All Mixed Up: A Solution to the Evolutionary Mystery of Humor

David Pinsof

Department of Psychology, University of California Los Angeles

Author Note

The author declares no competing interests. All correspondence should be sent to David Pinsof, UCLA Department of Psychology, Franz Hall, 502 Portola Plaza, Los Angeles, CA 90025. Email: david.pinsof@gmail.com.

Abstract

Human social life is filled with coordination problems: passing each other in a hallway, taking turns talking and listening, differentiating the meanings of "hook up with" and "meet up with," gathering at the same time and place, etc. But what happens when we suffer a mix-up—for instance, we get stuck dancing back and forth in the hallway, or I casually mention that I "hooked up" with your mother last night? Here, I argue that such mix-ups posed a significant adaptive problem for our ancestors, disrupting cooperation, damaging reputations, fomenting needless conflict, and destroying valuable relationships. Natural selection favored three solutions to this adaptive problem: 1) a sense of humor (i.e., the ability to detect, anticipate, and avoid mix-ups), 2) mutual laughter in response to humor (which creates common knowledge of the mix-up and defuses its costs), and 3) joking as a hard-to-fake signal of one's ability to detect and avoid mix-ups (and thus one's value as a coordination partner). I present a formal model of these phenomena suggesting that agents with a sense of humor, the ability to laugh at (or mutually signal) the humor they've sensed, and a preference for partners with a 'good sense of humor' would have been favored by natural selection under conditions that were plausibly characteristic of human evolution. I use the model to shed light on the nature of wordplay, comic timing, practical jokes, awkwardness, embarrassment, mockery, absurdity, mirth, seriousness, impersonations, slapstick, satire, creepy clowns, and cringe comedy.

Humor is a Darwinian mystery. We do not know what humor is, why we evolved an apparatus to sense it, why sensing it is joyous instead of confusing or relaxing, why a 'good sense' of it is important in a romantic partner, why we emit panting and hooting sounds in response to it, why these sounds are contagious, why we try to fake them and detect others' fakery, why they can sometimes be creepy or threatening, why it is aversive when they are directed 'at us' but not 'with us,' what explains their variation in volume or duration, why sharing them brings people together and makes others feel left out, why they convey the opposite of seriousness (what is seriousness?), why they arise in awkward or embarrassing situations (what are awkwardness and embarrassment?), why they depend on comic timing (what is comic timing?), and why they are said to be the 'best medicine.' The complexity and apparent disutility of humor poses a deep puzzle for evolutionary theory. While there have been many promising insights so far (Hurley et al., 2013; McGraw & Warren, 2010; Flamson & Barrett, 2008), no single framework has been able to tie them all together.

Here I attempt to provide—and mathematically formalize—such a framework, drawing on the game theory of coordination and common knowledge (Pinker, 2025). I propose that humorous situations are *costly mix-ups*, defined as agents trying and disastrously failing to align their behaviors and mental states in mutually beneficial ways (e.g., mixing up each other's words, roles, intentions, identities, expectations, or emotional states). We evolved a sense of humor to detect mix-ups, laughter to signal them, mirth (or amusement) to defuse their costs, and contagious laughter create common knowledge of the defusal. We are attracted to funny, charming, witty people—the ones we laugh and banter *with*—because they are better able to avoid mix-ups, which makes them more profitable coordination partners. We are repelled by fools, boors, oafs, and klutzes—the ones we laugh and cringe *at*—because they are prone to

costly mix-ups, which makes them ruinous coordination partners. When we smoothly coordinate with someone, we vibe. When we struggle to coordinate with someone, we feel awkward, which is aversive to us but often funny to others. Jokes are attempts to simulate, role-play, recollect, or point out costly mix-ups to show off our sense of humor, make rivals look foolish, and covertly differentiate profitable coordination partners from fools, dopes, and philistines (Flamson & Barret, 2008; Smaldino et al., 2018).

In the following section, I review the logic of coordination and common knowledge. I then formally model the selection pressures for sensing mix-ups, mutually signaling them, defusing their costs, and favoring coordination partners who are better able to sense and avoid them—i.e., partners with a good sense of humor.

Coordination games

Coordination is the water we swim in. Showing up at the right times and places, using tools according to their proper functions, pausing at the right moments in speech, saying things that are relevant to the topic at hand, interpreting others' utterances correctly, and finishing each other's… All of these activities can be thought of as coordination games where interactants mutually benefit from aligning their behaviors and mental states.

Tables 1, 2, 3, and 4 provide examples of coordination games. Table 1 contains the payoffs for two people trying to pass each other in a narrow hallway, with the possibility of continually bumping into each other. Table 2 depicts two players trying to meet up at the same time—and the costs of mixing up "am" with "pm." Table 3 represents the dynamics of a conversation—and the perils of talking over each other or sitting in awkward silence. Table 4 represents semantic coordination, with two possible interpretations of LOL: "lots of love," or "laugh out loud."

	Move to my left (your right)	**Move to my right (your left)**
Move to my left (your right)	$1,1$	$-1,-1$
Move to my right (your left)	$-1,-1$	$1,1$

Table 1. A game where two people try to pass each other in a narrow hallway

	Meet up at 9am	**Meet up at 9pm**
Meet up at 9am	$1,1$	$-1,-1$
Meet up at 9pm	$-1,-1$	$1,1$

Table 2. A game where two people try meet up at the same time

	Talk	**Listen**
Talk	$-1,-1$	$1,1$
Listen	$1,1$	$-1,-1$

Table 3. A game where two people try to have a conversation

	LOL = laugh out loud	**LOL = lots of love**
LOL = laugh out loud	$1,1$	$-1,-1$
LOL = lots of love	$-1,-1$	$1,1$

Table 4. A game where two people try to communicate with the acronym “LOL”

One important feature of coordination games is that they are win-win. You cannot ‘beat me’ at passing me in the hallway, and I cannot defeat my dentist by showing up to my

appointment at 9pm. If you and I struggle to coordinate, that not only hurts me; it hurts you too. We win or lose together.

Another notable feature of coordination games is that winning together often requires *common knowledge*, defined as "the recursive belief state in which A knows X, B knows X, A knows that B knows X, B knows that A knows X, ad infinitum" (Thomas et al., 2014, p. 2). Common knowledge matters for coordination because in order to align my behavior with yours, I need to know that *you know* what to do; otherwise I could not expect you to do it. Likewise, you need to know that *I know* what to do; otherwise, you could not expect *me* to do it. This implies, in turn, that I must know that you know *that I know* (that you know that I know…) what to do (Pinker, 2025).

Coordination can be distinguished from *cooperation*, or social dilemmas, in which there is an opportunity for either partner to cheat, freeride, betray trust, or fail to reciprocate. One of the major obstacles to the evolution of cooperation is that any agent can profit at other agents' expense by defecting, but when all agents defect, cooperation unravels (Apicella & Silk, 2019; Barclay, 2013). Regarding the evolution of coordination, the primary obstacle is not the incentive to defect, but rather the improbability of any set of agents converging on a single, mutually beneficial point in the vast space of behavioral possibilities. Cooperation and coordination are not mutually exclusive; indeed, cooperative interactions often contain coordination games embedded within them and vice versa (Skyrms, 2004; Pinker et al., 2025, chapter 3).

Adaptations for coordination

Throughout human evolutionary history, our ability to survive and reproduce strongly depended on our ability to coordinate with each other for mutual benefit. According to a growing body of research in evolutionary social science, coordination is the superpower of homo sapiens

—a skill we are more adept at than any other animal, the foundation of human culture, and the secret to our success (Chwe, 2013; Tomasello, 2009, 2022; Pinker, 1997, 2025; Henrich, 2015). We have evolved a cornucopia of adaptations to facilitate coordination, including:

1. **Conformity**. If coordination usually requires doing the same thing as others, then a good default strategy is to do the same thing as others—to, when in Rome, do as the Romans do (Grueneisen et al., 2015; Krupka et al., 2022).
2. **Theory of mind (or mind-reading)**. In order to know that you know that I know X, I need to know what it means to know something—to have a theory of mind, which includes concepts of knowledge, doubt, and belief, including thoughts about whether you doubt that I think you know X (Sperber & Wilson, 2002; Tomasello, 2022).
3. **White sclera**. The whites of our eyes are much whiter than they are in other primates, and it is easy to 'see' why. To create common knowledge, I need to know what you are looking at, because if you are looking at X, that implies you know X (Kobayashi & Kohshima, 1997; Tomasello, 2009).
4. **Preference for similarity in social partners**. We have an easier time coordinating with people who are similar to us—who speak the same language, are drawn to the same focal points of common salience (Schelling 1980, chapter 3), and share the same social conventions and expectations (Cole & Teboul, 2004).
5. **Language**. Language is a massive coordination game in which we are all constantly agreeing (and occasionally disagreeing) about which sounds or signals correspond to which informative intentions (Scott-Phillips, 2025). It is a game that you and I are playing right now, with each word I use—and each sentence I string together—constituting a unique opportunity for coordination or confusion.

6. **Norms.** Norms (including linguistic constructions) can be understood as equilibria in coordination games (Scott-Phillips, 2025; Binmore & Samuelson, 1994). Norms confer mutual benefits by reducing conflicts and collisions, aligning expectations, and facilitating role-based cooperation and division of labor (Binmore & Samuelson, 1994; Krupka et al., 2022).
7. **Leadership**. Leaders resolve conflicts, enforce norms, and create common knowledge of collective actions, so that everyone knows that everyone knows what to do and how to do it (Van Vugt et al., 2008; Pietraszewski, 2020).
8. **Ceremonies.** One of the best ways to create common knowledge of X is to get people to form a circle around X. That way, everyone can see that everyone can see X. Another way is for the person in the center of the circle to repeat X over and over again, or for everyone to chant X in unison, so that there are no doubts (or doubts about doubts) that everyone has heard (and knows) X (Chwe, 2013).

The rest of this paper will argue for the addition four more adaptations to this list: 9) the sense of humor, 10) contagious laughter, 11) mirth (or amusement), and 12) the meta-sense of humor (or the ability to sense others' senses of humor). In the next section, I will discuss each of these adaptations, outline their design features, and present a formal model of the selection pressures that gave rise to them.

