On generalized reading

BY SUSPENDED REASON & CRISPY CHICKEN

8 Letters · 0 Subscribers · 293 Reads · Updated 07 Jan '22 · Started 02 Jul '21

Subscribe Discussion

01 Letter 1

BY SUSPENDED REASON Created 02 Jul '21

I.

Sure, in the beginning was the word—but it wasn't written, it was read.

Let us suppose that meaning is entailment, a pragmatic implication. *So what? What's it to me?* And for an organism struggling to survive—struggling to parse its environment for resources which it can take within its boundaries to improve its condition, or for dangers which upon penetration (a claw, a chemical) weaken it—salient information abounds.

How do we know that there is reading without writing, reading before writing? Because we read the rocks and stars, the sun and moon, the tides and rivers—forces which, even if they could know our presence, and become bards for us, would be too indifferent to bother. In the beginning, there was information: a difference that made a difference, a sign that was salient.

(What makes a sign salient? It alters the probability distribution of events which concern us.)

A salient sign acts as a decision-rule: *if this, then that*. Behaviorism erred because it misunderstood and underestimated the complexities and contextualities of our stimulus-response patterns, mediated as they are by extensive experientially-built schemas which charge the inputs with meaning. But its basic instinct—that we observe in order to act, and act on the basis of our observations as they are interpreted by pre-existing beliefs—that was sound. What follows next is that our perception and ontologies are twin guides to decisions and actions: a field's technical vocabulary is precise or ambiguous exactly according to its need for distinction or its leniency to conflate. A difference that makes a difference; American pragmatism meets semiotic *différence*. We sense signs that point to some taxonomic identity; this identity carries a set of affordances or constraints in relation to our goal.

What happens is an animal learns to eat a leaf, based on its signs, its appearance, and eventually, through evolution, the leaf in turn "learns" to write itself differently, to hide. An insect "learns" through evolution that its predator "reads" by motion detection, and so starts writing out its existence through stillness. Evolution turns one of its great tricks, invents intelligence that is adaptive, dynamic, full-bodied strategic—that can encounter brand-newness and tailor a strategy to its newly learned attributes—*learned*, without scare quotes. We end up, in Spielberg's *Jurassic Park*, with the paleontologists standing stock-still in front of a T-Rex that is breathing down their necks, promising to devour them if they

flinch. There is no co-evolution here to teach them; how could they write this scene for their predator, unless they knew and knew *how* they were being read, in other words, could read their reading? Theory of mind, the reading of reading, gives way to the ability to dynamically, adaptively write. In lower organisms this stimulus-response pattern—if X, then Y—is "dumb" and automatic; in higher organisms, it is "smart" and adaptive, can pick up on subtle nuances of the stimulus in relation to a (historical, cultural, psychological etc.) context, can enter recursive hypotheticals modeling how they are read, how their reading is read, how their reading is written against.

II.

And as Sontag has pointed out (among countless others), "no style" is a style; there is no such thing as a "neutral" pose because there is no escape from being read. We must make choices in order to live—we must leave the privacy of our bedrooms, and choose clothes to put on, and go out into the world; we must take the metro, or a bus, or a town car, or walk; we can drive ourselves, or be driven; we can purchase coffee (so many options, on the menu board) or we can make it ourselves (drip, machine, K-cup, Chemex; hand-crank or electric grinder; whole-bean or pre-ground; bleached or unbleached filters). When we look across the train car at a young woman in silver moon-boots, when the magician performs a card trick requiring misdirection, our reading is read; this second-order reading, like all reading, serves the purpose of acting. I point at a spot on the map, and make sure your eyes follow, to ensure you understand my reference, to determine whether I need to re-iterate the pointing.

We are not yet even at the workplace and we have produced an enormous informational field to be interpreted, information that, even if at the object-level it is irrelevant to others—bleached or unbleached filters matter little to anyone—reflects deeper attitudes, preferences, personal mythologies (to Barthes)—a comfort with artifice, a preference for the natural. And so long as we may run the possibility of being advantageous or disadvantageous to others—as long as our behaviors and beliefs are salient to the preferences, desires, and goals of others—and that possibility is always latent in social space; ecologically co-present organisms' outcomes are always interdependent; we are always potential threats or allies—then our actions, tailored even as they may be primarily to some pragmatic purpose like staying warm or keeping dry, caffeinating or saving time,

beating traffic or saving money, will inevitably be read, are sources of salient information for those around us.

And it is the next step in this social or ecological interdependence that, even as we attempt to accomplish asocial ends, we will constrain ourselves, and allow our actions to be influenced, by considerations of how we (believe we) will be read. Pragmatic and interpersonal value combine to form a single payoff function, and others' actions—which are salient to our own interests—are premised on *their* perceptions and beliefs, which we are partly responsible for, which include as a subset others' readings of us. And there is no escape from getting read, just as abstaining from a vote is a kind of voting, and nudity in public is either a fashion scandal or a sign of insanity (a signal to others that we *ought* to be locked up, *lest* we cause harm to others). *If this, then that*.

02 Letter 2

BY SUSPENDED REASON Created 02 Jul '21

III.

