ve said elsewhere (Beach, 2019, Beach, Bissell, & Wise, 2016), imagination is not well understood, it gets mixed up with all sorts of shadowy stuff. But it is, in fact, a useful, rather commonplace tool for embellishing what we say to ourselves and others and, most especially, for anticipating the future as an extension of the narrative past and present. Indeed, anticipation of the remote future is almost completely a function of being able to use language to construct plausible scenarios, or stories, about how the future might play out (Beach, 2021).

Together, language and imagination are the engines of human thought. And the thought they create remains true to the experiential structure that provides its content—the underlying temporal/causal prime narrative.[[6]](#footnote-6) In this sense, the rules that structure the prime narrative also structure thought and communication with others; the events and causal relationships among those events give meaning to our experience by providing context for new experiences and thoughts and stories to tell others, giving form to how we understand what has happened to us in the past, what is happening now, and what will happen next.

Which brings us back the last two categories of rules, imitation and intellect. Any time you learn a rule from someone else, it is imitation. If they demonstrate or describe or simply prescribe, the acquired rule falls in this category—it comes from someone else. So, for example, outside the classroom, TV is the greatest popular source of imitated rules; books come next but far more people watch TV than read books. TV is credited with having changed many people’s rules about homosexuality, for example. ‘Will and Grace’ and ‘Glee’ made gay characters less alien and more sympathetic. Soap operas, whatever one may say about them, have had a huge impact on attitudes about formerly-undiscussed topics, mental illness being a prime example. Perhaps the starkest example is the difference between CNN (or MSNBC) and Fox in reinforcing their viewers rules for interpreting politics and in some cases changing then substantially. And, it can be argued that social media is replacing TV as the great educator, perhaps to society’s disadvantage because (as is also true of CNN and Fox) we only consult congenial views—rules similar to those we already have, thus strengthening our rules and our belief in them.

Any time you come up with a new rule that is sufficiently congruent with your other rules and, perhaps, follows as a consequence of them, it is an intellectual rule. It may seem as though it appears in a flash, an insight, out of the blue. But, it isn’t. It is rooted in your other rules, in your prime narrative, and elaborated by language. Elaborated in that it is couched in words and is constrained by the rules of syntax; it names are used and its causality is given substance in terms of transfer of energy, meaning, or some other form of leverage. People like to speak of flashes of inspiration, but you’ll notice that no plausible inspiration is without ties to other aspects of experience. This is what is meant by inspiration being a gift to those prepared to accept it; it isn’t isolated from everything else.

**Rules and Uncertainty**

No rule is completely reliable. Reflexes usually do what they are designed to do, but not always. Past experience does not always provide a sound prediction of future experience. Rules others teach you are not universally applicable and don’t always work. And, intellectual rules, even the most brilliant inspirations, are not sure thing. So, part of rule acquisition is an acquisition of an understanding of the rule’s reliability. This doesn’t necessarily mean repeated trials to see how well it works, although past successes and failures with the rule certainly are important. Frequently, it depends upon how closely the present situation (S1) approximate the situation for which the rule was formulated (stimulus discrimination). Other times it depends upon an understanding (prediction) of the interference of outside agents and similar contingencies—in effect, the resiliency of the rule. And, sometimes it is a matter of ‘luck’ or fortune or chance for which you have no foresight or control.

However it arises, reliability translates into *certainty*, into the degree to which you are willing to depend on the rule. Reliance means you think the future predicted by the rule (S2) will, in fact, occur—and that this future can be used as a starting point to predict what happens after that, and after that, and after that, on into the remote future. Each stage in this projection, where a predicted event is predicted to occur and cause the next event, is another point for certainty—or its complement, *uncertainty*. In principle, overall certainty about the entire projection is the result of these individual points of certainty.

