

**The Evolution of Social Paradoxes**

David Pinosof

david.pinosof@gmail.com

Department of Psychology, University of California, Los Angeles

**Abstract**

Human behavior is often paradoxical. We show humility to prove we're better than other people, bravely defy social norms so that people will praise us, and donate to charity anonymously to get credit for not caring about getting credit. Here, I argue that these and other social paradoxes have a common thread: they are all attempts to signal a trait while concealing, from both the signaler and the recipient, the fact that a signal is being transmitted. Such concealed signals emerge from the interaction of two cognitive abilities: cue-based inference and recursive mindreading. If agents can model each other's mental states, including their desire to signal their positive traits, then signals of positive traits can, themselves, become cues of negative traits. The result is that status seeking and virtue signaling are forced to occur covertly, without becoming common knowledge among the interactants. Status games can collapse when players gain common knowledge that they are playing a status game, licensing negative inferences about the players or unveiling deceptive signals. Social paradoxes may play a crucial role in enabling collective action in intergroup conflict, by socially incentivizing exploitative behavior that would otherwise elicit negative inferences. Social paradoxes can explain a variety of puzzling aspects of culture, including moralistic pretexts, volatile status symbols, and the function of sacred values.

*Keywords:* Signaling, Mindreading, Theory of Mind, Status, Values, Social Judgment

**Public significance statement**

We compete to be uncompetitive and seek status by not caring about status. Unraveling these and other social paradoxes offer powerful insights into human culture, explaining why status symbols vary across time and space, why deception can be mutually beneficial, and why communities orient themselves around sacred values and ideals.

### The Evolution of Social Paradoxes

Consider the following statements:

1. We try to gain status by not caring about status (Choi & Anderson, 2024; Feltovich et al., 2002; Hoffman et al., 2018).
2. We rebel against conformity in the exact same way as everyone else (Feltovich et al., 2002; Williams, 2022).
3. We show humility to prove we're better than other people (Costello et al., 2023; Hoffman et al., 2018).
4. We avoid being defensive to defend our reputations (Yandell, 1979).
5. We don't care what people think, and we want them to think that (Hoffman et al., 2018; Leary et al., 2003).
6. We donate to charity anonymously, to get credit for not caring about getting credit (De Freitas et al., 2019; Hoffman et al., 2018).
7. We bravely defy social norms so that people will praise us (Bellezza et al., 2014; Williams, 2022).
8. We avoid being manipulative to get people to do what we want them to do (Pinker et al., 2008; Powers & Altman, 2022).
9. We compete to reassure each other that it's not a competition (Barclay, 2013; Benenson, 2013).
10. We help those in need, regardless of self-interest, because being seen as the type of person who helps those in need, regardless of self-interest, is in our self-interest (De Freitas et al., 2019; Newman & Cain, 2014).
11. We consume anti-consumerism (Heath & Potter, 2004).

## SOCIAL PARADOXES

12. We make subversive art that only high-status people appreciate (Smaldino et al., 2018).
13. We make fun of ourselves for being uncool to prove we're cool (Gkorezis, & Bellou, 2016).
14. We help our friends without expecting anything in return, because we know they would do the same for us (Szynter et al., 2019).
15. We show everyone our true, authentic self—not who society wants us to be—because that is who society wants us to be (Bailey & Iyengar, 2023; Beer, 2020; Guenther et al., 2024; Potter, 2011).

These are all examples of *social paradoxes*: interpersonal signals designed to conceal, from both the signaler and the recipient, the fact that a signal is being transmitted. The virtue signaler does not believe they are virtue signaling, and neither does the recipient who awards them virtue (if they did, they would not award them virtue; Newman & Cain, 2014). The “brave” norm-violator does not believe they are seeking praise, and neither does the audience who praises them. The “authentic” person does not believe they are behaving how society wants them to behave, and neither do the members of society (Beer, 2020; Potter, 2011).

Once we see what the signalers and recipients fail to see—that genuine (and unflattering) signals are being transmitted—the paradoxes dissolve. The “subversive” artist is, in fact, catering to the tastes of artistic elites. The “rebellious” nonconformist is, in fact, conforming the norms of their particular subculture. The anonymous donor is, in fact, signaling her superior virtue to savvy observers who uncover her identity (Hoffman et al., 2018). These desirable effects are not accidental: they are strategically sought out by the artist, the rebel, and the donor—often unconsciously. What’s puzzling is *not* that people possess cognitive systems designed to produce these social benefits: such systems would have been favored by natural selection (Barclay, 2013;

## SOCIAL PARADOXES

Hoffman & Yoeli, 2022; Kurzban & Aktipis, 2007). Rather, what's puzzling is that the attempt to produce these benefits is not explicitly acknowledged by either the signaler or the recipient—and is often vehemently denied by both.

Social paradoxes raise a variety of interesting questions beyond their surreptitious and contradictory nature. Why do observers negatively judge people for possessing fundamental motives—e.g. seeking status—that all humans possess (Anderson et al., 2015; Storr, 2021)? Why do observers positively judge people for *appearing* to possess motives—e.g., violating social norms—that no humans possess (House et al., 2020)? How do signalers conceal the element of signalhood from both themselves and recipients, and why do they seek to do so?