Let us not speak of signals, then, but of interpretation and metonyms, ecological co-presence; choices which will either convey to our benefit or detriment. "Signaling" implies an out-of-the-way action whose primary or entire purpose is communication of some attribute. We pay exorbitant medical bills for an elderly parent (the argument goes) to signal that we care. But equal in this equation is the corollary: we can choose to forgo paying, *and this too will "signal*." The option is always in front of us, and there is a first-order benefit—the asocial, optics-free, intrinsic payoff, irrespective of others' notice—and a second-order benefit—the social, opticratic, extrinsic payoff, irrespective of reality. That is, the fuller picture of human social life, stretching beyond so-called signaling regimes, is one in which all our actions *already* communicate attributes, and this secondary aspect—atop a more pragmatic aspect, like comfort or mechanical function—is factored into our choosing. Every move we make leaks information; the question becomes which information we wish to

reveal (or project, falsely). We watch for, and mitigate, our “cues” as much as we “construct” “signals”—which is why the simplistic signal-cue dichotomy of ethology fails to suffice.

The informational regime that “leaks” from our behaviors is anti-inductive—any newly discovered pattern of *behavior → information* becomes “priced in” to individual actions. We'll use your own example: a classroom of young children beginning around ages nine and ten. Inevitably, there is a first crush: one student develops a fascination with another, and begins watching—reading—its object incessantly. (*Fascinare*, for “witchcraft”; the subject is hypnotized, captivated attentionally—in this case by a sudden surge in the salience of an information source.) At first, the object of fascination may not know she is being watched, or if she does realize, she may not know the “meaning” of the watching—what its fact entails for her—and eventually, when this too is learned, she is *still* unable to write for her fascinated subject until (1) she knows what she herself desires out of him, if only at an unconscious level, and (2) has some sense of what will “steer” him towards this desired outcomes. Over her life she will build this model of male desires, of gestures and actions, signs and signals she can pass along which will either encourage or discourage them—*in order to* escape harassment or clinch a reciprocated crush or preserve a friendship—as she goes from a crude model of general human behavior (perhaps premised primarily on her own psychology) to something more specific and fine-grained, and ability to size up “types of guys” and then interrogate their preferences, desires, properties at an individual level. What she learns from her marks are patterns that are meaningful for the affordances or obstacles they provide in getting what she wants.

As soon as other students learn that fascination—as publicly observable through gaze, or general attention—is a sign of a crush, they will begin to watch others for it, and to hide or flaunt the signs of fascination in public according to whether they wish to conceal or make a point of their interest. They may put on a show of the signs even when they do not feel interest, to accomplish some other strategic end. And individuals will still need to stare *somewhere*, and will sometimes need to stare *at each other*, and (in the case of a crush) will want to watch one another purely to gather information, *separate from the communication of romantic interest*. Ethologists would call this staring a cue—the subject pragmatically wishes to accomplish X (gather information), and in achieving X lets off information Y (gaze as sign of interest)—but this simple classification hides how much leeway of maneuvering is left up to the subject, the way the shape and style of his endeavouring for X changes as he

learns the kinds of information (Y1, Y2…) typically conveyed by the different modes of endeavouring. In other words, externally observable choices (actions, results of actions) develop reputations. *If this, then that*.

IV.

> *Moreover, fashionability, or cool, or style, acts as a passphrase, a shift key, a phase shift, a valuable proxy for speaker identity which then allows the speaker to communicate complexly, reflexively, with reference to self and modified by self. It is reliable because of the intense difficulty of faking fashion, which requires so much insider knowledge that any successful impostor is arguably no longer a fake. There's a reason it's tough to get into Berghain. Consider, by way of another example, the way true upperclass belonging, or highbrow aesthetic taste, is so impossible to convincingly fake for those outside the caste. So much of fashion is unquantified, subconscious, and ambiently soft that we are not even explicitly aware of the class and taste signals we send, or what alternative signals we might send to signal an alternative identity, and yet we send them constantly, through word choice, attitude, posture, interest, palette, reference, familiarity.*

Why, in ethology, is the dominant frame "signaling," with its implication of out-of-the-wayness—and why do ethological signals tend to be either costly or indexical—when in human social life, reading and writing—that is, the interpretation and creation of information—form an inseparable part of the fabric of action, and "mere talk" can effectively certify? The answer, I think, lies in the flow of information among human-level intelligences.

Much like the height of a tiger's scratch marks, on a tree trunk, is indexical to the size of the tiger, the choice of expression in anti-inductive games is an index of the upper limit of an individual's cultural exposure and intelligence; moreover, if expression is performed at the right point in the anti-inductive curve of adoption, the reading subject will not perceive it as conventional, and this lack of conventionality makes it more effective and reliable.

A game is anti-inductive if, once a salient piece of information becomes mutually known (among players in a round, or the larger playing population) it is priced in. That is, no individual player can gain an advantage from leveraging the information. The economy, and the fashion landscape, are two examples of anti-inductive games.

In this economy—that is, in the world of anti-inductive games—it is the rolling cycle of fashion (discovery, adoption, abandonment) which determines the value of "mere talk," determines whether a writing is read as one wants it to be read. And it is information—where, in the slow processing of pricing in novelty—which allows some to write in a successful manner.