The subjective probability research of the 1960’s and 70’s, and the subsequent work on causal Bayesian models of probability in the 90’s and 20’s are attempts to understand uncertainty. Doing so is important because uncertainty plays such a crucial temporizing role in behavior, dictating how boldly or cautiously we face the future, how willingly we commit to our expectations, how much we are prepared to invest and how much we hold back. Much has happened in the past 60 years on this front, particularly with Bayesian theory providing the theoretical foundation for modeling uncertainty (not to mention its successes in political election predictions, disease treatment efficacy, business planning, and even underpinning quantum mechanics).

The ‘story line’ of the prime narrative is the chain of its most reliable, strongly linked causal rules. If even these strong links are not particularly strong, if they are of low to moderate

reliability, the story line is weak; things do not hang together well, life is confusion and you don’t have confidence in your understanding of how you got where you are, what is going on now, or what will happen in the future, Consider how you feel in a wholly new, exotic, situation for which your past experience has not prepared you; you usually try to escape but, failing that, you tread cautiously because your familiar rules may not work. (On the other hand, a little strangeness can be exhilarating, which is why some of us like to travel to foreign lands.)

So, from the TNT point of view, narrative consists of temporal/causal rules and the accompanying uncertainty about the efficacy of those rules. Efficacy for providing a ‘tight’ prime narrative; uncertainty makes weak causal links. Efficacy for providing dependable predictions about the immediate future and beyond; weak causal links make the whole chain weak.

It is important to be clear. Yes, the narrative consists of rules and uncertainty about those rules. But the rule is a construct and while uncertainty is an emotion related to the reliability of that construct, it is not a property of that construct. Which is to say, the rule connecting two events in the prime narrative does not include uncertainty; uncertainty doesn’t enter the picture until the rule is applied and is a function of the context as much as it is a function of the rule’s history of accuracy or any other indication of reliability. Rules and situational reliability may be learned in tandem, but they nonetheless are learned separately. Moreover, while causal Bayesian models are a method of accounting for and predicting uncertainty, their architecture is not necessarily the same as that of subjective uncertainty.[[7]](#footnote-7), [[8]](#footnote-8)

*An Aside on Emotions*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_­

*Uncertainty is an emotion but, needless to say, it isn’t the only emotion. I’ve discussed this topic at length in Beach (2019), so I won’t repeat it here. But my colleague, Jim Wise, recently sent me the following in the context of discussing an article we’d read about the neurology of threat detection. Because emotion is interlaced throughout TNT, his comments are especially enlightening:*

*“Emotions are the great mammalian invention. They freed mammals from a hard linkage between the environment and their behavior. Remember that life did not drop into the environment. It sprung from it. Life may not have a purpose, but it does have a trajectory, and that is to get as far away, and as quickly as possible, from the crystalline matrix from which it emerged.*

*“The first stages of life, fossilized as stromatolites, were a total symbiosis between proto-bacteria and the rock on which it lived. More a smear than a succinct organism.*

*Then came the Ediacaran fauna, matts and fronds with no sensing or digestive organs, absorbing nutrients from the waters within which they lived, in total equipoise.*

*Then came the “Cambrian Explosion” of life some 350 MYA. All this due to energy, fueled by oxygenation of Earth’s atmosphere, which substituted aerobic energy exchange for anaerobic – a 17-fold gain. (Life is an energy game. How it is captured, transferred, concentrated, put to work).*

*Then came reptiles and insects, which are tied to the environment energetically, even their gender being determined by outside temperature. They are really good at having a long list of chained reflexes, but they’re still tied to conditions of the surrounding environment. It is as if the environment actively resisted loosening its hold on life.*

*“Then came mammals and something completely new, (evolved) emotions, which put the break on environmental control by loosening it. Instead of tying specific environmental conditions to specific responses, emotions produce a more global ‘state’, a flush of chemicals and activations, that affects lots of internal processes, putting the organism in a complementary ‘state’ in which it can respond in a broader variety of ways: flight or fight or mate or do nothing.*

*“By the way, findings of studies of emotions boil down to three dimensions, valence (positive-negative), arousal (high-low), and magnitude (a lot-a little). In TNT, valence is related to violations of standards being good or bad and magnitude is related to the size of the violations, which dictates the motivation to act.  Arousal is related to how much action is required. All the different names we give to emotions are merely labels we learn to attach to various combinations of valence, magnitude and arousal.*

*--Jim*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Rules and Causality**

Causality is fundamental to rules because rules exist solely to explain (answer the question, ‘Why?’) and to predict (answer the question, ‘What next?’). Forests have been sacrificed to tomes on causality, one of philosophers’ favorite topics. But, I will spare us both a review of all they’ve said and boil it down to what is relevant to TNT.