In this paper, I attempt to answer these questions using the logic of evolutionary game theory and cultural evolution (e.g., Hoffman & Yoeli, 2022). I propose that mutually concealed signals emerge from cognitive systems for recursive mindreading and cue-based inference. When these two systems interact, unintended cues and intended signals become unstable, with the former sliding into the latter and vice versa, giving rise to social paradoxes. In the following sections, I review research on recursive mindreading, cue-based inference, and their interaction. Then, I use the logic of social paradoxes to shed light on the nature of status symbols, sacred values, and political ideologies.

### **Recursive Mindreading**

Mindreading (or mentalizing or Theory of Mind) is the ability to infer others' mental states from their behavior, or to anticipate their behavior from their mental states (Sperber & Wilson, 2002; Tomasello, 2022). Mental states include beliefs, desires, and intentions, and these states can be represented in the negative as well—as in *not* believing there's food in the fridge or

## SOCIAL PARADOXES

*not* intending any harm. Humans appear to be uniquely adept at mindreading compared to nonhuman primates, which enables complex forms of cooperation (Tomasello, 2022).

*Recursive* mindreading (or metarepresentation) is the ability to model mental states *about* other mental states—for example, a belief that Sara wants to get lunch, or a desire that she believe that my intentions are good (Sperber, 2000). Second-order intentionality refers to two levels of embedding (“Sam wants me to believe he’s smart”), third-order intentionality refers to three levels (“Sam wants me to believe that Sara wants to go to the restaurant”), fourth-order refers to four levels (“Sam wants me to believe that he doesn’t care if I believe he’s smart”), and so on. Most people can keep track of mental states at four levels of recursion, and roughly 20% of people can keep track of mental states at five levels of recursion (Krems, Dunbar, & Neuberg, 2016; Stiller & Dunbar, 2007).

### **Cue-Based Inference**

Cue-based inference is the ability to use behaviors to infer underlying traits. For instance, we reliably infer that a person who takes dangerous risks is formidable (Fessler et al., 2014), that a person who delays gratification is trustworthy (Fitouchi et al., 2021), and that a person who believes that fetuses are people is a Republican (Williams, 2021).

It is important to make a distinction between *cues* and *signals*. Signals are behaviors that evolved to, or are intended to, convey information to others. Cues are behaviors that did *not* evolve to, or are not intended to, convey information to others—but are nevertheless used by others to guide their decision making (Laidre & Johnston, 2013; Smith & Harper, 2003). The rustling sound a squirrel makes while moving through the grass is a cue used by predators. The chest-thumping of a gorilla is a signal designed to intimidate rivals.

## SOCIAL PARADOXES

Some cue-based inferences may be innate or especially easy to learn. For instance, humans may have evolved cue-based inference systems for detecting competence (Bor, 2017), alliances (Pietraszewski et al., 2015), and formidability (Fessler et al., 2014). Other cues may be learned or calibrated throughout development, as people detect correlations between behaviors and group membership (people who wear yarmulkes are Jewish), between behaviors and traits (people who recycle are conscientious), or between mental states and traits (people who want attention are self-centered).

### **The Interaction Between Recursive Mindreading and Cue-Based Inference**

When cue-based inference interacts with recursive mindreading, cues and signals become unstable. Cue-based inferences are themselves mental states, which means they can be anticipated, and manipulated, by recursively mindreading agents. The manipulation of others' cue-based inferences would count as a signal—an attempt to convey information. Yet what is being manipulated is, from the recipient's perspective, a cue.

For example, suppose I pick up on a valid, naturally emitted cue: people with disheveled hair are lazier than people with neatly combed hair. As soon as I begin using this cue, I might engage in a bit of mindreading and expect you to use it as well. Now I will want to neatly comb my hair before seeing you, to appear more diligent. What began as a cue became a signal.

Normally, the slide from cues to signals happens slowly, over evolutionary time, in a process known as ritualization or the “signaler precursor route” (Laidre & Johnstone, 2013; Smith & Harper, 2003). In the first step of the process, an animal might naturally display its canines, horns, or claws in preparation for an attack. In the second step, intruders evolve to use such displays as cues of impending aggression—and respond by fighting or fleeing. In the last

## SOCIAL PARADOXES

step, the displays evolve to become more conspicuous and stereotyped, so intruders are more likely to detect them. Over many generations, attack cues evolve into attack signals.

In humans, such transitions occur in real time, giving rise to what we might call *intentional signals*—or information transmitted because of its anticipated effects on recipients’ mental states (Scott-Phillips, 2014; Sperber, 2000; Sperber & Wilson, 2002). Recursive mindreading lies at the heart of these signals—and often, their acknowledgement by recipients. Without recursive mindreading, thoughts like the following would be unthinkable:

1. “If I apologize, then she will *believe* that I *intended* no harm; therefore, I ought to apologize” (second-order intentionality).
2. “He must have apologized because he *wanted* me to be *believe* that he *intended* no harm” (third-order intentionality).