In anti-inductive games, it is not just that information leaks from our behavior, but that information about which moves are winning leaks—in other words, it is not just that we understood X as a sign of Y—X trait as implying, metonymically, a private quality Y—but that we also have a sense, from watching the resultant payoffs, of whether displaying Y is a winning or losing move, and by extension, whether X is a winning or losing metonym. (Bikhchandani, Hirschleifer, & Welch 1998: "The propensity to imitate is presumably an evolutionary adaptation that has promoted survival over thousands of generations by allowing individuals to take advantage of the hard-won information of others.") By making a winning move in public, we cannot help but share our winning moves with others, which allows their adoption, which leads to "solution fads." We are read not just for the purpose of acting (adversarially, against us, or cooperatively, with us) but in order to glean information about winning moves—we watch one anothers' strategies and their payoffs in order to determine whether we ought to follow in footsteps, or choose a different path.

Just as inflation—the devaluation of a unit of currency—results from an increase in money supply, in anti-inductive games, many strategic moves are devalued as they popularize. As a given solution—a given tactic of writing—spreads throughout a population of players, and readers are increasingly deceived, readers in turn realize they are being deceived, and update adversarially "against" the solution, developing means of interrogating or discounting the solution. A plane flying over enemy territory might use the metonym of "the shape of tanks from above" as a cue to the presence of enemy tanks proper. So the enemy sets up rubber or wooden tanks which are indistinguishable from the air, causing false beliefs to form. As soon as this practice becomes widespread, air surveillance will develop techniques for distinguishing dummies from the real deal. Such tactics do not work over the long-term of evolution, because as soon as they are found out, they are countered. But they make up the vast majority of moves in human games, *because* they are short-term. We ride the informational adoption curve until it loses enough value to merit hopping off.

03 Letter 3

BY CRISPY CHICKEN Created 15 Aug '21

In your first two letters, you setup the foundations for Generalized Reading and Writing, and I would like to take this opportunity to show how very complex ecosystems of reading and writing might develop—one egging on the other.

What is a signal?

When I look at society from an appropriately ridiculous level of abstraction and defamiliarization I think to myself: "What's all this?"

All the levels of intertwined groupings, rituals, roles, moves, and narratives are constantly screaming out to everybody, saying "I'm here!", but it's incredibly difficult to actually figure out how the ecosystem functions, not least because it relies on lots of private rituals that only have to make sense to those involved, e.g., how individuals in a couple mutually manage each other's emotions.

The second question that pops up if I look a minute longer is: "How did we get here?"

The answer to that question is necessarily long and tedious. Given the right representation of the answer, it will be filled with interesting patterns and dynamics, but the reality is that society is complicated and specific and it evolved from something complicated and specific so the transition must be a function of many inputs. There is no neat regular structure to it because even if the underlying forces acting on humanity were uniform and static (they're not), there are still so many little things that were specific from the very beginning—for instance, the ecology of earth before Agriculture—whose complexity must be factored in.

I think it's more fun (at least at this stage in our ability to describe social dynamics) to think about how new social strategies get made in the general parameters of the societies we already have an intuitive understanding of.

A very common "primitive" in the way we tend to describe social dynamics is that of the "signal". *Signaling Theory* has become popular in economics, e.g. Robin Hanson and Kevin Simler's *Elephant in the Brain* suggests that much of all human intention in a given action is essentially signaling. As Robin Hanson says in an interview, "scratching your butt" is one of the few things you can do that is *not* signaling, emphasizing the pervasiveness of the frame.

While I am in agreement with Hanson that signaling essentially saturates our action space, I feel like his framing ends-up betraying how insufficient it is. During the interview, when Hanson refers to "scratching your butt" the audience laughs and Cowen, the interviewer, avoids repeating the string verbatim. This is because the idea of scratching one's butt makes people uncomfortable, at least in something public and published.

Indeed, the first thing I think of when Hanson says that "scratching your butt" isn't signaling is: Why don't I constantly see people scratching their butt? Or at least, why don't I *ever* see people scratching their butts in public? Butts itch sometimes, so this fulfills a real need.

Okay, okay before all the normies burst through the door and tell me "scratching your butt is embarrassing": yes, this is obviously the answer! But that's exactly why Hanson's answer is so ridiculous. He chooses this example precisely because he knows people don't do it a lot. He does not choose this example because it's unimpressive, but because it *signals social malfunction*.

Is this signal intentional? The answer to that certainly isn't a binary "yes" or "no", but what does a good answer even look like? "Somewhat", for instance, would be a very unsatisfying answer.

Let us look at the case of poker.

Vague Poker

Poker is a game with many variations in which people bet increasing amounts on their hands which they are dealt randomly. We will deal with it from a 10,000ft aerial view.

In poker, everyone can only see their own hand, but they can see what other people choose to bet under different circumstances, as well as shared cards that are revealed as the rounds proceed. Interestingly, the bets of different players are formally connected. Over

various rounds where information is revealed the players go around and *match* or *raise* the bet. If a player is not willing to bet at least as much as the previous bet they must quit (or "fold"). However, even if a player believes that they are actually unlikely to win, they can bet an amount that suggests they are likely to win, causing others to fold. Further, the amount of money one can bet is not uniform across players and changes across a session, allowing for complex leverage dynamics.