A Theory of Humor, Laughter, and Mirth

A model of the sense of humor

In tables 1, 2, 3, and 4, there is the possibility of a coordination failure or *mix-up* between the interactants, in which both players incur a cost. The interactants might mistake "am" for "pm" or vice versa, continually talk over each other or sit in awkward silence, dance back forth

trying to pass each other in a hallway, or mistake “laugh out loud” for “lots of love” (and disastrously end their funeral texts with “LOL”).

Other examples of mix-ups include thinking that you’re waving at me when you’re waving at the person behind me, thinking that “hook up with” means “meet up with” (and saying you “hooked up” with my mother), mistaking an insult for a compliment (or vice versa), laughing at a grave admission (or remaining stone-faced after a joke), using “literally” to mean “figuratively” (and then being taken literally), mistaking “reply” with “reply all,” treating a superior as a subordinate (or vice versa), etc.

As we can see, there are as many ways to be mixed up as there are ways to coordinate. There is thus an immensely vast range of coordination failures we are vulnerable to as hypersocial animals—a vastness that is fully commensurate with the range of humorous situations we can entertain.

The confusability of a mix-up. Insofar as two things are analogous, perceptually similar, structurally isomorphic, or otherwise *easily confusable*, the two things more likely to be mixed up than other pairs of things. For example, homonyms (identical-sounding words with different meanings) are more likely to be mixed up than other pairs of words. Voices or faces that sound alike or look alike are more likely to be mixed up than other voices or faces. Substances are more likely to be confused when they look alike (e.g., ketchup and blood) than when they look different. Roles, standards, or situations are more likely to be mixed up when they license similar inferences or are triggered by similar stimuli. Alternative interpretations of an utterance are more likely to be mixed up when both interpretations are plausible given the context.

The costliness of a mix-up. We can also see that some mix-ups are more *costly* in terms of damaged reputations, damaged bodies (as in a hallway collision), or thwarted opportunities for

cooperation. Avoiding these costs would have been a major selection pressure for our ancestors. Mixing up words or intentions could render communication impossible. Mixing up the social conventions surrounding hygiene or pathogen avoidance could result in illness, poisoning, or ostracism. Confusing the proper uses of tools, the intended destinations for travel, or the goals of a collective action could waste precious time and energy. Mix-ups in the mating domain could be catastrophic for fitness—e.g., mistaking an insult for a come-on, collaboration for romantic entanglement, or a wardrobe change for seduction.

Mix-ups would be particularly costly in cases where there is a *large misalignment* in mental states between the two interactants—where they are on opposite ends of a continuum in their understanding of the situation. Smaller misalignments (e.g., mistaking a gentle pat on the back for a firm pat on the back) are less costly than larger misalignments (e.g., mistaking a gentle pat on the back for a physical assault). Insofar as the goal of communication is to get as close as possible to a shared interpretation of an utterance, wider deviations between intended and received meanings will be costlier than narrower deviations. Mix-ups can therefore have immediate, observable costs but also inferred costs related to the degree of misalignment between the interactants.

A simple coordination game. Consider the table below, modeled after tables 1-4. Instead of am vs. pm or talking vs. listening, I have added a generic "Option 1" and "Option 2." Think of these as any pair of behaviors, mental states, or communicative acts that would be mutually beneficial for two agents to align. Instead of (1) and (-1), I have replaced them with (*V*) and (-*C*), meaning that the benefits of successful coordination (*V*), and the costs of a mix-up (-*C*) are open-ended: they could be anywhere from minor to life-or-death.

	Option 1	**Option 2**
Option 1	V, V	$-C, -C$
Option 2	$-C, -C$	V, V

Table 5. A coordination game with open-ended parameters

Let *e* be the *confusability* of option 1 with option 2, or the degree to which mistaking 1 for 2 (or vice versa) is cognitively *tempting* to the interactants or particularly likely to occur. This parameter can range from high (e.g., "pianist" and "penis") to low (e.g., "pianist" and "saxophonist"). When *e* is high, mix-ups are common and tantalizing; when *e* is low, mix-ups are rare and easily avoidable. Since coordination games are not one-shot but repeated over long stretches of time, we will let *w* be the probability that the interaction will continue, with the costs of the mix-up continuing to be incurred, or the benefits of successful coordination continuing to be reaped ($0 < w < 1$). When there is a mix-up, which occurs with probability e, the partners pay the cost of being mixed up, C, yielding a payoff of $\frac{e(-C)}{1-w}$ over time. When there is *no* mix-up, which occurs with probability $1-e$, the partners gain V each round, yielding a payoff of $\frac{(1-e)V}{1-w}$ over time. Putting these outcomes together, we get the expected payoffs to coordination for the two partners:

$$\frac{(1-e)V - eC}{1-w} \quad (1.1)$$

Selection for a sense of humor. Let *p* be the probability that a mix-up, when it occurs, is detected by a focal agent. We assume that detecting a mix-up allows that agent to coordinate successfully on subsequent rounds by switching to the alternative option. When a mix-up occurs

and is detected, which occurs with probability e multiplied by p, the partners pay the cost of the initial mix-up, and then reap the benefits of successful coordination on all subsequent rounds yielding $ep\left(-C+\frac{V}{1-w^2}\right)$. Detecting mix-ups, with whatever probability, will therefore confer an advantage when:

$$\frac{(1-e)V}{1-w}+ep\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)C}{1-w}>\frac{(1-e)V-eC}{1-w} \quad (1.2)$$

This inequality can be simplified to the following (see Appendix A for the derivation):

$$e(V+Cw+Cw^2)>0 \quad (1.3)$$

Agents with an ability to sense mix-ups will therefore invade a population of coordinating agents. Moreover, the ability to sense mix-ups confers an advantage even when it is frequent in the population, suggesting continual selection on—and elaboration of—this ability (see Appendix B). As we can see in the above inequality, the selection pressure for mix-up detection (or a sense of humor) will be strongest insofar as: 1) there are significant benefits to aligning one's behavior or mental states with other agents (V is large), 2) mix-ups are particularly costly to coordinating agents (C is large), 3) mix-ups are easy to fall prey to or likely to occur (high e), and 4) interactions are ongoing and likely to continue into the future (high w). In addition, we expect selection to favor an ability to detect mix-ups insofar as this ability can be put to use in many different coordination games with many different partners.

Given the hypersociality of ancestral humans, their elaborate gestural and vocal communication, their reliance on pedagogy and socially transmitted information, their extensive cooperation and division of labor, their inclination toward pairbonding and intensive biparental investment, their long lifespans and enduring relationships, and the inherent confusingness of

reality (including the reality of others' mental states), the conditions favoring a sophisticated sense of humor would have been plausibly characteristic of human evolution (Apicella & Silk, 2019; Csibra & Gergely, 2011; Dunbar, 2009; Henrich, 2015; Kaplan et al., 2000; Pinker, 2010, 2025; Scott-Phillips, 2025; Tomasello, 2009, 2022).

How the sense of humor works. Since coordination depends on shared expectations between interactants, the sense of humor may employ predictive models of minds, actions, utterances, and social roles (Tamir & Thornton, 2018; Jara-Ettinger & Dunham, 2025). It may scan the social landscape for outsized prediction errors and, if any are found, search for plausible explanations of them. The search may involve simulating counterfactual interactions in which easily confusable behaviors or mental states are substituted for observed behaviors or mental states (Quillien & Lucas, 2024). If, in a simulated interaction, the outsized prediction error is reduced or eliminated, the system may terminate the search for explanations and switch to the counterfactual behavior or mental state. We will now turn to the question of how, when, and why laughter, joy, and amusement are added to these computations.

Modeling contagious laughter

If I think you "hooked up" with my mother or "laughed out loud" at the death of my beloved uncle, it scarcely matters if you sense your mistake and switch to the alternative meaning of "hook up with" or "LOL" in our next conversation. The damage has already been done. Thus, in addition to sensing mix-ups, natural selection would have plausibly favored the ability to *defuse the costs of mix-ups*—to undo whatever damage was done.

This is harder than it sounds. One cannot merely announce that a mix-up has taken place, because the announcement would represent another coordination game—one between the intended meaning of the announcement and its received meaning. For example, suppose you say:

“I didn’t mean that.” What did you mean by “I didn’t mean that?” Did you mean that you laughed to yourself and not out loud at the death of my beloved uncle?

Or suppose you are more specific. Suppose you say: “I meant that *I met up* with your mother last night.” There is still the possibility that I don’t believe you. Maybe I think you’re trying to backtrack from your admission—to cover up what you’ve done. If so, you can expect me to be hostile to you, which gives you a reason to avoid me—and, perhaps, my mother. To defuse the cost of the mix-up, I need to know there was a mix-up.

But interestingly, that is not enough. Suppose I’m certain you didn’t sleep with my mother, but *you’re* not certain if *I’m* certain. Maybe you still suspect that I am seething in quiet resentment. In that case, you might avoid me, which might make me actually resent you, which might confirm your suspicions, which might further drive us apart, gradually unraveling our relationship. The only way out of this mess is through *common knowledge* of the mix-up. You need to know that I know (that you know that I know) there was a mix-up.

More than that, however, we need common knowledge that the costs of the mix-up have been defused—that no grudge is being held, no resentment is being harbored, and no damage has been done. To take a simpler example, suppose I mispronounce your name. Even if we achieve common knowledge that it was a mispronunciation, I might still wonder whether you were offended by the mispronunciation, and you might still wonder whether *I think* you were offended by it. Thus, what is needed beyond common knowledge of a mix-up is common knowledge that its costs have been mutually defused, nullified, or removed from future interactions.

Co-opting the play signal. Many animals have *play signals* that they use to differentiate play interactions from non-play or ‘serious’ interactions. Cetaceans use an open-mouth display (Maglieri et al., 2024), kea parrots use a warble (Schwing et al., 2017), canids use a bow (Bekoff,

1995), and rats use 50-kHz ultrasonic vocalizations (Kisko et al., 2015). More relevant to our purposes, chimpanzees use a panting sound, uncannily reminiscent of human laughter, during bouts of tickling, chasing, or rough-and-tumble play (Matsusaka, 2004). If we could translate this panting sound into words, it might be something like: "I understand that this is play aggression and not real aggression. I am not mad at you or afraid of you."