Recursive mindreading thus enables apologies, requests, insults, admonitions, or any other intentional signal. Such signals require a metarepresentation—what the recipient knows about the signaler’s mind—and an attempt by the signaler to bring about, or modify, that metarepresentation. Moreover, the *receipt* of an intentional signal requires an additional level of recursion: if you send me a signal, I must infer that you *wanted* me to *believe* something about your *intentions* (or beliefs, desires, hopes, etc.).

Here is where cues and signals become unstable. As soon as there are programs for recursive mindreading and cue-based inference running inside the same mind, there is nothing to stop that mind from using the output of the former as an input to the latter. That is, one might use intentional signals *as* cues. For example, if I flamboyantly signal to you that I am morally righteous, then you might use my flamboyant signal as a cue of my underlying vanity or self-absorption. My behavior was both a signal *and* a cue.

## SOCIAL PARADOXES

But now the opposite transformation can occur. Once the flamboyant virtue signal becomes a cue of vanity or self-absorption, then that *very cue* can transform back into a signal, via the same process described above. That is, if I anticipate that my flamboyant virtue signal will make me look vain and self-absorbed, then I might conceal or “bury” my virtue signal to show off my humility and selflessness (Hoffman et al., 2018). The paradox is that my attempt to conceal my virtue signal would itself count as a virtue signal.

To take another example, suppose I discover a new cue: people who use “big words” (e.g., “reductionist,” “hegemony”) are smart. As soon I discover this cue, I will do a bit of mindreading and expect others to use it as well. Now I will want to pepper my speech with big words to look smart, transforming the cue into a signal. As such peppering becomes widespread, it might begin to make one look pretentious, transforming the signal into a cue (i.e., of pretentiousness). At that point, people might avoid or deride the “big words” to show they are not pretentious, transforming the cue into a signal (i.e., of being down-to-earth).

Humans can form generalizations about people, and so there is the possibility of signaling in general becoming a cue in general. That is, generally caring what others think, and contorting one’s behavior to win their approval, might become a cue of insecurity and low status. That cue might then become a signal, as people put on a performance of “not caring what others think” to (ironically) make others think they are self-assured and high-status. Observers may then see through the performance, causing it to revert back to a cue of insecurity and low status.

Similar chicanery may result when we focus on the recipients of these signals. Seeing through a performance (e.g., of “not caring what others think”) may be used by others as a cue of one’s savviness or perceptiveness. But as soon as this cue becomes widely used, it may transform into a signal, with people actively showing off their savviness by snidely calling out

## SOCIAL PARADOXES

performances that do not exist. If the signal slides into a cue—e.g., of cynicism or combativeness—it may then slide into a signal, as people show off their idealism and agreeableness by kindly playing along with others' performances (see also Altay et al., 2023).

**Arms Races and Symbiotic Deception**

Given this analysis of cues and signals, and the ease with which they slide into each other, it is important to get a handle on when they are valid or honest. When a cue correctly tracks a feature of reality, the cue is *valid*. When a person signals the possession of a trait that they truly possess, the signal is *honest*. Cues are invalid when they fail to track what they are cueing, and signals are deceptive when the signaler belies their signal.

For example, suppose that people who naturally use big words are smart, but people who pretentiously use big words to look smart are *not* smart. In this case, we would have a valid cue of intelligence that slides into a deceptive signal of intelligence. The deceptive signal would exploit the valid cue, analogous to Batesian mimicry in evolution (Mokkonen & Lindstedt, 2016). As the deceptive signal increases in frequency relative to the valid cue, an arms race will emerge between ever-more vigilant recipients and ever-more cunning deceivers, either over evolutionary time or over individual lifetimes (in the case of humans).

But valid cues can also transform into honest signals. For example, suppose people with disheveled hair are truly lazy, and people who comb their hair to avoid looking lazy are truly diligent. Here, the valid cue of laziness slides into an honest signal of diligence. But now suppose that neatly-combed hair becomes a cue of something else—say, nervousness. Maybe people who try to look diligent are nervous about making a good impression. If so, the honest signal of diligence would slide into a valid cue of nervousness. However, if a diligent, nervous wreck tries to avoid looking nervous by intentionally leaving his artfully tussled, then we would

## SOCIAL PARADOXES

have to add another link to the causal chain: 1) a valid cue of laziness that 2) slides into an honest signal of diligence that 3) slides into a valid cue of nervousness that 4) slides into a deceptive signal of nonchalance. Yes, social life can be very complicated.

At the risk of disorienting the reader, allow me to add one more link to the causal chain. Suppose people who are inclined to engage in all this recursive mindreading—who can anticipate what inferences others will draw about their hair—are socially competent. If so, the deceptive signal of nonchalance would contain a valid cue of social competence. The benefit of the valid cue could outweigh the cost of the deceptive signal, creating a kind of *symbiotic deception*. The signaler benefits from being favored by the recipient (e.g., for a job or relationship), while the recipient benefits from partnering with a socially competent person—one likely to make a good impression on others. If the benefit of the signaler's social competence outweighs the cost of his deceptive nonchalance, then both sides profit from the deception.