First, Bayesian models of causal probability, which provide an abstract way of thinking about the architecture of uncertainty, use a network of events (nodes) connected by probabilistic causal links (edges). If this brief description is accurate, the models conflate causality and uncertainty. This doesn’t negate their value, but it is important not to make the conflation when discussing TNT.

As I said above, the events and links (nodes and edges) in the network are separate from uncertainty; uncertainty (an emotion) is mapped onto the network as a function of the context. That is, the links between events exist prior to the appraisal of uncertainty; their links are why they are included in the network in the first place. Then uncertainty is appraised; uncertainty which is conditional upon the circumstances in which the network is deployed. If this is the case, then we still have to understand why some events are linked to some events and not others; merely linking everything and then assigning zero probability (certainty) to some links may work mathematically, and would be okay for TNT because the prime narrative is ‘structured everything’ in your experience, but it really doesn’t get to the heart of the issue. It doesn’t say much about the nature of the link, just how uncertain the link is. Indeed, the last sentence doesn’t even make sense because links aren’t themselves certain or uncertain—certainty is not a property of events, it is an emotion held by a sentient being, by the actor whose cognitive structure is being modeled by the network. After all, the point of the exercise is to understand subjective uncertainty, not to indulge in modeling for its own sake.[[9]](#footnote-9)

So, what is causality and how does it apply to rules and links? I suspect that the most basic kind of causality involves the transfer of physical energy from one object to another (think croquet). Although that certainly applies to some of the rules in your prime narrative, as you matured the relative proportion of such rules decreased—social rules do not involve a transfer of energy in the same sense as a croquet ball hitting another croquet ball.

Some have suggested that causality in Bayesian networks are *implications*, X implies Y, and that is the way that it has been used in TNT; the past and present imply the future.

Another suggestion, by my friend Larry Phillips, is that causality is *relevance*. That is, S2 is relevant to S1 and therefore it is a better bet that it rather than an irrelevant event will occur when S1 occurs. Similarly, S2 is relevant to something you do, A1, so it is a better bet than an irrelevant event if you do A1.[[10]](#footnote-10)

Indeed, each of these concepts of causation (transfer of energy, implication, and relevance), and perhaps others as well, may be a profitable way of characterizing causation under different circumstances--with probability (or something like it) being a measure of the user’s certainty about the reliability of the rule in all circumstances. Note that probability is not causal, it is a measure of uncertainty about the strength or reliability of the causal relationship between events, not the strength or the reliability itself.

But, we are not philosophers, so, while nice discriminations among different kinds of causality are interesting, perhaps even important, it is convenient when discussing TNT to simply lump them into one term, causal. So too, as intriguing as the nuances of Bayesian theory may be, it is convenient when discussing TNT to simply refer to certainty about the reliability of causal links as rules for predicting future events.

**References**

Beach, L. R. (2019). The structure of conscious experience. Newcastle-upon-Tyne, UK: Cambridge Scholars.

Beach, L. R., Bissell, B. L., & Wise, J. A. (2016). A new theory of mind: The theory of narrative thought. Newcastle-upon-Tyne, UK: Cambridge Scholars.