The panting sound is thus a kind of mix-up signal—designed to avoid the confusion between play and non-play—a message one might benefit from sending in the event of a slip-up, misunderstanding, or faux pas. To return to the miscommunication about the "hookup" or the meaning of "LOL," the panting sound would send the ideal message: "I understand that this is play communication and not real communication. I'm not mad at you or afraid of you." It can therefore be slotted into the familiar payoff matrix for a coordination game:

	Play communication	**Real communication**
Play communication	V, V	$-C, -C$
Real communication	$-C, -C$	V, V

Table 6. A game where two players try to take each other's words seriously or unseriously

If you reciprocate my play signal, then you have not only acknowledged the receipt of the signal; you have confirmed that you—like me—are not taking the communication 'seriously' as real communication. Repeated exchanges would confirm the confirmations and create common knowledge of the communication's unseriousness. Since the play signal would have likely already existed in ancestral populations (Ross et al., 2009), the play signal would have been mutually recognizable to both parties and less vulnerable to the problem of meta-mix-ups (or mix-ups about the mix-up signal itself). Its emotional and cognitive underpinnings—in terms of

not taking play behavior 'seriously'—would also be ideally suited to solving the adaptive problem of mutual cost defusal.

Interestingly, the problem that evolution appears to have solved with reciprocal play signals is analogous to the problem that engineers solved when trying to link two electronic devices. In his book on common knowledge, Pinker (2025, pp. 56-57) describes the history of these devices and how they are analogous to the game theory of coordination:

> "A set of 'handshaking protocols' were implemented in which one device would send a message to the other, wait for a confirmation to come back, send a confirmation of the confirmation, and so on—not ad infinitum, but some number of times which depended on the clarity of the channel and the cost of an error. When the channel was a telephone line, the handshaking was audible as the call-and-response of hisses, howls, and hums that many of us remember from the era of dial-up modems and the movie You've Got Mail. Nowadays we see it in the blinking LEDs of a Bluetooth device as we try to pair it with a computer or smartphone."

We thus have a parallel between engineering and evolved design. Instead of "blinking LEDs" or the "hisses, howls, and hums" of a dial-up modem, mother nature made use of the hoots, pants, and giggles made possible by the primate vocal apparatus. We thus laugh to signal that 1) there has been a (potential) mix-up, 2) we know that the other person knows (that we know that they know) about the mix-up, 3) we are not processing the (potential) mix-up's costs —i.e., we are treating them as mere play, and 4) we know that the other person knows (that we know that they know) the costs are not being taken 'seriously.'

Formalizing cost defusal. Let us assume that a focal agent is able to initiate an exchange of play signals, and defuse the costs of a mix-up, in cases where both agents have detected the

mix-up—and thus already have private knowledge of it. When a mix-up occurs with probability e, both agents will detect it with probability p^2, allowing them to defuse its costs and continue reaping the benefits of coordination, yielding a payoff of $e\,p^2\left(\frac{V}{1-w^2}\right)$. When a mix-up occurs and only one partner detects it, which occurs with probability $2ep(1-p)$, both partners incur an initial cost, and the partner who detected it switches to the alternative action, thereby avoiding the mix-up in subsequent rounds, yielding a payoff of $2ep(1-p)\left(-C+\frac{V}{1-w^2}\right)$. Assuming an ability to sense mix-ups is already present in the population (as in inequality 1.2), the ability to defuse the costs of a mix-up will confer an advantage when:

$$\frac{(1-e)V}{1-w}+e\,p^2\left(\frac{V}{1-w^2}\right)+2ep(1-p)\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)(1-p)C}{1-w}>¿\frac{(1-e)V}{1-w}+e\,p^2\left(-C+\frac{V}{1-w^2}\right)+2ep(1-p)\left(- \tag{2.1}$$

This can be simplified to the following (see Appendix C for a derivation):

$$e\,p^2C>0 \tag{2.2}$$

Agents with an ability to defuse the costs of their mix-ups will therefore invade a population of coordinating, mix-up sensing agents. We can also demonstrate that selection for this ability increases as the play signal becomes more frequently used—and begins causing the other partner to sense the mix-up as well (see appendix D). As we can see in the above inequality, the selection pressure for mutual cost defusal (or shared laughter) will be strongest when: 1) the mix-up was especially tempting or likely to occur (large e), 2) the mix-up was especially costly (large C), and 3) both agents were likely to have sensed it (high p^2).

The cognitive architecture of mirth

We can think of the costliness, confusability, and mutual recognition of a mix-up as inputs into an emotional system: *mirth* or amusement—a system that likely overlaps with neural systems for play (Panksepp et al., 1984). The outputs of mirth might include: 1) an urge to laugh, 2) a heightened sensitivity to others' laughter, 3) a motivation to reciprocate others' laughter to the degree that it is sensed, matching the observed intensity, 4) feelings of reward in proportion to the magnitude of the costs defused by the reciprocally emerging laughter, as well as in proportion to the updated value of the coordination partnership, and 5) a deactivation of emotions that process costs, to ensure that the (potential) costs are not taken seriously by either party, and that the process of common knowledge generation is not disrupted.

This last output is especially important, because the costs of the mix-up cannot be defused if they trigger any observable fear, anger, or regret in either party. To display even a hint of these emotions—or to detect them in one's coordination partner—is to unravel the process of common knowledge generation, stoking doubts about whether you know that I know (that you know) the costs have truly been defused. If mirth is well-designed, it should prevent illusory dangers (e.g., a mix-up between ketchup and blood) from sending either party into a panic or meta-panic; it should prevent illusory transgressions (e.g., "hooking up with" my mother) from devolving into brawls; and it should prevent illusory insensitivities (e.g., "laughing out loud" at the death of my beloved Uncle) from destroying relationships. It should ensure that one does not regret a mistake that was never made or respond to a pat on the back with a roundhouse kick.

Unfortunately, represented costs spread through the brain like wildfire (Sznycer & Lukaszewski, 2019; Sznycer, 2022), making their mutual defusal a difficult adaptive problem. It is often unclear what all the relevant costs to any mix-up might be (e.g., relational, reputational, physical, hygienic, economic), or all the relevant emotions the costs might feed into. A perceived

insult could trigger anger, shame, guilt, sadness, regret, disgust, and fear in either the insulted party, the victim, or third parties, depending on the nature of the insult and its social context—and on what actions or events might be expected to follow from it. Insofar as mirth is well-designed, it might produce a *general* deactivation of emotions that process costs, in order to stop the wildfire of negative representations from spreading throughout the brain and disrupting the process of mutual cost defusal.

Mirth should also be designed to *invert* the costs of mix-ups insofar as they are being successfully defused. A larger fitness cost avoided is a larger fitness benefit gained. If the mix-up would have destroyed one's relationship, for example, then defusing its costs would be analogous to the joy of reuniting with a loved one. The defusal should be especially rewarding if it enhances—or honestly signals—the potential for mutually profitable coordination in future interactions. Insofar as the mix-up would have gone undetected or un-defused by less suitable partners, its mutual defusal should update estimates about the value of the coordination partnership, triggering additional positive affect.

Mirth is menacing. An implication of this analysis is that mirth can transform a person into something rather frightening. It may defuse their fear, making them impossible to threaten or deter. It may defuse their empathy for others' plights, transforming others' suffering into a joke. Scowls of disapproval would be all but invisible; threats of punishment and cries for help would fall on deaf ears. It is nearly impossible to get through to a mirthful individual or negotiate with them for better treatment. The only thing they can do laugh in our faces.

This might explain why mirth can, if one is not sharing it, feel hostile, creepy, or even terrifying. The best example of the menacing nature of mirth comes from the character of the Joker in *The Dark Knight*, whose mirthful disposition conveys a sense of fearlessness and

heartlessness: he cannot be bought, reasoned, or negotiated with because he takes nothing seriously. He just wants to watch the world burn, unsaddened by—or perversely delighted by—the sight of a world in flames. Other examples of scary mirth come from the *It* film series, the *Smile* film series, the cackling witch, the creepy doll, the eerie laughter that fills the halls of asylums, and the 'muahaha' of the supervillain.

Taking things seriously. We can think of the phenomenology of seriousness as the opposite of mirthfulness—a state in which social or physical costs, either potential or actual, are being carefully attended to. If I'm angry with you, then you need to process the costs that I'm threatening to inflict on you (Sell et al., 2017). If something terrible has happened, we need to take that seriously and figure out what to do about it. To take something seriously is to devote non-mirthful attention to it—to be sensitive to its actual or potential costs.

But then what is a 'serious person?' It is a person who demands non-mirthful attention—a person who can inflict costs on others, either directly through reputational or physical attacks, or indirectly by withholding valuable knowledge or resources. A serious person is someone whose interests must be respected, whose threats must be heeded, whose absence is greatly felt. In the show *Succession*, Logan Roy tells his children they are not serious people. We can now see why his words cut so deep.

And we can also see why humor is so often political. To laugh at something is to not take it seriously—to turn off our fear in the face of a threat, our anger in the face of a provocation, or our empathy in the face of a suffering victim. Politics revolves around what we ought to take seriously as a society—what problems we must work together to solve—and mirth turns these problems into jokes. Authority is maintained by stern threats of punishment and disapproval, and mirth deflates it like a whoopie cushion (Pinker, 2007, chapter 8). Politicians wield negative

emotions as political weapons, and mirth leaves them weaponless. It is therefore unsurprising that people with stronger moral identities are less able to appreciate humor and generate jokes (Yam et al., 2019).

A model of the meta-sense of humor

It is worth reflecting on the etymology of the word "fool." A fool used to refer to a clown or jester, but a more common usage refers to mix-ups. One might be fooled by a mirage, deception, or disguise, mixing up the real with the illusory. To be a fool or a foolish person is thus to be prone to these sorts of mix-ups—someone who 'can't tell their ass from a hole in the ground' or 'can't tell shit from Shinola.' To laugh *at* someone is to view them as the sort of person who is prone to disastrously mixing things up.