Another kind of symbiotic deception occurs when allowing oneself to be deceived elicits positive inferences among observers. As we saw previously, playing along with others' performances might be a valid cue of idealism or agreeableness. Believing a propagandistic narrative might be a valid cue of partisan loyalty (Pinsof et al., 2023). Believing a dubious accusation might be a valid cue of trust in the accuser. In each case, the valid cue could slide into an honest signal, as recipients gullibly accept implausible information to signal their idealism, loyalty, and trust. Insofar as the benefits of sending these signals outweigh the costs of accepting false information, the recipients profit from being deceived.

Under what conditions do we get social paradoxes? In two conditions: 1) where a valid cue slides into a deceptive signal (e.g., naturally using big words → needlessly using big words), and 2) where an honest signal slides into a valid cue of a negative trait (e.g., an honest signal of

## SOCIAL PARADOXES

diligence → a valid cue of nervousness). Under either of these conditions, agents benefit from concealing their signals, either to successfully dupe recipients with their deceptive signals, or to avoid eliciting negative inferences about their honest signals. If concealing a signal from oneself allows one to better conceal it from others (Smith et al., 2017), then we expect the signaler and the recipient to be unaware of the information being transmitted between them. As we will now see, signalers may also be unaware of the *status games* (Storr, 2021)—i.e., the rules for gaining and conferring status—they are playing.

**Not Caring about Status to Gain Status**

There is a contradiction at the heart of social status, defined here as widespread recognition of one's superiority as a social partner (e.g., Sznycer, 2017, 2018). On the one hand, there is abundant empirical evidence that status is a basic human need—one of our most powerful motives (Anderson et al., 2015; Storr, 2021; Sznycer et al., 2017; von Rueden, 2024). We endure costly trials and tribulations to win the esteem of our peers, from religious rites to yam-growing contests to death-defying risks (Storr, 2021). Across cultures, feelings of pride and shame are exquisitely sensitive to others' approval, accurately tracking the judgments of local audiences (Sznycer et al., 2017, 2018). Social rank is correlated with reproductive success among a wide variety of mammalian and primate species, including humans, suggesting that status-seeking is a deep part of our evolutionary heritage (Cowlshaw & Dunbar, 1991; Ellis, 1995; Huchard & Lukas, 2022; von Rueden & Jaeggi, 2016).

On the other hand, status-seekers are repulsive. We have a plethora of terms for such people: braggarts, blowhards, bigshots, bullies, posers, know-it-alls, holier-than-thous, name-droppers, narcissists, snobs, showoffs, sycophants, grandstanders, social climbers, egomaniacs, and petty tyrants. We occasionally view them as symbols of the male reproductive organ, the

## SOCIAL PARADOXES

female hygienic product, or the anus (and those who kiss it), implying a disgust response (Sharpe et al., 2022; Urban Dictionary, 2006, 2013). Research indicates that overt status-seekers are judged as selfish, disingenuous, and uncaring, causing them to lose the very status they seek (Choi & Anderson, 2024; Kim & Pettit, 2015). Norms against bragging and dominance are common around the world and may have been an important selection pressure for our species (Boehm, 2009, 2012; Brown, 1991; Sznycer et al., 2017; von Rueden, 2024). In sum, the contradiction is this: we must seek status without being a status-seeker. How do we do it?

The answer, I propose, is via social paradoxes—the concealing of status signals from recipients and signalers (the better to conceal them from recipients; Smith et al., 2017). Recall that intentional displays of virtue can backfire and damage one’s moral reputation—that is, virtue signals can slide into vice cues (Newman & Cain, 2014). Something similar can occur with showy displays of wealth, skill, knowledge, or connections to high-status people. Status signals can slide into valid cues of egotism, mendacity, or callousness—and thereby lower one’s status. Actions that make one look superior can rarely be taken for the sole purpose of looking superior (Choi & Anderson, 2024). Objects that confer status often do so implicitly and discreetly. Status games are rarely played “out in the open.” If they were, the prevailing status signals would slide into negative cues, inverting—or at least altering—the payoffs of the game.

The phrase “out in the open” is used metaphorically to describe a concept in game theory known as “common knowledge” (e.g. Thomas, DeScioli, Haque, & Pinker, 2014). For example, if I’m trying to coordinate with you for mutual benefit—say, meeting up with you in a park without communicating—then it is not enough for me to know that the fountain is the best place to meet up. I must also know that *you know* the fountain is the best place to meet up; otherwise, I could not expect you to be there. But interestingly, that is also not enough. I must also know that

## SOCIAL PARADOXES

you know *that I know* the fountain is the best place to meet up; otherwise, you could not expect *me* to be there. In order to successfully coordinate, we both must know that the other one knows that the other one knows (that the other one knows, and so on) to meet up at the fountain.

Common knowledge, in this sense, can erase a status signal. For example, if I privately know that Harvard is impressive, it might boost my status to tell people I went to Harvard, in case they find it impressive too. If I also know that *others know* Harvard is impressive, then I can be even more assured that revealing my alma mater will impress them. But if I know that others know *that I know* Harvard is impressive—then I might expect them to infer that the reason I namedropped Harvard was to impress them. After all, my knowledge of the university's impressiveness would, from their perspective, plausibly explain my behavior (what else would explain it?). After engaging in this sort of higher-order mindreading, I might ultimately refrain from bragging about my alma matter, to avoid looking pretentious. In this case, common knowledge of the status signal transformed it into a negative cue—and made me conceal it (or, perhaps, reveal it more slyly; Hoffman et al., 2018).