I will call this Vague Poker, so it is not confused with any specific version of that beautiful game.

Level 0

Imagine we have no way of reading each other's intentions or thoughts, no Theory of Mind, no reading of "tells" that reveal how good your hand is, no eye-tracking. If you were to play poker under these circumstances then you must look at your own cards (your "hand"), look at the shared cards, and make a guess as to (a) how much your hand will improve with the yet-to-be-revealed shared cards and (b) whether that will beat the best shared hand of the _N_ other players that are still in the game. Since you have no idea if they will fold or not, their actions are irrelevant.

This is level-0 reading, as you are reading the objective phenomena that are relevant, but you are not reading anything (like a person's facial expressions), that could "write back".

With this reading in hand our poker player, henceforth PP, can simply calculate their expected gains/losses from folding, matching, or raising the bet and act accordingly. A simple matter, though it is unclear how difficult it is to actually find this probability for a sufficiently precise and accurate model of the game.

Level 1

While a naïve player with a computer might simply invent an algorithm to decide whether their hand is likely to beat the _N_ other hands that it is competing against and then be satisfied once they have a proof that they are playing "optimally" for some silly notion of optimal that can't take advantage of the tiny subset of strategies it must actually contend with, a more pragmatic player might start to keep a record of (a) their wins and (b) what likelihood they predicted for winning under a given round. With this information PP can tell whether their model is well-calibrated.

When PP realizes the model is not well-calibrated, they might start snooping around for other kinds of information to condition on: after all, if the model is optimal, maybe they're just playing the wrong game?

But, of course, poker *does* provide an additional source of information: the bets other players place and their choice to fold. Armed with this information, PP might decide to start considering

(i) The likelihood of beating less and less hands as they continue to fold.

(ii) The likely composition of hands, given the choice to fold and the shared cards available.

(iii) The likelihood of another player having a powerful hand, given PP's own hand, the shared cards, and the way other players raise their bets.

...and many other possible signals.

Note that this reading is a "reaction" to the previous writing not having the desired effect—the miscalibration between PP's model of the game and the outcomes PP managed to achieve. Such effects are common: when reading or writing fails, a new source of information in the environment must be harnessed or new choices must be made to poke the environment into yielding such information.

The choice of what information to condition on and how to represent game states is left up to PP, and there is no one right answer, though there are representations that make it impossible to choose a winning strategy in certain cases. Certainly there are ways of representing the information that subsume all other possibilities, e.g., considering everyone's moves, the shared cards, and one's own hand a unique state is the "full representation", but the question then becomes "How do we compare all these unique states to each other in order to leverage our knowledge form previous games?" Often simpler, more reductive models are better at doing this, for instance just keeping track of which recognized Poker hands PP is one-card away from.

With PP's fancy new conditional model of poker outcomes, they will surely win some more rounds—and thus develop a taste for more winning. Since PP's conditional models yield probabilities, they will quickly notice something strange: when *other* players reveal their

hands they often have hands that are much too poor for the bets they raised—at least according to PP's model. What on earth could be incentivizing other players to do this?

04 Letter 4

BY CRISPY CHICKEN Created 15 Aug '21

Level 2

After entertaining a number of hypotheses our dear PP will realize the simple truth: other players are manipulating his naïve model of probabilities by leaking information that they knew PP would interpret "incorrectly", causing PP to make predictions about an opponents hands that are inaccurate. By doing this, opponents can push PP out of the game, causing PP to fold, without what PP views as the "requisite" cards that should make that possible

This is generally called "bluffing".

Bluffing is a funny thing—because most of the time in poker (when it's played seriously) no one makes any sort of claim about what their hand is. Rather the "bluff" is considered to be a bluff against the naïve model in which people bet if they really think their hand is better than everyone else's. Yet, poker is a game where one is precisely supposed to use such techniques to get an edge, making the negative/unfair connotation of bluff feel quite absurd.

The thing we are always fighting, however, is our lack of complete control over our actions. Humans are social and highly communicative animals, and we are always telling each other what we think, consciously and unconsciously, both with and without words. After all, why do people smile at text messages, alone in their rooms? There are many answers to this question, and a complete answer is far from being nailed-down by hard science, but let us simply accept that we find it difficult to stop ourselves from communicating our thoughts.

Sure enough, we can find plenty of ways in which people "communicate" whether they are bluffing or not through their habits, demeanor, and words—altogether these are known as

"tells", which [Mike](https://www.imdb.com/title/tt0093223/) informs us is because we are "telling" everybody else what we're thinking.

After a few frustrating losses PP will notice that people who "shouldn't" be winning are, and after a few dozen more losses PP might start to notice that these people tend to act slightly different when they win "fairly" vs. when they win by "bluffing". PP has discovered tells and will naturally add the observation or non-observation of tells to their predictive model of predicted gains and losses.

Armed with this new model, PP will make a bunch of money, much to the chagrin of other players...

Level 3

...who will quickly look inward and realize they are giving away information. Aha! We have come to the most basic pattern in Generalized Reading: Reaction Repurposing.