1. I put ‘discovered’ in parentheses because both laws were simply common sense to anyone who had ever trained a pet, had children, learned to ride a bicycle, or learned to skate. What makes him psychology’s Darwin is his recognition of their ubiquity and his attempt to formulate them precisely. Unfortunately, his penchant for stating them in slightly different ways from one time to another rather reduced their precision, but the general idea was what was important and what proved formative for experimental psychology. [↑](#footnote-ref-1)
2. The term ‘conditioned’ derives from a mistranslation from the Russian; Pavlov meant ‘conditional’ but it was translated as conditioned. He wanted to say that the bond between the S and the R was conditional upon the reward. So, we got stuck with the awkward and obscure ‘conditioned reflex’, both parts of which are misleading. [↑](#footnote-ref-2)
3. In previous essays and books on TNT, I have used S for situation and R for action when discussing rules. But, starting now, I will replace the R with A for action to emphasize the difference from S-R psychology’s S1→R1 reflex responses. [↑](#footnote-ref-3)
4. Jim Wise points out that this is pretty much the definition of a feedback loop. [↑](#footnote-ref-4)
5. I should mention that I think the computer metaphor has run its course. It may be useful at some level of enquiry, but not at every level. As we move from the basic neurological architecture of cognition to the architecture of action, the metaphor wears thin. I think the narrative metaphor is better for these higher-level conceptions of cognition. [↑](#footnote-ref-5)
6. Recall that the prime narrative is the temporal/causal structure of your reality, the record of your past and present experience as well as the future that they imply. It is, in effect, your intuitive knowledge about how you got where you are in your life and what you can expect to happen next. Derived narratives for thought and communication with others draw their content and structure from the prime narrative but are abbreviated versions of it that fit the context. [↑](#footnote-ref-6)
7. Just as the architecture of TNT is not necessarily the same as that of structured experience. We must not conflate theory and reality. We know about the theory but all we know about reality is what our experiments reveal, and those experiments are shaped by our theory—even if that theory is not explicitly stated. The result is what we know is biased by our theory, but what else can we do? We limp on, but even then must never believe our theory is gospel, is somehow The Truth, because it makes us inflexible and unwilling to change when change is required. In short, we take our story line too seriously and are overconfident and blundering when caution is required.  [↑](#footnote-ref-7)
8. Larry Phillips reminds me that physicists keep trying to make probability into a concrete thing when it is, in fact, an emotion, it is uncertainty. Nobody can point to probability in the physical world, and it is substantively different from the other ‘things’ in their theories. As Larry says, “If probability is a thing, I’d like some please.” [↑](#footnote-ref-8)
9. Jim Wise suggests the following alternative for this paragraph: “And, here’s how I would rewrite it:

   As I said above, the events and links (nodes and edges) in the network are distinct from the feeling of uncertainty, which is emotionally and deep conditionally based. That feeling of uncertainty is mapped onto the causal narrative network as a function of the experienced context. That is, the links between events exist prior to the appraisal of uncertainty; those links are why they are included in the network in the first place. Then uncertainty is appraised; uncertainty which is conditional upon the circumstances in which the network is deployed. If this is the case, then we still have to understand why some events are linked to some events and not others; merely linking everything and then assigning zero probability (certainty) to some links may work from a relative frequency perspective in getting to probability. But there are other ways of doing that, through relative similarities, for example. (See Wise, 1970, Wise and Mockovak, 1973.) And judgement of similarity could well depend on relative activation recruitment of nodes into flexible networks, again depending on the experienced context. Relative similarities can meet the mathematical requirements for probabilities given that certain invariance conditions are met in their generation. So, getting from an emotion based ‘feeling of uncertainty’ to the exposition of probabilities by a sentient being is not an unbridgeable gulf. And doing so may be exactly what makes us human. What appears certain is that as the late naturalist, Loren Eiseley observed-- “In becoming human, we have given up the one thing that can never be regained: The certainty of the animal in that what it senses is exactly there in the shape the eye beholds.” In that case, by creating Narrative Thought, telling stories about the world that we then live in, we have forever removed ourselves from it.” [↑](#footnote-ref-9)
10. Jim Wise comments: ‘It seems to me that ‘implication’ is an outside→ in view, and ‘relevance’ is an ‘inside→ out’ view.’ [↑](#footnote-ref-10)