This raises an interesting question. What happens when players of a status game gain common knowledge that they are, in fact, playing a status game? How do members of a culture react when the proverbial boy declares that the emperor is naked? Just as the emperor seeks to cover his dignity, the community might seek to cover theirs. If Wall Street investors are accused of greed, selfishness, or conspicuous consumption, they might point out how investment is actually good for the economy and promotes the national welfare—prosperity trickles down. When a status game is exposed as a mere status game, its players can be expected to react with disbelief, skepticism, or outrage. They might argue they are not interested in petty things like status or resources, but nobler things like innovation or the national welfare.

## SOCIAL PARADOXES

But what if their arguments are unpersuasive, and everyone sees through them? Or what if their outrage is too weak, and nobody is deterred? At that point, common knowledge will set in, and the status game will *collapse*. Status-enhancing signals (e.g., working on Wall Street) will morph into cues of greed and stuffiness. Coveted objects (e.g., enormous yachts) will become seen as disgusting and selfish. Status-enhancing narratives (e.g., the power of free markets) will become the tired old “conventional wisdom”—a sport for mockery. Rebellion will become the new conformity. What once was cool will no longer be cool.

When a status game collapses, the players experience something like an inversion of the social hierarchy. The higher one’s rank, the more people one must have stepped on to acquire it. The more time and energy one invested in the competition, the more obsessed one must have been with winning it—and the more cutthroat and egotistical one must be. When people gain common knowledge that a status game *is* a status game, the social hierarchy becomes—if not entirely inverted—significantly rearranged. The winners look conniving and entitled; the losers look humble and modest. The top gets lowered; the bottom gets lifted.

In the aftermath of a collapsed (or rearranged) status game, the players often gain status by doing the opposite of what was done previously. If neatly-combed hair and crisp, black-and-white suits become cues of petty tyranny and snootiness, then long, messy hair and flowing, colorful outfits become cues of the opposite—rebelliousness and authenticity (Heath & Potter, 2004). Whenever members of a subculture get outed as puffed-up status-seekers, it creates an opportunity for everyone else to conspicuously differentiate themselves, transforming the negative cues (crisp, black-and-white suits → pretentiousness) into positive signals (flowing, colorful outfits → authenticity). Acting in defiance of a collapsed status game signals that one doesn’t care about status—which, paradoxically, raises one’s status.

## SOCIAL PARADOXES

This suggests an explanation for the seemingly endless variation in status symbols observed by anthropologists across cultures and throughout history. Status games are *volatile*: they collapse and re-emerge in antithetical forms. Baroque music and ornate architecture may give way to modernism and minimalism, as vainglorious complexity gives rise to humble simplicity (Hughes, 2024). Artists may create intricate paintings to show off their skill and years of practice, only to change course when the showy techniques are outed as a form of bragging. Fashions fluctuate as new styles become cool, mindlessly conforming to what's cool becomes uncool, and deftly figuring out what's cool or uncool becomes a valid cue (and resultant signal) of social competence. Long-enduring practices like dueling or powdered wigs may morph into symbols of fragile male egos or haughty aristocracies (Festa, 2005; Stevens, 1940, pp. 280-83, as quoted in Mueller, 1989, p. 10). Gilded Age cultures of conspicuous consumption and social Darwinism may get outed as greedy and heartless, giving rise to their opposites: inconspicuous consumption and social justice (Currid-Halkett, 2017; Eckhardt et al., 2015).

### **Variation in Status Game Stability**

Not all status games are equally volatile. There is a continuum between overtness and stability on the one hand, and slyness and volatility on the other. What explains the variation? We can move toward an answer by considering the nature of social paradoxes and the conditions that give rise to them. Recall that these conditions are: 1) when valid cues slide into deceptive signals, and 2) when honest signals elicit negative inferences about the signaler. In the context of a status game, we expect greater slyness and volatility when there is 1) greater opportunity for deception (Mokkonen & Lindstedt, 2016), and 2) the benefits of status signals are low relative to the costs of the negative inferences they elicit.

## SOCIAL PARADOXES

Regarding 1)—the opportunity for deception—we might expect greater volatility when the status-conferring traits are hard to infer or easy to mimic. Whatever cues observers use to infer humility, sensitivity, or innocence—such cues will tend to transform into deceptive signals as charismatic phonies find ways to pick the locks of others’ cue-based inferences. When such signals are spotted, they may slide into antithetical cues, as when showy displays of humility morph into cues of arrogance, or when panicky displays of innocence morph into cues of guilt (Yandell, 1979). Such spotting and morphing may lead to arms races between deceivers and deceived, and cultural variation in signals and cues. By contrast, traits like brute strength and physical attractiveness are easier to infer and harder to mimic. Though humans will go to extraordinary lengths to mimic them (e.g., performance-enhancing drugs, plastic surgery), status games surrounding these traits are more universal and less volatile.