It is easy to see why such foolish people are undesirable as social partners. The long-run benefits of coordinating with them are low or even negative: they might waste one's time and energy on 'fool's errands' or wild goose chases. They might inflict costs in the form of injury, illness, conflict, disorganization, failed rendezvous, and destroyed possessions. To be widely recognized as a fool, oaf, klutz, slob, or ignoramus is to suffer profound damage to one's reputation—to become everyone's last choice as a coordination partner.

At the other end of the continuum are the witty wisecrackers and smooth talkers who always come in with the right zinger at the right time. It is easy to see why these charmers are so desirable as social partners. The long-run benefits of coordinating with them are high—much higher than they are with anyone else. Steering clear of the fools and flocking to the charmers would have been a major adaptive problem for our ancestors. We are descended from the early humans who chose wisely in the social marketplace, who profited the most handsomely in the economy of social life—and who were profitable choices themselves.

We can add some precision to these musings by returning to the algebra. Recall that p was the probability that a mix-up would be detected by a focal agent. This parameter would have been variable among ancestral humans—indeed it must have been variable if natural selection increased it from low to high levels. To understand how important this variation would have been for partner choice, consider a situation where one is choosing between two suitors for marriage. Let p_1 be the mix-up detection rate (or sense of humor) of the first suitor, and let p_2 be the mix-up detection rate of the second suitor ($p_1 > p_2$). These parameters could either reflect the suitor's global mix-up detection rate or their local mix-up detection rate that is specific to the marriage. Assuming we are in a post-laughter world (see inequality 2.1), and mix-up detection by one party is sufficient for mutual cost defusal (see Appendix D), a preference for the first suitor over the second will confer an advantage when:

$$\frac{(1-e)V}{1-w}+\frac{ep(1-p_1)V}{1-w}+\frac{e\,p_1(1-p)V}{1-w}+\frac{ep\,p_1V}{1-w}-\frac{e(1-p)(1-p_1)C}{1-w}>¿\frac{(1-e)V}{1-w}+\frac{ep(1-p_2)V}{1-w}+\frac{e\,p_2(1-p)V}{1-w}+\frac{ep\,p}{1-} \quad (3.1)$$

This can be simplified to the following (see Appendix E for a derivation):

$$e(p¿¿1-p_2)\frac{V+C}{1-w}>0¿(3.2)$$

As we can see in the above inequality, selection for an ability to favor the partner with the better sense of humor will be strongest insofar as: 1) each relationship is highly valuable (large V), 2) there are large costs to being mixed up with either partner (large C), 3) mix-ups are tempting or likely to occur (high e), 4) relationships are long-lasting (high w), and 5) there is a large discrepancy between the two suitors' sense of humor, either globally or locally (large p_1-p_2). Moreover, a meta-sense of humor will be particularly valuable insofar as coordination is common, there is a market for coordination partners (Barclay, 2013), and the meta-sense of

humor can be put to use in evaluating the coordination ability of multiple partners across multiple relationships.

It is worth noting that even a tiny discrepancy between the two suitors' senses of humor can be of tremendous significance over the long run. Imagine marrying someone who mixes up one-in-a-hundred phrases compared to two-in-a-hundred phrases. If we assume a million phrases exchanged over the course of a decade-long relationship (a conservative estimate; Mehl et al., 2007), that implies that favoring the first suitor will spare one the costs of 10,000 verbal mix-ups compounding over time. And that is just verbal coordination. There is also mental, gestural, tonal, physical, sexual, hygienic, navigational, coalitional, and role-based coordination, as well as the ability to laugh off whatever mix-ups have occurred instead of disastrously taking them seriously—itself a kind of mix-up. It is no wonder that a sense of humor is one of the most important characteristics in a long-term mate worldwide (Lippa, 2007), and it is no wonder that romance and comedy so often go together.

Detecting and signaling the sense of humor. Given that natural selection favored the ability to evaluate others' sense of humor, the better to choose more profitable coordination partners, the question arises: how do we do this? How do we signal to others that we have a good sense of humor in a way that is hard-to-fake?

The answer is all too familiar to us as humans: we tell funny stories, play with language, impersonate people, expose others' foolishness, use irony and innuendo, poke fun of ourselves, engage in social play, mock rivals, tease friends, and deliver funny monologues about the tempting mistakes in our thought processes. People who can skillfully perform these feats can attract many friends and allies, because one would be unable to perform them without a good sense of humor.

To tell a funny story, one needs to know which mistakes are humorous—i.e., the ones that are particularly costly and tempting—and which ones are not. To expose the foolishness of a plan or argument, one must be able to spot foolishness in the first place and avoid being foolish oneself. To skewer a rival, one needs to pinpoint all the ways they are failing to verbally, mentally, or physically coordinate with others. To be good at vibing, getting along, and wading through minefields is to know what it means to be bad at these things—and thus, to have a good sense of humor.

In the modern world, we have transitioned from social economies to market economies, where professional comedians compete to make us laugh. The modern world has given us a cornucopia of jokes in their purest form—a treasure trove of empirical data from which to draw. Let us turn to this cornucopia and see how well the theory can explain it.

Explaining the joke

Wordplay

Linguistic communication contains a minefield of misunderstandings. Words and phrases can sound similar but mean different things, as in the confusion between "The Gothic Castle" and "The Gothic Asshole" when said in a cockney accent (which were mixed up in *Arrested Development*), or the confusion between "Colonel Angus" and "cunnilingus" when said in a southern drawl (which were mixed up in a popular SNL sketch). Interpretations of utterances can be even more perilous. I can intend one meaning while you hear a very different one, and we can easily 'talk past each other' without realizing it, as in the famous Abbott and Costello routine, *Who's on First*?

These linguistic bunglings vary in how costly they are for the interactants, ranging from mild to life-ruining. I might say "pretty unlikely" while you hear "very unlikely"—a minor mix-

up. On the other hand, you might say "one in a million" while I hear "a meaningful chance," causing me to wildly overestimate my chances of mating with you, as in the following bit of dialogue from *Dumb and Dumber*:

LLOYD: I drove a long way to see you, Mary, the least you can do is level with me. What are my chances?

MARY: Not good.

LLOYD: You mean "not good" like "one out of a hundred?"

MARY: I'd say… more like one out of a million.

LLOYD: *(Pause. Excited.)* So you're telling me there's a chance! YEA!

Here, the cost of the mix-up is quite large, potentially causing Lloyd to waste time and energy pursuing a mating opportunity that will never pan out, and Mary to waste time and energy deflecting his foolish pursuits. It also reflects poorly on Lloyd's character, which is part of where the humor comes from—we get to laugh *at* his inability to take the rather large hint that Mary is uninterested. He is clearly the type of person who is prone to these sorts of mix-ups—the "dumber" of the two dumb protagonists.

Verbal mix-ups also vary in how similar or easily confusable the mixed-up options are. Some utterances, like the imperative to "Make sure he's dead," have two very different yet easily confusable interpretations: 1) "Check his pulse" or 2) "Kill him." These two meanings were disastrously mixed up in the highest-rated joke from Wiseman's (2002) international survey, in which a hunter fires a gun at his collapsed friend at the apparent behest of an emergency services worker. Verbal mix-ups may also stem from a failure to triangulate on what noun a person is describing. For example, in *Naked Gun 2½,* a detective asks for a description of an assailant, and

the witness replies, “A white guy. A mustache. About six-foot-three.” The detective replies, “Awfully big mustache.”

Malapropisms are also popular sources of comedy, as in the comic relief moment from *The Shawshank Redemption* when an inmate picks up a library book and struggles to read the cover: “The Count of Monte… Crisco by Alexandre… Dumbass.” The inmate falls prey to two malapropisms: “Cristo → Crisco” and “Dumas → Dumbass.” They are both tempting mistakes, but the second is funnier, because it’s costlier. Were the dignified Mr. Dumas present, he would likely take offense. Other humorous malapropisms include “I’m your destiny” → “I’m your density” (which got mixed up by George McFly in *Back to the Future*), “thesaurus → clitoris” (which got mixed up by a Russian obstetrician in *Nine Months*), and “rapier wit → rapist wit” (which got mixed up by Lloyd in *Dumb and Dumber*).

The written word can also invite many tempting mistakes. A recurring gag from the SNL parody of Celebrity Jeopardy featured humorous misreadings of the categories, in which the contestants mixed up “The Pen Is Mightier” with “The Penis Mightier,” “An Album Cover” with “Anal Bum Cover,” and “Catch These Men” with “Catch the Semen.” Then there are the mix-ups that arise when writers forget to use the oxford comma, and a list of three things gets mixed up with one thing and its two descriptors. For example:

1. “Highlights of his global tour include encounters with Nelson Mandela, an 800-year-old demigod and a dildo collector.”
2. “I’d like to dedicate this book to my parents, Ayn Rand and God.”
3. “I love my pets, chocolate and Elvis."

Why are 1) and 2) funnier than 3)? Because the first two are costlier mix-ups—more damaging to the speaker’s reputation and perceived sanity (not to mention the reputations of

Mandela and the Lord almighty) than the third. There is also a *larger misalignment* between the intended and received meanings in 1) and 2) then there is in 3). Loving chocolate and Elvis and naming one's pets "Chocolate" and "Elvis" are conceptually similar: we love our pets, and we name our pets after the things we love. But including Mandela in a list of memorable encounters is wildly different from transforming him into a dildo-obsessed deity, and dedicating a book to God and Ayn Rand is wildly different from claiming to be the Messiah and anointing a libertarian figurehead as the Virgin Mary.

Physical comedy

Many humorous situations involve failures to *physically* coordinate with others. This might involve trying to high-five a teammate and missing, dancing out of sync with one's partner, or dancing out of sync with the music itself (as in Elaine Bennice's spasmodic dancing style in *Seinfeld*). If you and I are passing a ball back and forth, that requires mental coordination. We both need to know that the other one knows that we are ready to catch the ball. But if I suddenly get distracted by a pretty woman walking by, and the ball hits me in the face, our failure to coordinate becomes costly for me (especially if I'm trying to impress the woman) and humorous. Similar coordination failures occur when we wave at the wrong person or endlessly teeter back and forth trying to pass each other in a hallway, as we saw in table 1.