This can be described very simply as:

> If a behavior an agent engages in elicits a reaction from their environment (including other agents), they will naturally come to use that behavior to the manipulate the environment to their benefit.

Here are five examples:

1. Children learning to keep track of their smiles, sighs, etc. in public so as not to be misinterpreted (or correctly but undesirably interpreted) as directing their attention at someone around them.
2. People learning the topics that put someone in a good mood and indulging in them talking before asking for something or saying something negative.
3. Learning to dress in a way that gets you compliments from the people that you want compliments from.
4. Stores having a greeter who stands at the door and says "hello" and "have a nice day" to discourage theft using purely social mechanisms.
5. Companies creating new versions of products at quicker and quicker cycles because certain groups use having the latest gear as a status signal.

there are millions more. I mean millions—and easily billions when you look at non-human behavior, e.g., fruits becoming more and more delicious over their evolutionary history because it helps them spread their seeds.

So, when poor PP starts very obviously watching for the tells of his opponents, it's obvious what they must do: trick PP by exhibiting the tell (or not) in a situation that is at odds with the information PP expects the tell to correlate with. Many humans have a tell for when they are bluffing, because this kind of "lying" is extra cognitive load, and humans exhibit lots of interesting behaviors (e.g. sticking out the tip of our tongue) when engaging in high cognitive load tasks.

When our opponent, henceforth O, decides to fake a tell PP will lose a significant sum of money—but will start updating the model of tell-hand correlations, making the tell will become less and less effective. This is anti-inductivity in a nutshell: moves that leak information about underlying strategy can usually be used to make that strategy less effective.

Level 4

But let's think on this: O might never have even been aware of their tells if PP hadn't used them against O. So what if PP didn't "leak" this information?

If we assume that PP eventually learns to be stealthy and isn't just making it obvious what tells are being read, then O can still eventually find out by simply looking in the mirror and correlating events with outcomes. But PP can do something dastardly back: simply bet against the information presented, decorrelating wins and losses with the tells themselves.

This is tricky—because ultimately PP still wants to win money, so they can't just always do this without being at disadvantage. Instead, PP can do it when the stakes are low, and only slowly leak information out when the stakes are high, or perhaps when PP's opponents are inebriated and less likely to notice.

This, of course, still leaks information. As O realizes PP is only using tells when the stakes are high they will keep track of the leaked information in their own model and eventually learn the tells.

Level 5

In poker, people usually put a certain amount of money down. For most games of poker that I've been privy to this is "enough to be interesting, but not enough to ruin a friendship". Depending on how much money players put in, they can "leverage" this sum to make moves that other players will be unlikely to retaliate against because of how little money they have left to bet. Various social contexts deal with this differently, by forcing players to quit when they lose all their money, allowing them to put more money down, etc. and these choices—as well as the financial situations of the players—determine how this leverage must be interpreted.

As PP gets better, they will start looking for more information and eventually find that being low on leverage forces people to fold when they might be called out on their bluff. The simplest method to take advantage of this is to bluff harder when others will be unwilling to call such a bluff, though there are more complex ways of using this information both in the game and through social presence.

O will not leave this unnoticed and will attempt to price how over-confident PP is at any given moment, though there is always a danger of information asymmetry: What if PP knows that another player just lost a lot of money to their mother's medical bills and is much more "price sensitive"?

Level 6

Of course, any game of great enough stakes cannot be confined to a single room. In their unstoppable quest for power PP might engage in social engineering, for instance by encouraging O to buy an expensive car when chatting in non-game settings, then using the pressure of debt against O during a game of Vague Poker.

This, however, might be recognized by O, and so PP could go even further: they could hire someone else to rob O's house, stealing their most expensive possessions which PP is aware of after years of friendship with O. While this might make O suspicious of all his friends, it would be difficult to narrow it down to just PP, unless only PP knows about a certain possession or he is aware of PP's gambling addiction.

Readers interested in taking this to its logical end are encouraged to watch *House of Games*.

05 Letter 5

BY CRISPY CHICKEN Created 15 Aug '21

Humans Don't Stack Things That High

If it feels like most poker players would not go as far as PP does, it's because that's true.

Allow me introduce another little conceptual tool, *Sardine's Law*:

> *Humans don't stack things that high.*

This is shorthand for the fact that humans tend not to build models with many layers, where each layer requires the lower layer to be quite accurate. Instead, people tend to build very *wide* models, like spiderwebs, where information about the environment can be gleaned through the vibration of a number of different threads and cross-referenced. The world is ever changing, so building-up incredibly complex and brittle formal models tends to not be of much use.

Relatedly, level 6 is clearly in "Movie Territory", i.e. I doubt most poker players would be willing to screw over their friends *that* badly just to win a few rounds. Despite that fact, I think strategies such as leveraging shared history are used all the time—in both microgames between thoroughly entwined individuals and macro-scale games of global power struggles. They are simply used in ways that we don't feel are quite as bad, "Honey, I know we just renovated the bathroom, but I think that's exactly *why* we should wait a while before buying a car. Are you even sure you really want it?" Even this example is far more overt than most actual strategic interaction in daily life.

The less overt a strategy, the more context needs to be built-up to describe it—and the harder it becomes for moralists to moralize who is really getting what out of it. We live in a complex ecosystem of give and take we are only partially aware of. Thus, it is easier to describe in a game of Vague Poker.