But perfectly honest status signals can still collapse under the weight of mutual awareness. For example, if I publicly utter the words, “I’m better than you,” that could elicit a variety of negative inferences about my character—and feelings of threat—even if I truly *am* better than you. This brings us to 2), the negative inferences drawn about status signals, even when they are perfectly honest. Here, stability depends on two variables: signal benefits and inference costs. When the former is higher than the latter, we expect greater overtness and stability. When the latter is higher than the former, we expect greater slyness and volatility.

Consider a community living in poverty, where wealth is of primary importance in a social partner, and where the negative traits associated with status-seeking are a secondary concern. In such cultures, people will display their wealth overtly, because it is more important to get the message across (“I’m rich!”) than to look selfless and modest. Flashy displays of largesse may be more eye-popping than distasteful to poor audiences, and more distasteful than

## SOCIAL PARADOXES

eye-popping to rich audiences. Conspicuous consumption is thus expected to be a preoccupation of the “nouveau riche” (or newly rich), who may still have lower-class peers to impress (Galbraith, 1998, as cited in Currid-Halkett, 2017, p. 9). The old rich, by contrast, will prefer to engage in more inconspicuous consumption or “stealth wealth,” because the benefits of getting the message across will be outweighed by the ensuing inferences of haughtiness, insecurity, or “nouveau riche” status itself (Eckhardt et al., 2015; Trebay, 2023).

Another domain with low inference costs—and high signal benefits—is athletics. It doesn’t matter so much if LeBron James is an egomaniac; what matters is that he is good at basketball. Insofar as the costs of looking selfish and braggadocious are low, we expect more overt status-seeking and trash-talking to occur. Since athletes are not forming deep relationships or monogamous pairbonds with their opponents, there is no need to scrutinize each other’s character. Then again, inference costs may be higher between teammates, where such trash-talking may foster resentment or thwart opportunities for leadership. The inference costs may be higher outside the court or the stadium, or in interviews with the press, where the machismo might give way to idealism about the value of teamwork, giving 110%, or the love of the game.

On the other end of the continuum, inference costs may be especially high in female intrasexual competition, where the goal of the competition is often to present oneself as less competitive than one’s competitors (Benenson, 2013). Compared to men, women tend to have a smaller number of more intimate friends, built upon a sense of mutual trust, sympathy, and egalitarianism (Williams et al., 2022). This implies that the costs of being seen as a selfish, competitive backstabber will be particularly high in female social networks, outweighing the benefits of any overt status-seeking tactics (e.g., bragging, dissing rivals, etc.). Of course, women still need to compete for status, as all social animals do, but the competition is often forced to

## SOCIAL PARADOXES

occur covertly, without becoming common knowledge among the competitors. Instead of overtly derogating rivals, for example, women may frame their derogation as “venting” (to release pent-up emotions) or “expressing concern” (to disguise their insults as efforts to help)—tactics to ensure plausible deniability of the derogation (Krems et al., 2024; Reynolds et al., 2025).

### **Sacred Values as Plausibly Deniable Status Seeking**

Rather than overtly pursuing status, humans often frame their activities as the pursuit of a *sacred value* or *ideal* (e.g., equality, diversity, honor, temperance, holiness, the love of the game, giving 110%, innovation, the national welfare, etc.). Sacred values can be differentiated from other values by being apparently disconnected from rational self-interest (Atran & Axelrod, 2008; Ruttan & Nordgren, 2021). Many scholars have suggested that sacred values function to promote group cohesion (Atran & Axelrod, 2008; Haidt, 2012). I do not dispute that sacred values have this effect, but I will argue that their primary function is something different: to stabilize a status game by disguising it as the pursuit of a non-status-related end.

This makes sacred values deceptive, albeit of the symbiotic variety discussed previously. The winners of the status game, together with their allies, benefit from the deception by preventing their status game from collapsing and inverting. Aspiring winners benefit from ensuring their chosen status game remains stable into the future. Professed devotion to the sacred value may thus become a valid cue, and resultant signal, of commitment to a culture and loyalty to its higher-ups. As we saw previously, it may also become a valid cue, and resultant signal, of idealism, agreeableness, and trust. The cost of the false belief may be outweighed by what it reveals about the believer, including the believer’s status, loyalty, idealism, and commitment.

But the sacred value may not benefit everyone. For those at the bottom of the social ladder, who have little hope of rising up, the sacred value may be exploitative, entrenching a

## SOCIAL PARADOXES

status game they are destined to lose. It may behoove them to expose the hollowness of the sacred ideal and trigger an inversion of the hierarchy—to bring others down and lift themselves up. Unfortunately for these low-status people, their attempts to expose the status game may become a valid cue of low status itself (as well as disloyalty, alienation, cynicism, and combativeness). Sacred values can thus be self-reinforcing, with any attempt to challenge them becoming a negative cue, and any attempt to affirm them becoming a positive signal.

## SOCIAL PARADOXES

Sacred values should, if well-designed, be as conceptually distant from status as possible, because any confusion between the two could cause the underlying status game to collapse and invert. The urge to jockey for relative standing should be reframed as the opposite: a yearning for spiritual purity, self-abnegation, or moral progress. Competitive urination contests should be reimagined as exalted and selfless journeys. Dominating subordinates should be reconceptualized as shepherding followers toward wisdom or absolution. Dissing rivals and upstarts should be reinterpreted as defending the sacred value from corruption or perversion.