The bungling of tools and gizmos is another popular source of physical comedy. Human-made artifacts represent a kind of coordination game between the designer and the user, and physical comedy often derives from mix-ups about what a tool is for or how to use it. For example, a cigar is meant to be lit on one side and inhaled from the other. A bidet is meant for washing one's rear-end. Potpourri is meant for decoration. But the user might get mixed up and insert the lit end of a cigar into his mouth (as Kramer does while trying to appear classy in an

episode of *Seinfeld*), innocently drink from a bidet (as Mick Dundee does in *Crocodile Dundee*), or casually munch on a handful of potpourri (as Joey does in an episode of *Friends*). Other artifact-related mix-ups include missing the chair as the waiter pulls it out for you, continuing to pour wine into a glass after it overflows (perhaps because one is distracted by cleavage), or forgetting to pull one's thumb away before hammering a nail.

Physical comedy is also rife with substances, sizes, or bodily states that get mixed up. One might confuse a bit of semen dangling from a man's ear for a bit of hair gel (as in the gross-out moment from *There's Something about Mary*), a bottle of urine for a bottle of beer (as in *Dumb and Dumber*), congenital dwarfism for childhood (as in the darkly comedic finale of *In Bruges*), a 'pants tent' for an erection (as in the pilot of *Curb Your Enthusiasm*), a small man when seated for a large man when standing (as when Marty McFly unwittingly picks a fight with the enormous Biff in *Back to the Future*), and an inadequate-sized penis for a normal-sized penis afflicted with 'shrinkage' (as in the classic episode of *Seinfeld*).

An additional layer to physical comedy might derive from the struggle to coordinate among observers in the aftermath of a fall, collision, or projectile impact. Has a serious injury taken place? Is the person dead or unconscious? Or have they merely slipped on a banana peel, walked into a glass door, or gotten hit in the back of the head with an orange (as in the alleged 'run-by fruiting' in *Mrs. Doubtfire*)? Here the humor may derive from an anticipated mix-up, either between the stunned victim and their audience, or between each member of the audience and each other, about whether a severe injury has taken place. Either mix-up—between the severe and the minor, or between the minor and the severe—would be costly. One does not want to call the paramedics in the aftermath of a run-by fruiting, nor point and laugh at the fallen

clown on stilts, as blood comes out of his mouth, as in the darkly comedic moment from *Billy Madison.*

The final layer of physical comedy may come from a schism in the human mind—a conflict between two ways of representing reality—described by Pinker in How the Mind Works (1997, p. 550):

> "Slapstick humor runs off the clash between a psychological frame, in which a person is a locus of beliefs and desires, and a physical frame, in which a person is a hunk of matter obeying the laws of physics."

Pinker used the word "clash," but I think it is more insightful to use the word with "mix-up." Mentally framing a situation in the same way as others is a coordination game with significant costs: a surgeon in the operating theater would not want to tearfully apologize to the body he is cutting open, and a pair of dear friends would not want to treat each other as coat hangers, footrests, or bongo drums. When we sense a potential mix-up between two compelling frames of a situation—like an important businessman on his way to an important meeting and a hunk of matter obeying the laws of physics—the instinctive response is to laugh.

Breaking all the rules

Rules have functions as well as artifacts, and one can become just as easily mixed-up about them. One might confuse the letter of the law for its spirit, as when the duo from *Dumb and Dumber* deny stealing cash from a briefcase because they replaced the stolen bills with IOUs. Or one might confuse the spirit of the law for its letter, as in the awkward scene from *Office Space* where the manager of a Tchotchke's restaurant tries to reprimand a waitress for wearing the minimum number of pieces of flare. Or one might mix up an actual law with a nonexistent law, as when the feckless attorney from *Arrested Development* confidently declared

that a husband and wife cannot be accused of the same crime—a mix-up with spousal testimonial privilege, where a husband cannot be forced to testify against his wife.

Morality is another set of rules that are rife with the possibility of perilous misunderstandings. Moral rules are (supposedly) designed to be followed and enforced to benefit the larger community. But there's a tempting mistake here: one can foolishly think a moral rule is designed to be enforced (but not followed) to enrich oneself (but not the larger community). This can lead to one's peers being mixed-up about one's virtue: they might think one is upstanding for enforcing the rule when one is treacherous for flouting it. These coordination failures are part of what we call *hypocrisy*, and their confusable nature explains why hypocrisy can sometimes be humorous (assuming the mix-up is made salient)—and is often the centerpiece of political satire.

Moral mix-ups might also involve the *magnitude* of a transgression. One might mistake a grave crime for a minor faux pas, as occurred in the action-comedy *True Lies*, when Harry Tasker (played by Arnold Schwarzenegger) is outed as a spy to his wife Helen of 15 years. Helen, dumbstruck by the revelation, breathlessly asks Harry if he has ever killed anyone. Harry casually replies: "Yea but they were all bad." The opposite mix-up—where a minor faux pas is mixed up with a grave crime—was played upon by George Carlin when he joked about "one of those farts" that could "end a marriage." Here the humor comes from how tempting it is, in the midst of a truly horrendous fart, to consider divorce an appropriate response.

Seinfeld was filled with similar mix-ups between unforgivable sins and minor annoyances (e.g., being a close-talker, double-dipping a chip). But the main comedic innovation of *Seinfeld* was a mix-up about the unspoken rules of entertainment. Other sitcoms of the time featured characters recovering from setbacks, learning valuable lessons, and growing as people. Seinfeld

featured characters learning nothing, remaining shallow and cynical, and ending every episode more miserable than when they began. Designed to be a show about nothing, *Seinfeld* was itself mixed up about the purpose of a show—namely to be about something meaningful, like love or friendship, and impart wisdom to its viewers.

Perhaps the most famous mix-up about the rules of entertainment comes from *The Producers,* the story of a duo of scheming investors trying as hard as they can to engineer a Broadway flop—a deliciously inverted mix-up about the purpose of producing a musical. As if the premise weren't funny enough, a second mix-up gets layered on top of it: when the musical premiers at a packed theatre, the audience gets mixed up about its quality, mistaking the tasteless trainwreck for an ingenious satire. In one of the most classic comedic lines of all time, Max Bialystock bemoans: "We got the wrong play, the wrong director, the wrong cast. Where did we go right?"

Breaking all the roles

Social life is just as awash with rules as with *roles*—producers, neighbors, lovers, professionals, user and owner, expert and novice, doctor and patient, buyer and seller—where partners mutually benefit from coordinating on what roles they have been assigned and what is expected of each role (Jara-Ettinger & Dunham, 2025). But it is all too easy to get mixed-up here, as when a moocher (e.g., Kramer from *Seinfeld*) takes advantage of his generous neighbor who tells him to think of his apartment as his own (as Jerry regrettably says in *Seinfeld*), or when a depressed man confesses his deeply felt insecurities to his doctor, albeit a doctor of dentistry (as in the awkward scene from *Punch Drunk Love*).

Social roles pertaining to age, gender, race, and class are even more popular sources of comedy. We laugh when babies act like adults (as in *Boss Baby, Family Guy,* and *Look Who's*

Talking), when men act like little boys (as in *Billy Madison*, *Big*, and *Jack*), when women act like little girls (as in *Freaky Friday* and *13 Going on 30*), when men act like women (as in *Tootsie*, *Mrs. Doubtfire*, and *Some Like it Hot*), when women act like men (as in *Mulan*, *Twelfth Night*, and *She's the Man*), when black men act like privileged white women (as in *White Chicks*), or when upper-class people act like lower-class people (as in *Trading Places*, *Pretty Woman*, and *My Fair Lady*).

Absurdist humor

The model entails that we laugh to defuse the costs of mix-ups, but it does not entail that we always pinpoint what was mixed up with what. It is certainly helpful to figure this out, and we often do, but it is not strictly necessary. All that is necessary is that the costs of the mix-up—whatever its source—are defused, and whoever was mixed-up stops doing whatever foolish thing they were doing (see Appendix D). This brings us to a specific type of humor we might call *absurdist humor*: mix-ups whose source is unclear or overdetermined.

For example, consider the mix-up between "Dumas" and "Dumbass" described previously. Let us imagine someone who has never heard of Alexandre Dumas—and who therefore does not know what surname "Dumbass" was mixed up with. Such a person may still find the moment funny. Why? Because they will make the following inference to the best explanation: it is very unlikely that any real person has the surname "Dumbass" (that would be absurd), so there must have been a mix-up.

Or consider the joke from *Arrested Development*, where the feckless attorney declares that a husband and wife cannot be accused of the same crime. Even if one has never heard of spousal testimonial privilege, one could reasonably infer from the absurdity of the lawyer's claim that he was mixed up about something—and find the moment funny. Mix-up Theory therefore

converges with incongruity theories of humor (see Monroe, 1967), but it specifies *when* incongruous situations will be judged as humorous: when observers infer that foolishness is the best explanation for the incongruity.

As another example, consider the beginning moments of *Naked Gun 2 ½* (itself a bit of absurdist humor about how sequel numbers work). The opening credits are displayed as a police car cruises through urban streets at night, while jazzy music plays—a good way to set the tone for a noir film. But then the montage transitions into the police car driving through a bullfighting arena, bumping into a series of bumper cars, driving down a bowling lane and knocking down the pins, and emerging from a woman's birth canal. It is apparent from the sequence that someone is deeply confused about something. Maybe there has been a mix-up about the purpose of an opening credit sequence, what kind of genre noir is, where police cars are permitted to drive, how large they are, or where they come from. Regardless, the viewer does not need to know exactly what was mixed up with what to have a good laugh. The film itself is playing a character—a fool—and insofar as the audience is in on the joke of the film's foolishness (as revealed by its very title), they will find the absurdity humorous.

Comic timing

Consider all the subtle coordination games we play when we have a conversation.