What the above explanation brushes over is that

(a) humans have a funny way of keeping models of the world in their head we don't entirely understand yet, and won't fully understand by the end of the century either.

(b) humans can go through the above steps much more quickly than the information channels I've described suggest, largely because of *Theory of Mind*—our remarkable ability to imagine what other people are thinking given our knowledge of them and the situation they're in.

These are really where all the interesting bits come in: How do humans factorize their models? How do they leverage internal simulations? How similar do they assume other people's internal simulations are, or do they simply assume they must perform above certain levels to have functioned in society so far and in a neat little error term?

However, the above narrativization is meant to give a sketch of the stages we might at least consider in our internal simulations of a game, while showing how we *inevitably* leak information that gets read by our environment, which then chooses to write us different kinds of messages.

06 Letter 6

BY SUSPENDED REASON Created 07 Jan '22

Crispy,

I want to dig in to the first half of your letter, and your feeling that something about ethology's signaling theory, when applied to human affairs, becomes confused. Recursive phenomena (e.g. Sicilian reasoning, after *Princess Bride*) seem to play a non-trivial role in strategy games, but I'm not sure that stacking layers beyond standard theory of mind stuff

("I expect you to expect me"; "I expect you to interpret") is fundamental for formulating a theory of generalized reading and writing. I think you're also skeptical of this, given your letter's conclusion; I also know you've been considering writing a rebuttal or refactoring of *Elephant in the Brain*, so let's stay with Hanson for now.

The first error we make, in talking about signaling, is using language like "X is a signal of Y" to describe any situation where you can attribute plausible positive connotations to X. This description collapses a buncha different dynamics: whether X has Y effect on its audience; whether X is intended by the signaler to have Y effect; whether the signaling (i.e. communicative) intent is the primary purpose of X or merely a happy byproduct; etc. Rather than reify an effect (desired or actual) as the "is" essence, in a to-be form, of a given action, let us treat it only in verb form ("X signals Y"), and further, always ground it in either intentionalist, prospective terms ("X is performed with the aim of signaling Y") or reader-response terms ("X is taken to signal Y"). There are additional (Gricean) combinatorial layers: "X is taken to be intended to signal Y," "X is intended to be taken as intended to signal Y," etc.

And the same arguments, and the same linguistic refactorings, hold for a word like "meaning." There is no third property of language that is "meaning" outside of or in addition to intent and interpretation; there is only meaning as a process of intending and interpreting.

This taxonomy isn't perfect—for instance, intent is arguably a muddled concept in its own right, collapsing the conscious-unconscious binary. The presence and absence, let alone the exact parameters, of intentionality is notoriously difficult to assess, both for outsiders (opacity between minds) and insiders (opacity between conscious and unconscious), forcing us to rely on historical patterns as a way of projecting desired effect. (Even as I still believe that something like "desired effect" is real; perhaps you disagree.) But, despite all these problems, I think verb-ifying "signal" and "meaning" is an improvement on the status quo, at least insofar as it helps us begin to break down the butt scratching example.

Next, there is the problem that, typically, it's assumed we only signal things which improve our sexual or evolutionary fitness, and thus, in a crude way, we can explain much or most culture in terms of "maximizing fitness" (perhaps with one or two causal syllogisms to get there, e.g. "the signal helps form an alliance; alliances improve survival chances"). This is not the view of all evolutionary psychologists, but seems true of many of the field's less-sophisticated advocates.

I think we can clear up both this naive evolutionary view, and the “signal” reification (nounification? essentialization?) issue, by anchoring ourselves in the metaphor of strategy games. Strategy games are an established framework for human behavior dating at least to the 1940s (see e.g. Jessie Bernard 1952, “A Theory of Strategy Games”) but it was not until Schelling that the closed-world mathematical view of strategy games was modified to better describe human interaction. These modifications included: an emphasis on mixed games of coordination and conflict; a re-emphasis on supposedly “arbitrary” features of the map or territory as “focal points”; the merging of economic “signaling games” into strategy games, etc.

Very simply, we will say a game exists, at the very least, between an agent, a goal, and an environment. A goal emerges any time that the agent has a preferred state of being—what we might call “desires” in more mystical, humanist terms—which he may influence through his actions. The game’s goals transform the environment into a set of affordances and obstacles. Since agents cannot exist outside environments, and since agents are defined by the possession of preferences (they must, at the very least, maintain homeostasis, reproduce, and must maintain boundaries and take in energy, to do so etc), a game is, very simply, the default and inescapable state of agents.

In single-agent or “one-player” games, the game is one of engineering; physics alone is theoretically sufficient to design a solution. Once a second agent is introduced, the situation becomes a game of strategy; theory of mind, generalized reading, and generalized writing suddenly come into play. Since humans spend most of their lives around other human beings, pursuing goals whose attainment is altered by those other agents’ actions, games of strategy are endemic to human life.