At the same time, the sacred value should awkwardly track real status acquisition and conferral. Everywhere the sacred value appears, the competition for superiority should follow closely behind. Pursuit of the sacred ideal should, beneath appearances, be indistinguishable from the pursuit of social rewards. If authenticity is a sacred ideal, then the pursuit of authenticity will (ironically) track the pursuit of others' approval (Bailey & Iyengar, 2023; Beer, 2020; Potter, 2011). If humility is a sacred ideal, then displaying humility will (ironically) function to make one look superior to others (Costello et al., 2023). The same logic applies to other sacred values, like bravery, subversiveness, non-conformity, and truth-telling. "Brave" norm-violators should be praised by their peers, "subversive" artists should cater to the tastes of the intelligentsia, "unpopular opinions" should be met with enthusiastic agreement, and spin doctors should gain followers for having the "courage" to "tell it like it is."

In some cases, status itself may be given a new label and reconceptualized as something internal to one's essence or soul, as opposed to something dependent on others' approval. One example of this re-labelling may be the concept of "self-actualization." Research indicates that the pursuit of "self-actualization" is conceptually interchangeable with the pursuit of status, with the former functioning as a kind of spiritual reframing of the latter (Krems et al., 2017). When

## SOCIAL PARADOXES

asked to describe what self-actualization meant to them, participants' responses most commonly referred to having high status, e.g., "Making seven figures," "getting a 4.0 GPA," "working at Wall Street and making tons of money," and "writing the great American novel" (Krems et al., 2017, p. 6). Self-actualization may be a secular sacred value.

Another way to stabilize a status game is to make the sacred value resemble—or mimic—the actual values people hold. For example, we value love (for family, mates, and close friends), altruism (toward our allies and cooperative partners), and knowledge (for the immediate, the concrete, and the actionable). Sacred values like universal love, helping humankind, and the quest for worldly knowledge thus have an element of plausibility to them: they are similar to real, evolved values we actually have. The sacred versions of these values are different from the real ones, of course. Our actual love, altruism, and knowledge-seeking are more circumscribed and contextualized than their sacred counterparts. But because sacred values *have* real-life counterparts, they appear more plausible to communities of status-seekers—and more effective at stabilizing their status games. If deceptive signals of sacred values contain valid cues of real values, the deception can symbiotically persist.

### **Establishing Dominance to Promote Equality**

There is a contradiction at the heart of *tribalism*, defined here as the pursuit of ingroup benefits at outgroups' expense. On the one hand, we are a tribal species. Like our chimpanzee cousins, we compete over territory and attack rival groups when we have a numerical advantage (Wrangham & Glowaki, 2012). We form intergroup dominance hierarchies and privilege our ingroup over our outgroup (Tooby & Cosmides, 2010). We spread propaganda and scurrilous rumors to mobilize outgroup aggression (Horowitz, 2001; Petersen et al., 2020). Alliances and

## SOCIAL PARADOXES

coalitions are ubiquitous among primates and social animals, suggesting that intergroup competition is a deep part of our evolutionary heritage (Harcourt & De Waal, 1992).

On the other hand, the word “tribalism” is pejorative. Propaganda fails when it is outed as propagandistic (Biddle, 1931; Mercier, 2020), and dominance often backfires, fomenting rebellion against the would-be dominator (Boehm, 2009; Powers & Altman, 2022). Murder and violence are universally proscribed (Brown, 1991), and attacks against outgroups can cause feeling of guilt, depression, or “moral injury” among the attackers (Zefferman & Matthew, 2020). When one group expropriates another’s resources, or privileges themselves at others’ expense, our moral psychology is likely to flag it as unjust, casting the privileged as exploitative and undeserving (Baumard et al., 2013; Boehm, 2009). In sum, the contradiction is this: we must prevail in intergroup competition without being tribal. How do we do it?

The answer, I propose, is via social paradoxes. Rather than pursuing tribalism “out in the open,” we conceal it with sacred ideologies. We rationalize intergroup dominance hierarchies as karmic orders, divine rights, or great chains of being. We disguise group-serving propaganda as the unbiased truth and group-serving agendas as altruistic or spiritual imperatives. Just as sacred values stabilize interpersonal status games, sacred ideologies stabilize intergroup dominance games. And just as sacred myths are symbiotically deceptive, so too are sacred ideologies, enabling groups to coordinate against rivals in intergroup conflict—and signal commitment to their group. Without a moral pretext, physically or verbally attacking the outgroup could elicit negative inferences about the attackers, socially disincentivizing the attack.