- **The volume game**. If we speak too softly, we won't hear each other. If we speak too loudly, we might frighten each other.
- **The pacing game**. If we speak too quickly, we won't understand each other. If we speak too slowly, we might bore each other.

- **The gesture game**. If we hold our arms stiffly at our sides, we might struggle to communicate. If we wildly flail our arms around, we might disorient each other or accidentally hit each other.
- **The facial expression game**. If our faces are blank or expressionless, our emotions will be ignored or misjudged. If we contort our faces in exaggerated grimaces, our emotions will be overestimated—or will make us seem odd or mentally unstable.
- **The proximity game**. If we stand too far away from each other, we might signal we are disgusted or afraid of each other. If we stand too close to each other, we might misinterpret the proximity as romantic or threatening.
- **The tone game**. If we speak in an overly monotone way, the emotional undercurrents may be lost or underestimated. If we speak in an overly sing-songy way, we may distract the listener from the words themselves or appear mentally unstable.
- **The timing game**. If we respond too quickly to each other's utterances, we might accidentally interrupt each other or convey hostility. If we respond too slowly or insert an awkward silence, we might appear distracted or dimwitted.

Comedic actors often have exaggerated mannerisms, from the deadpan delivery of Charles Grodin to the exaggerated clownishness of Jim Carrey. Good comedic actors can make the jokes they are delivering even funnier by layering in subtle mix-ups on top of the joke itself—mix-ups involving pacing, body language, facial expressions, volume, tone, or proximity to the person they are talking to. Some comedians have good 'comic timing,' which may refer to a subtly discoordinated sense of timing that is temptingly out of sync with dialogic conventions—a kind of rhythmic mix-up akin to dancing badly (like Elaine in *Seinfeld*). Such rhythmic mix-ups might accentuate the "punch" of an unexpected punchline or the twist to a convoluted scenario.

In some cases, however, good comic timing might refer to the opposite strategy: a kind of pacing that is smoothly in-sync with dialogic conventions, in a way that allows a whopping absurdity to casually slip out. This might add a different layer of humor to the joke: a mix-up between the nonchalance of the delivery and the shock of what is being delivered. Navigating between such comedic strategies—and understanding which jokes are better served by which strategy—can explain much of the art of comedic acting.

Mistaken identities

If there is one shtick that has stood the test of time, it is mistaken identities, explaining why this plot device still makes people laugh at Shakespearean comedies centuries after they were written. The art of *impersonation*, similarly, can be explained as the art of engineering confusability between the impersonator's voice, vibe, or visage and the target's, made all the more costly—and funny—if the impersonator says or does foolish things that damage the target's reputation. Similar mix-ups are nearly everywhere in comedic entertainment. One might mistake a lipstick-besmirched Jon Lovitz for Adolph Hitler (as in *Rat Race*), male strippers dressed as police officers for actual police officers (as in *Arrested Development*), a liberal gay father for a straight conservative mother (as in *The Birdcage*), a holocaust survivor for a contestant on the reality show *Survivor* (as in *Curb Your Enthusiasm*), a fully-grown man for a rhinoceros newborn (as when Jim Carrey emerges naked from a broken-down mechanical rhino in *Ace Ventura: When Nature Calls*), and one's own identity as a black person for one's own identity as a member of the Aryan race (as in the *Chapelle Show* sketch about a blind, black white supremacist).

Status games, embarrassment, humiliation, and cringe

Social status is, in large part, a coordination game. We all benefit by coordinating on who is at the top or bottom of the social ladder, and we pay a cost if we're mixed up about it. For instance, if I fawn over a social pariah, then my reputation will be tarnished by association. If I diss a beloved and venerated figure, then I'll suffer the wrath of that figure and their many loyal allies. When it comes to coordinating on who's high or low status, the stakes are high.

How do we coordinate? One way is to use widely observed or gossiped-about actions as focal points of common salience (Schelling, 1980, chapter 3). If someone behaves in a way that creates common knowledge of their ineptitude or disgustingness, it represents a kind of 'green light' to impugn, exclude, or side against them. If someone behaves in a way that creates common knowledge of their sagacity or integrity, it represents a kind of 'green light' to praise, include, and side with them.

But of course, in the dance of social hierarchy formation, it is easy to get out of sync with others. What if someone elicits others' disgust, but in a way that is commonplace and relatable—e.g., they pass gas? What if a woman exposes her nipple in an office Christmas card, but accidentally and unknowingly (as in the classic episode of *Seinfeld*)? These sorts of actions create a coordination problem: do we, as a community, lower the person's status or not? In addition to whatever confusability is inherent to the faux pas itself, there is the added confusability about whether the transgressor's status has plummeted. Much in the same way slapstick plays off of a mix-up between severe vs. minor bodily damage, farts and wardrobe malfunctions play off of a mix-up between severe vs. minor reputational damage. Disgraces are to embarrassments as drive-by shootings are to run-by fruitings.

It is worth comparing the function of shame to the function of embarrassment. If shame is a response to *clear* social devaluation (Sznycer et al., 2016), then embarrassment is a response to

tempting, ambiguous, or confusable social devaluation. Whereas shame functions to conceal or mitigate *clear* status loss in the wake of a serious transgression, embarrassment functions to tip the scales away from *potential* status loss in the wake of a slight, faux pas, or Freudian slip. If the embarrassed person succeeds and the coordinated equilibrium stays at the "status" quo, then everyone can have a good laugh together and move on. But if the embarrassed person fails and the coordinated equilibrium drops the person down a rung on the social ladder, then the embarrassment sours into painful *humiliation*, which we might define as the kind of shame that flows from unsuccessful embarrassment, or the kind of shame that flows from being laughed *at* —from being seen as a fool, oaf, slob, klutz, stinker, or ignoramus.

Tipping the scales away from humiliation and toward shared laughter (or laughter *with* others) is thus a major adaptive problem for the embarrassed person, with large, cascading effects on their reputation. What can the embarrassed person do to ensure that the scales tip in their favor? They might 1) try to persuade others that the faux pas was an honest or forgivable mistake that anyone might have made, 2) frame the faux pas as nondiagnostic of their intentions or character (which is easier if they are recounting a faux pas from the distant past), 3) nervously laugh at the faux pas themselves in an attempt to make others do the same (to defuse whatever social costs may have been incurred), or 4) signal to others that they are, in fact, painfully embarrassed.

This last tactic is especially important, because a painfully embarrassed person is unlikely to commit similar faux pas in the future. Additionally, the fact that the person is aware of the faux pas (and aware that others are aware) suggests that they have at least some ability to coordinate with others. Alternatively, if the target is utterly oblivious to the mistake they just blundered into, then that suggests they are a very poor coordination partner—a boor, an oaf, a

buffoon—because they will surely blunder into many more mix-ups in the future. Blushing may have thus evolved as an honest signal of self-aware, painful, and thus deterrent embarrassment, with its visually conspicuous and hotly felt character increasing the likelihood that the target and each member of their audience—as well as each member of the audience and every *other* member of the audience—will achieve common knowledge of the target's painful embarrassment (see Pinker, 2025, chapter 6 for a similar account). Such common knowledge would be advantageous for the target insofar as it lowers the probability and severity of their social devaluation, and it would be advantageous for the audience insofar as it prevents them from coordinating against an otherwise valuable social partner.

This might shed light on the nature of nervous laughter. A person who is prone to nervous laughter is someone who is too easily embarrassed or humiliated—constantly thinking they've blundered into various slights and peccadilloes—and in constant need to defuse their imagined costs. Insofar as one's nervous laughter is unreciprocated, it may increase the perception that one is failing to coordinate with others, creating feelings of *awkwardness*. We might thus define an awkward situation as one in which the interactants are failing to coordinate but unable to create common knowledge of the failure, either because one or both parties is too inept or oblivious to recognize it, or because creating common knowledge of it would have negative repercussions for one or both parties (see also Pinker, 2007, chapter 8).

As an example of the latter type of awkwardness, consider the nightmarish meal from *There's Something About Mary,* where Ben Stiller's character is unable to acknowledge that Mary has his dried semen in her hideously stuck-up hair. If Mary knew that he knew that her hair was sticking up, and why it was sticking up, it would destroy the evening and probably their relationship. An awkward situation is thus the opposite of vibing—a state of painful

discoordination, often perched on the brink of perilous coordination. Such a situation will feel aversive to the interactants but potentially humorous to outsiders, who can create common knowledge of the mix-up(s), even though the interactants could not.

But then what does it mean to be cringe? We can think of the cringey feeling as a combination of awkwardness and vicarious humiliation. The feeling arises from someone who is oblivious to their ongoing devaluation as a coordination partner (e.g., someone who is coming off as foolish or boorish) because they are too inept to recognize it—an ineptitude which itself contributes to their devaluation. We cringe at the wannabe pop star singing her heart out but sounding like a strangled cat, or the 'cool teacher' who turns his cap backwards and delivers a lesson plan in the form of a rap. The cringey feeling combines the internally simulated humiliation of the target with the awkwardness of being unable to coordinate with the target or with others. No one wants to be 'the asshole' who tells the tone-deaf singer to get off the stage, yet everyone wants the cat to stop being strangled. The failure to coordinate on who should be the bearer of bad news, or whether the bad news should be delivered, is where the feelings of awkwardness and vicarious humiliation converge. Again, the cringey feeling will be aversive to the interactants but potentially humorous to outsiders, assuming they don't empathize too much with the interactants.

Discussion

A theoretical synthesis

Mix-Up Theory offers a synthesis of existing evolutionary theories of humor. It clarifies the precise nature of the information being encrypted by our inside jokes (Flamson & Barrett, 2008; Smaldino et al., 2018), the specific type of error that humor evolved to correct (i.e., mix-ups in iterative coordination games; Hurley et al., 2013), and the particular trait (i.e.,

coordination ability, both in general and with specific people) that joke-telling advertises to potential mates and friends (Bressler et al., 2016; Driebe et al., 2021). It explains why humor and play go together (Simler & Hanson, 2017, chapter 8), and why turning a serious situation into a humorous situation can be threatening to those with power or authority.