Within strategy games, we must distinguish between the intrinsic and extrinsic effects that result from our actions. That which is intrinsic exists separate from outside observation, for instance, the desire to relieve an itch. The physical sensation of relief does not (provided one can reach the itch) require other agents. The intrinsic game is won by “work” in both the thermodynamic and colloquial sense: it is a game of physics first and foremost. Such games are rarer in modern life, where manual manipulation and ability has lost ground to inter-subjective manipulation and ability, but solo tasks such as making a fire, fashioning a stone knife, or collecting fallen wood are all intrinsic.

With extrinsic effects, the situation is entirely different. The effect is never physical, the result of energy, inertia, and heat—rather, extrinsic effects are always upon and through other agents, who observe the actor and alter their own behaviors around what they observe. Recall e.g. Bateson's "Form, Substance, and Difference" on the difference between effects born of information, and those born of literal force. We use "free energy" to describe both informational and thermodynamic transformation potential—to Friston, surprisal, that is, deviation from expectation, *is* information, which causes us to update our behavior and model (in other words, transforms us). But the function and structure of these energies seems meaningfully different.

(The differences and similarity between such effects, as types of "manipulation," are important to sus out, and I welcome that direction in future letters if you wish to take it.)

In one-player games, only intrinsic effects exist. In multi-player games—games of strategy—extrinsic effects become possible and even begin to dominate the landscape of moves.

All actions, even those whose desired effect is purely intrinsic, produce information—for instance, the footprints one leaves walking in snow, mud, or sand. In other words, every action "signals" something insofar as every difference (of decision, of movement) makes a difference (in environment). Some of these differences persevere longer, or become more permanent in the environment; or are more immediately temporally and attentionally salient; or are more pragmatically relevant to observers; or have more or less possible origin stories, making them more or less ambiguous as testimonies to their cause or source. But all actions "add information" and disambiguate the situation; that is, all actions testify or signal something about both the abilities and choices, and by extension the beliefs and preferences, of the acting organism.

Moving on to the next item. There is a somewhat mistaken belief in many otherwise savvy groups of thinkers that most signals in the social, like the natural, world, are necessarily costly or hard-to-fake (and thereby honest). It is true that, in reading one another, we are aware enough of the potential for misrepresentation that we carefully monitor and verify representations, and that further, in writing to one another, we are aware that we (our audiences) are liable to monitor and verify, and through this, are kept tethered near the truth, at least most of the time. But insofar as we consider something an expressive action, an action which adds information and thus alters an observer's interpretation, it is far more common in human social life that a signal is easy to fake, than that it is hard. (And also,

many of these easy-to-fake signals are honest, enough that they can be relied upon.) A display of sophisticated aesthetic or culinary taste, a printed diploma mounted on the wall, the pronunciation of a word, the correct garment available at a common clothing store—none of these are especially difficult to fake. Many of them can be picked up and imitated from an n=1.

[...]

07 Letter 7

BY SUSPENDED REASON Created 07 Jan '22

[...]

Rather, it is the total assemblage of signals, all of which add up to create an impression, that is difficult to forget, especially when audiences will disqualify (or at least begin to suspect our efforts) upon a single contradictory signal, a leak of information we did not realize would undermine the impression we are attempting to create. It is rarely a sign and more often the assemblage, with all its possible contingencies and reactive decision trees, in the face of normal interaction—let alone concerted, skeptical interrogation—which is so difficult.

Goffman believes that this—the creation of an impression—is the core purpose of our actions in public. But it is not solely an impression of oneself one seeks to create, not always a self-expressive performance one puts on for others. As performers, he writes, we are always attempting to establish and maintain a "definition of a situation," for instance, that a romantic date is casual and low-stakes, that the courting person is a desirable prospect and worthy catch, that what they are engaging in is, in fact, dating. The plumber performs not just that he is competent but that he is beginning and finishing his work, that he will be entering the bathroom now as part of his professional inspection (and not for other untoward) reasons, etc. Goffman settles on the metaphor of the stage, with his dramaturgical theory; I believe he is right to call this dramaturgical, but that we can clarify his argument by zooming out and contextualizing it within the game playing metaphor.

That is, I believe that we do not perform as an end in itself, and that we are not, at our core, performers. When we regularly devote time to intrinsic activity and rewards, such as fixing a sink, we may carry over the habits of comportment that we have established within a life in public, but we are not performing. There is a meaningful, and phenomenologically salient, difference between how it feels to be within view (especially within view of those whose judgments matter dearly to the observed actor) and how it feels to be alone. But despite the burden of performance, the burden even of performing one's "true self" (that is, in a way which minimizes conscious dissimulation), we slowly build up our conditioning and tolerance for performing (such that, when we have spent long periods in isolation, we become "out of shape," are more quickly and easily exhausted by social activities then we were before entering isolation).

It is specifically within extrinsic games, and in service of extrinsic effects, that we take up performing *as an instrument* to game-winning. And though we often idealize ourselves in our performances, this too is primarily instrumental and on the margin, reified autotelically; there may well be a "fetishistic," i.e. for its own sake, and separate from its pragmatic contribution to our goals, desire to be liked and looked well upon; but it is clearly the case that we can and do take on definitions of a situation which *fail* to flatter us, if they accomplish our goals. For instance, in therapeutic settings, patients may begin to confide information that is embarrassing or taboo, out of a desire to make psychological progress by cooperating with the therapist via disclosure. (Goffman, e.g. focuses more on the idealized nature of performance, and not on its pragmatics—my gut holds that this is a mistake; perhaps you differ in opinion.)

Even actions which are meant primarily or solely to convey information, that is, to manipulate an agent, which is to say, to create extrinsic effects, typically accomplish some secondary, intrinsic effect which is a plausible desired effect of the actor. The ambiguity creates plausible deniability, and when we are believed to be performing without constraint by intrinsic goal, others rationally distrust us. There are of course exceptions, in which it is understood that one is being asked to, essentially brag, or to verify some information in a hard-to-fake way. (Selection games such as job interviews and secret society membership tests spring to mind.) But when it comes to easy-to-fake signals, we must be believed to release them as "cues"—that is, as unintended or secondary effects attached to a primary, first-order, intrinsic effect we undertook the action to accomplish.

It is certainly true that “drama” springs from the pursuit of goals, and the conflicts which arise when the drama’s characters have goals which are at least partially mutually exclusive. (Or which make other goals more difficult to achieve.) But “drama” in the sense of theater is only a third-person simulation of these goal-driven conflicts. The actual actors view their work as predominantly cooperative (the performance of a script to entertain and move an audience) and only adversarial on the margin, and where there is inevitably conflict (for instance, one actor may support a more “reverent” reading of the script, and another actor a more “avant-garde” interpretation) the conflicts are not those presented to the audience as driving the overtly visible drama. Occasionally, the desires and conflicts and goals of the actors do come to partially resemble the diegetic plot; indeed, this happens more often by chance, likely because (1) the roles we perform are the result of selection and self-selection games whose outcomes are determined by our real character and capacities, and because (2) in playing these roles we are put into plot-approximating relationships saturated my simulated emotional displays that cannot help but have real effects (see e.g. the many well-known cases of the lead actor and actress, who are romantically involved on-screen, becoming similarly involved off it).

[Indeed,“shoots” (“real” drama which bleeds through the kayfabe, in wrestling speak) are immensely popular inside and outside, giving rise to the concept of a “worked” (i.e. performed) shoot—a performance which pretends to be a back stage reality leaked onto the front-stage. Sports fans enjoy when athletes who are rivalrous on-court extend that rivalry off-court into mutual antagonism (drama) because it adds a layer of (what appears to be) human reality to the more contrived sports game.]

But by and large, in literal theater performances, actors embody the staged conflicts only from the third-person; internally, their state of mind differs; moreover, their state of mind is a blackbox, such that the dramaturgical metaphor fails to explain the motivation or psychology of its performances. As a metaphor, it gets us the instrument or tactic—both theater actors and everyday performing game-players are attempting to “write” effects to reader observers, and generally to persuade through naturalism the veracity of their performance. (Creation not just of effects, but of their veracity; writing which is believed.) But it does not explain the urges which give rise to drama. In this sense, games, for instance sports and poker, serves as a better analogy. It is clear why a basketball player fakes his shot, or why a poker player bluffs; they are attempting to secure desired outcomes. The stage actor of course is also attempting to secure *some* desired outcome, such as

entertaining the audience, impressing his peers and director and any watching theater critics, feeling to himself that he has done a good job, etc. But that is only insofar as he is playing a specific kind of game, “theater acting,” which involves performance. And this game’s specific goals are narrow, irrelevant to most human activities, and detached from the fault lines of drama the show claims to present.

08 Letter 8

BY SUSPENDED REASON Created 07 Jan '22

I’ll return, before this letter finishes, to think through the relationship between intent and effect. To scratch one’s butt, or genital region etc, may be done to satisfy intrinsic goals (that is, out of callousness to the observer, or even out of the belief that one is not presently observed) but its effect on the observer may be pronounced, appearing as a show of indifference or dominance. As for the relationship between intent and effect: these effects on the observer will differ depending on whether the observer believes the scratcher believes himself observed, i.e. the appearance of intentionality strongly shapes effect; further, the appearance of intentionality correlates statistically with (but has no necessary, inherent relationship to) the reality of intention.

At the same time, intended meaning can be dramatically altered when a reader speaks a different “language” (bears a different interpretive schema) then that which the writer envisions in writing his text. Our efforts to "mean" are always bottlenecked by our reader's capacity to mean, his grasp of the possible entailments and causal origins of the symbolic displays he encounters. As Ben Hoffman recently wrote me in a letter about torque dynamics:

> *a Mohist (Chinese utilitarian) is invited to help defend a city, but gradually discovers the belligerents on both sides are not actually acting on self-interest or trying to win the conflict, but are instead committed to playing out their roles, even when this kills them.*

> *They interpret his constructive attempts to save lives as power grabs, and the regime he's trying to help repeatedly acts to thwart him. His attempts to save the lives of the enemy soldiers and leaders are also thwarted, partly by their own actions. By the end of the movie the city has been burnt to the ground by the armies supposedly fighting over it, and the Mohist hero is leading away the local children, who aren't old enough to have been initiated into a Hegelian death cult.*

The tragedy, of course, is that the language of power, native to those playing the power game, cannot help but interpret all actions as moves in this game. This is how the Mohist's readers won their power games to begin with.

Yrs,

SR