The theory also provides an ultimate foundation for more proximate-level theories of humor. It explains why humorous situations are simultaneously perceived as wrong and okay (McGraw & Warren, 2010). That is, if laughter and mirth evolved to defuse the costs of mix-ups, then laughter and mirth will be associated with perceiving the mix-up's cost (the wrong part) and its defusal (the okay part). It also explains why humorous situations are often judged as incongruous. Incongruous (or absurd) situations arise when foolishness is judged as the best explanation for the incongruity. The theory sheds light on the nature of vibing, awkwardness, faux pas, cringe comedy, humiliation, mockery, satire, status games, comic timing, slapstick, puns, malapropisms, impersonations, Freudian slips, practical jokes (or engineered discoordination), seriousness, mistaken identities, why laughter is contagious, why romance and comedy go together, the joy of a good laugh, the 'muahaha' of the supervillain, the attractiveness of a witty joke-teller, the repulsiveness of a foolish oaf, the adaptive function of mirth, the adaptive function of embarrassment, the precise part of reality the humor apparatus evolved to sense, and why it is of the utmost importance—in terms of our survival and reproduction as hypersocial animals—to sense it.

Future directions

The model here makes a host of convenient, simplifying assumptions. It assumes two players when most real coordination involves entire communities. It assumes symmetry in costs, benefits, and confusability when most real interactants face different benefits from successful

coordination (e.g., if one partner values the relationship more), face different costs if there is a mix-up (e.g., if one party is judged as the foolish one), and have different probabilities of mixing up the two options (e.g., if one party is a cultural outsider). It ignores the correlations between the objective confusability (or similarity) of the two options, the subjective probability of a specific agent sensing when they are mixed up, and the likelihood of agents developing workarounds (e.g., using different words) when mix-ups are especially likely. It assumes there is one game, when in real life, the coordination games we play are nested—games within games within games. It assumes a single mix-up between two options, when many humorous situations are messy tangles of mix-ups, deriving from many interleaving behavioral and mental options. The model ignores the underlying signaling dynamics of laughter, the cognitive challenge of differentiating a laugh 'with' and a laugh 'at,' the ensuing inferences about the insiders who get a joke and the outsiders who don't (Smaldino et al., 2018), and the arms race between fake laughers and fake-laugh detectors (Bryant et al., 2018). It assumes two identical equilibria, when real coordination games have multiple equilibria, with some more preferable to all players (as in the stag hunt game), and with some more preferable for some players than others (as in the battle of the sexes game; Pinker et al., 2025, chapter 3). The model has barely scratched the surface of the underlying complexity of coordination, mix-ups, and common knowledge. Nevertheless, I hope to have convinced you that humor is *about* these things. If I have succeeded in my goal, then humor will seem less like a Darwinian mystery and more like a tractable scientific problem.

The model implies a pitfall for humor research itself: the inability of scholars to 'take it seriously.' Mirth deactivates emotions that process costs—fear, anger, guilt—and we often gauge how important a topic is by the intensity of the negative emotions it can elicit. This may explain why psychologists have devoted so much attention to prejudice, inequality, and intergroup

conflict, while neglecting less serious topics like play, humor, and entertainment. The prestige economy in academia often centers around how serious one's research is. As a result, centuries of scholarship have failed to answer the most basic question about humor: what is it *for*?

My goal is not only to provide a satisfying answer to this question, but to argue that social science has been wrong to ignore it. Humor is part of the fabric of social life, weaving together relationships big and small. It can reveal nude emperors, transform the sacred into the profane, and make our status games collapse and invert (Pinsof, 2026). It can turn acquaintances into friendships, enemies into friends, and trysts into marriages. Without humor, our ancestors would have been unable to survive in their hypersocial, linguistic niche: they would have been so mired in mix-ups that hypersociality and communication would be unprofitable—unable to evolve by natural selection. For many of us, humor makes friends worth having and life worth living: a good laugh is among the greatest joys we can experience.

Yet when we focus our attention on humor, it does not seem like a serious topic, does it? The seeming is an illusion—a byproduct of the psychology of humor itself. We must transcend this quirk of the mind, however sensible its adaptive logic may be. We must learn to take humor seriously.

Acknowledgements: Thank you to Pat Barclay, Julien Lie-Panis, and the Center for Evolutionary Psychology for helpful comments on an earlier version of this paper.

Appendix A: Derivation for Humor Sensing

$$\frac{(1-e)V}{1-w}+ep\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)C}{1-w}>\frac{(1-e)V-eC}{1-w} \quad (4.1)$$

Subtracting $\frac{(1-e)V}{1-w}$ from both sides yields:

$$ep\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)C}{1-w}>\frac{-ec}{1-w} \quad (4.2)$$

Multiplying out $e(1-p)C$ and collecting terms yields:

$$ep\left(-C+\frac{V}{1-w^2}\right)+\frac{epC}{1-w}>0 \quad (4.3)$$

Dividing both sides by p yields:

$$e\left(-C+\frac{V}{1-w^2}\right)+\frac{eC}{1-w}>0 \quad (4.4)$$

Multiplying out the left parenthetical yields:

$$-eC+\frac{eV}{1-w^2}+\frac{eC}{1-w}>0 \quad (4.5)$$

Multiplying each side by $1-w^2$:

$$-eC(1-w^2)+eV+eC(1+w)>0 \quad (4.6)$$

Multiplying out the parenthetical and collecting terms yields:

$$eCw^2+eV+eCw>0 \quad (4.7)$$

Factoring out e yields:

$$e(V+Cw+Cw¿¿2)>0¿ \quad (4.8)$$

Appendix B: Stability Conditions for Humor Sensing

I assume that when two humor sensors mutually sense a mix-up, they do not both switch to the alternative option (itself a mix-up worth sensing), but instead successfully coordinate on subsequent rounds by converging on one of the two equilibria (e.g., the more conventional one). Under this assumption, humor sensing will yield higher payoffs when common than when rare if:

$$\frac{(1-e)V}{1-w}+ep^2\left(-C+\frac{V}{1-w^2}\right)+2ep(1-p)\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)(1-p)C}{1-w}>¿\frac{(1-e)V}{1-w}+ep\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)}{1-w} \quad (5.1)$$

Subtracting $\frac{(1-e)V}{1-w}$ from both sides and collecting terms yields:

$$(ep^2+2ep(1-p))\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)(1-p)C}{1-w}>ep\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)C}{1-w} \quad (5.2)$$

Assuming monotonicity of subtraction, we can separate this into:
$(2ep(1-p)+ep^2)\left(-C+\frac{V}{1-w^2}\right)>ep\left(-C+\frac{V}{1-w^2}\right)$ and $\frac{e(1-p)(1-p)C}{1-w}<\frac{e(1-p)C}{1-w}$. Starting with the first inequality:

$$(2ep(1-p)+ep^2)\left(-C+\frac{V}{1-w^2}\right)>ep\left(-C+\frac{V}{1-w^2}\right) \quad (5.3)$$

Dividing both sides by $\left(-C+\frac{V}{1-w^2}\right)$, multiplying the parenthetical, and collecting terms yields:

$$ep - e\,p^2 > 0 \quad (5.4)$$

Dividing both sides by ep yields:

$$1 > ep \quad (5.5)$$

Since $0 < e < 1$ and $0 < p < 1$, the inequality is true. Moving to the second inequality:

$$\frac{e(1-p)(1-p)C}{1-w} < \frac{e(1-p)C}{1-w} \quad (5.6)$$

Dividing both sides by $\frac{e(1-p)C}{1-w}$ and collecting terms yields:

$$0 < p \quad (5.7)$$

The inequality is also true; thus, humor sensing is evolutionarily stable.

Appendix C: Derivation for Cost Defusal

$$\frac{(1-e)V}{1-w} + e\,p^2\left(\frac{V}{1-w^2}\right) + 2ep(1-p)\left(-C + \frac{V}{1-w^2}\right) - \frac{e(1-p)(1-p)C}{1-w} > ¿\, \frac{(1-e)V}{1-w} + e\,p^2\left(-C + \frac{V}{1-w^2}\right) + 2ep(1-p)\left(\quad (6.1)$$

Adding $\frac{e(1-p)(1-p)C}{1-w}$ to both sides, subtracting $\frac{(1-e)V}{1-w}$ from both sides, and subtracting $2ep(1-p)\left(-C + \frac{V}{1-w^2}\right)$ from both sides yields:

$$e\,p^2\left(\frac{V}{1-w^2}\right) > e\,p^2\left(-C + \frac{V}{1-w^2}\right) \quad (6.2)$$

Subtracting the right side from the left side yields:

$$e\,p^2\left(\frac{V}{1-w^2} + C - \frac{V}{1-w^2}\right) > 0 \quad (6.3)$$

Collecting terms yields:

$$e\,p^2 C > 0 \quad (6.4)$$

Appendix D: Stability Conditions for Cost Defusal

I assume that as co-optation of the play signal becomes more frequent, it becomes sufficient for only one party to sense a mix-up in order for its cost to be defused. Sending the play signal may alert the other party of the mix-up even if they had not sensed it initially. Or, if the other party still cannot sense the mix-up after receiving the play signal, they may still benefit from joining their coordination partner in defusing its cost, whatever the cost may be. Under this assumption, sending the play signal after sensing a mix-up (or sensing that one's partner has sensed one) will yield higher payoffs when common than when rare if:

$$\frac{(1-e)V}{1-w}+e\,p^2\left(\frac{V}{1-w^2}\right)+2e(1-p)\left(-C+\frac{V}{1-w^2}\right)-\frac{e(1-p)(1-p)C}{1-w}>¿\frac{(1-e)V}{1-w}+e\,p^2\left(-C+\frac{V}{1-w^2}\right)+2ep(1-p)\left(-C\right.$$

(7.1)

Adding $\frac{e(1-p)(1-p)C}{1-w}$ to both sides, subtracting $\frac{(1-e)V}{1-w}$ from both sides, subtracting $e\,p^2\left(\frac{V}{1-w}\right)$ from both sides, and subtracting $ep(1-p)\left(-C+\frac{V}{1-w}\right)$ from both sides yields:

$$ep(1-p)\left(\frac{V}{1-w}\right)>ep(1-p)\left(-C+\frac{V}{1-w}\right)$$

(7.2)

Dividing both sides by $ep(1-p)$ yields:

$$\frac{V}{1-w}>-C+\frac{V}{1-w}$$

(7.3)

Subtracting $\frac{V}{1-w}$ from both sides and multiplying each side by (-1) yeilds:

$$C>0$$

(7.4)

Thus, sending a play signal or “laughing” after sensing a mix-up, or sensing that one’s partner has sensed a mix-up, is evolutionarily stable.

Appendix E: Derivation for the Meta-Sense of Humor

$$\frac{(1-e)V}{1-w}+\frac{ep(1-p_1)V}{1-w}+\frac{e\,p_1(1-p)V}{1-w}+\frac{ep\,p_1V}{1-w}-\frac{e(1-p)(1-p_1)C}{1-w}>¿\frac{(1-e)V}{1-w}+\frac{ep(1-p_2)V}{1-w}+\frac{e\,p_2(1-p)V}{1-w}+\frac{ep\,p}{1-}$$

(8.1)

Subtracting $\frac{(1-e)V}{1-w}$ from both sides and pooling numerators yields:

$$\frac{ep(1-p_1)V+e\,p_1(1-p)V+ep\,p_1V}{1-w}-\frac{e(1-p)(1-p_1)C}{1-w}>¿\frac{ep(1-p_2)V+e\,p_2(1-p)V+ep\,p_2V}{1-w}-\frac{e(1-p)(1-p_2)C}{1-w}$$

(8.2)

Multiplying out the leftmost numerator on each side and subtracting $\frac{epV}{1-w}$ from both sides yields:

$$\frac{e\,p_1 V - epp_1 V}{1-w} - \frac{e(1-p)(1-p_1)C}{1-w} > \frac{e\,p_2 V - epp_2 V}{1-w} - \frac{e(1-p)(1-p_2)C}{1-w} \quad (8.3)$$

Factoring out $(1-p)$ on each side yields:

$$\frac{e\,p_1 V(1-p)}{1-w} - \frac{e(1-p)(1-p_1)C}{1-w} > \frac{e\,p_2 V(1-p)}{1-w} - \frac{e(1-p)(1-p_2)C}{1-w} \quad (8.4)$$

Dividing each side by $(1-p)$ yields:

$$\frac{e\,p_1 V}{1-w} - \frac{e(1-p_1)C}{1-w} > \frac{e\,p_2 V}{1-w} - \frac{e(1-p_2)C}{1-w} \quad (8.5)$$

Multiplying out the parentheticals, pooling numerators yields:

$$\frac{e\,p_1 V - eC + ep_1 C}{1-w} > \frac{e\,p_2 V - eC + ep_2 C}{1-w} \quad (8.6)$$

Adding $\frac{eC}{1-w}$ to both sides yields:

$$\frac{e\,p_1 V + ep_1 C}{1-w} > \frac{e\,p_2 V + ep_2 C}{1-w} \quad (8.7)$$

Removing $e\,p_1$ and $e\,p_2$ from the numerators yields:

$$e\,p_1\frac{V+C}{1-w}>e\,p_2\frac{V+C}{1-w} \quad (8.8)$$

Subtracting the right side from the left side yields:

$$e(p¿¿1-p_2)\frac{V+C}{1-w}>0¿(8.9)$$

REFERENCES

Apicella, C. L., & Silk, J. B. (2019). The evolution of human cooperation. *Current Biology*, *29*(11), R447-R450.

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior*, *34*(3), 164-175.

Bekoff, M. (1995). Play signals as punctuation: The structure of social play in canids. *Behaviour*, *132*(5-6), 419-429.

Binmore, K., & Samuelson, L. (1994). An Economist's Perspective on the Evolution of Norms. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft*, *150*(1), 45–63.

Bressler, E. R., Martin, R. A., & Balshine, S. (2006). Production and appreciation of humor as sexually selected traits. *Evolution and Human Behavior*, *27*(2), 121-130.

Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Amir, D., Chávez, B., ... & Zhou, Y. (2018). The perception of spontaneous and volitional laughter across 21 societies. *Psychological science*, *29*(9), 1515-1525.

Chwe, M. S. Y. (2013). *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.

Cole, T., & Teboul, B. (2004). Non-zero-sum collaboration, reciprocity, and the preference for similarity: Developing an adaptive model of close relational functioning. *Personal Relationships*, *11*(2), 135-160.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149-1157.

Driebe, J. C., Sidari, M. J., Dufner, M., Von der Heiden, J. M., Bürkner, P. C., Penke, L., ... & Arslan, R. C. (2021). Intelligence can be detected but is not found attractive in videos and live interactions. *Evolution and Human Behavior*, *42*(6), 507-516.

Dunbar, R. I. (2009). The social brain hypothesis and its implications for social evolution. *Annals of human biology*, *36*(5), 562-572.

Flamson, T., & Barrett, H. C. (2008). The encryption theory of humor: A knowledge-based mechanism of honest signaling. *Journal of Evolutionary Psychology*, *6*(4), 261-281.

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). Conforming to coordinate: Children use majority information for peer coordination. *British Journal of Developmental Psychology*, *33*(1), 136-147.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton University press.

Jara-Ettinger, J., & Dunham, Y. (2025). The Institutional Stance. *Behavioral and Brain Sciences*, 1–62. doi:10.1017/S0140525X25103427

Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *9*(4), 156-185.

Kisko, T. M., Himmler, B. T., Himmler, S. M., Euston, D. R., & Pellis, S. M. (2015). Are 50-kHz calls used as play signals in the playful interactions of rats? II. Evidence from the effects of devocalization. *Behavioural processes*, *111*, 25-33.

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767-768.

Krupka, E. L., Weber, R., Crosno, R. T., & Hoover, H. (2022). “When in Rome”: Identifying social norms using coordination games. *Judgment and Decision Making*, *17*(2), 263-283.

Hurley, M. M., Dennett, D. C., & Adams Jr, R. B. (2013). *Inside jokes: Using humor to reverse-engineer the mind*. MIT press.

Lippa, R. A. (2007). The preferred traits of mates in a cross-national study of heterosexual and homosexual men and women: An examination of biological and cultural influences. *Archives of sexual behavior*, *36*(2), 193-208.

Maglieri, V., Vantaggio, F., Pilenga, C., Böye, M., Lemasson, A., Favaro, L., & Palagi, E. (2024). Smiling underwater: Exploring playful signals and rapid mimicry in bottlenose dolphins. *Iscience*, *27*(10).

Matsusaka, T. (2004). When does play panting occur during social play in wild chimpanzees?. *Primates*, *45*(4), 221-229.

McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological science*, *21*(8), 1141-1149.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science*, *317*(5834), 82-82.

Monroe, D. H. (1967): Humor. In: Edwards, P. (ed.): Encyclopedia of Philosophy. New York: Macmillan.

Panksepp, J., Siviy, S., & Normansell, L. (1984). The psychobiology of play: theoretical and methodological perspectives. *Neuroscience & Biobehavioral Reviews*, *8*(4), 465-492.

Pinker, S. (1997). *How the Mind Works*. New York: Norton.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *PNAS*, *107*(2), 8993-8999.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. Penguin.

Pinker, S. (2025). *When Everyone Knows That Everyone Knows...: Common Knowledge and the Science of Harmony, Hypocrisy and Outrage*. Random House.

Pinsof, D. (2026). The evolution of social paradoxes. *American Psychologist.* Advance online publication. *https://doi.org/10.1037/amp0001692*

Ross, M. D., Owren, M. J., & Zimmermann, E. *(*2009*).* Reconstructing the evolution of laughter in great apes and humans. *Current Biology, 19(13), 1106-1111*.

Schelling, T. C. (1980). *The Strategy of Conflict: with a new Preface by the Author*. Harvard university press.

Schwing, R., Nelson, X. J., Wein, A., & Parsons, S. (2017). Positive emotional contagion in a New Zealand parrot. *Current Biology*, *27*(6), R213-R214.

Scott-Phillips, T. (2026). Communication and grammar: A synthesis. *Psychological Review, 133*(1), 81–93.

Sell, A., Sznycer, D., Al-Shawaf, L., Lim, J., Krauss, A., Feldman, A., ... & Tooby, J. (2017). The grammar of anger: Mapping the computational architecture of a recalibrational emotion. *Cognition*, *168*, 110-128.

Simler, K., & Hanson, R. (2017). *The elephant in the brain: Hidden motives in everyday life*. Oxford University Press.

Smaldino, P. E., Flamson, T. J., & McElreath, R. (2018). The evolution of covert signaling. *Scientific reports*, *8*(1), 4905.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity, and mind-reading. *Mind & Language*, *17*(1-2), 3-23.

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., & Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences*, *113*(10), 2625-2630.

Sznycer, D., & Lukaszewski, A. W. (2019). The emotion–valuation constellation: Multiple emotions are governed by a common grammar of social valuation. *Evolution and Human Behavior*, *40*(4), 395-404.

Sznycer, D. (2022). Value computation in humans. *Evolution and Human Behavior*, *43*(5), 367-380.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in cognitive sciences*, *22*(3), 201-212.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, *107*(4), 657.

Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.

Tomasello, M. (2022). The coordination of attention and action in great apes and humans. *Philosophical Transactions of the Royal Society B*, *377*(1859), 20210093.

Quillien, T., & Lucas, C. G. (2024). Counterfactuals and the logic of causal selection. *Psychological Review*, *131*(5), 1208.

Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: some lessons from the past. *American psychologist*, *63*(3), 182.

Wiseman, R. (2002). *Laughlab: The Search for the World's Funniest Joke*. London: Randomhouse.

Yam, K. C., Barnes, C. M., Leavitt, K., Wei, W., Lau, J., & Uhlmann, E. L. (2019). Why so serious? A laboratory and field investigation of the link between morality and humor. *Journal of Personality and Social Psychology*, *117*(4), 758.