But would such negative inferences be valid? The answer appears to be yes. Motivations to insult political outgroups are associated with psychopathy (Mackin & Roesse, 2024). People who engage in hostile political rhetoric online are also hostile offline—and are more likely to

## SOCIAL PARADOXES

have a criminal record (Bor & Petersen, 2022; Rasmussen et al., 2024). Support for political violence is associated with “callous, manipulative” personality traits (Gøtzsche-Astrup, 2021, p. 1; Nai & Young, 2024). Motivation to argue about politics is associated with status-seeking and “dominance-oriented moral grandstanding motivations” (Grubbs et al., 2022; Grubbs et al., 2019). Endorsement of conspiracy theories—which may function to mobilize the group (Marie & Petersen, 2022)—is associated with criminal behavior (Jolley et al., 2019). Even milder investments in intergroup competition—political interest and participation—are associated with narcissism and psychopathy (Chen et al., 2021; Rogoza et al., 2022). If people began using these cues in contexts of intergroup conflict, or suspected others to use them (or suspected others to suspect others...), the social incentives for collective action would collapse and invert.

The challenge for any coalition, therefore, is to cloak intergroup dominance in the garb of altruism and ethics—a kind of Orwellian doublethink. Groups must dominate rivals while posing as egalitarian revolutionaries (Storr, 2021, chapter 28) and quash dissent while masquerading as courageous truth-seekers (Kubin et al., 2024). Territorial conquest must be draped in virtue and ornately justified—e.g., as a denazification campaign (Wollast et al., 2024) or a quest to manifest “the excellence of divine principles” (Pratt, 1927). Wherever outgroup violence occurs, it is usually preceded by moralistic or religious pretexts—e.g., retaliation against outrages (Horowitz, 2001), possession by vengeful spirits (Hayden, 2018), the cracking of eggs for a utopian omelet (Mercier, 2020; Petersen et al., 2020), or the spilling of blood to appease the gods (Watts et al., 2016). By insulating groups from common knowledge of their motives, social paradoxes may enable and sustain the most unethical forms of behavior known to our species.

**Discussion**

## SOCIAL PARADOXES

I have argued that social paradoxes arise from two adaptations: recursive mindreading and cue-based inference. When these two adaptations interact, cues become unstable: as people recognize them as cues, they intentionally imitate them (or conceal them) to gain social rewards, thereby transforming them into signals. The volatile transformation from cues to signals, from signals to buried signals (Hoffman et al., 2018), and from signals back into cues (and perhaps back again), may be one of the fundamental engines of cultural variation. To protect our cultures from this volatility, we create and defend sacred narratives.

Social paradoxes imply a danger for social science itself. If humans are constantly seeking to conceal their motives and signaling strategies from one another, and even from themselves, then any science that relies on self-reported agreement with overly abstract statements is bound to be misleading. Political psychologists relying on such data might be misled into believing that politics revolves around sacred values, like equality, authority, or tolerance of outgroups (Pinsof et al., 2023). Social psychologists relying on such data might be misled into believing that humans pursue evolutionarily implausible goals like authenticity, self-affirmation, or self-actualization (Beer, 2020; Krems et al., 2017; Kurzban & Aktipis, 2007). Cultural psychologists might be misled into believing that cultures are more different than they really are, orienting themselves around the pursuit of different abstract ideals. Without knowledge of the paradoxical nature of the human psyche, social scientists seeking to understand human nature would be led astray. The mistake would be akin to developing an economic theory of corporations by relying on mission statements, ad copy, and PR as data.

It is worth reflecting on the fact that the practice we are currently engaged in—the scientific method—is itself undergirded by the sacred values of truth, knowledge, and discovery. To suggest that we are all just trying to show off our impressive intellects, jockey for status, play

## SOCIAL PARADOXES

academic politics, twist ambiguous data into glib advice or partisan cheerleading, and compete for an ever-dwindling number of cushy jobs, is to insult us in a very deep way. Such thoughts are, if not taboo to think, uncomfortable to think. That we are all striving to glimpse the subtle contours of reality, to push the frontiers of knowledge, to make the world a better place—these thoughts are the opposite of taboo: they are ennobling and inspiring. We need these thoughts to strengthen the institution of science and attract the energy of young scientists. Without such sacred ideals, this paper could not exist.

Another implication of social paradoxes is that nihilism, or skepticism of our sacred values, should be deeply threatening to humans. This very paper, by shining a light on what we continually strive to cover up, is predicted to be discomfiting to many of its readers. Nihilism threatens to unravel the status games and collective projects that add value to our lives, potentially leaving us adrift and disoriented. It is no wonder that nihilistic ideas are a weapon used by embittered or low-status people to attack the prevailing social order and bring people down to their level (Arceneaux et al., 2021). Regardless of whether nihilism is true as a philosophical matter, we have evolved to (in many contexts) fervently reject it—and shun anyone who might be tempted by its logic. It is this underlying tension between the looming threat of nihilism and the passionate intensity of our sacred ideals that underscores much of what is interesting, beautiful, and paradoxical about the human condition.

## REFERENCES

- Altay, S., Majima, Y., & Mercier, H. (2023). Happy thoughts: The role of communion in accepting and sharing (mis) beliefs. *British Journal of Social Psychology*, 62(4), 1672-1692.
- Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental

## SOCIAL PARADOXES

- human motive? A Review of the Empirical Literature. *Psychological Bulletin*, 141(3), 574.
- Arceneaux, K., Gravelle, T. B., Osmundsen, M., Petersen, M. B., Reifler, J., & Scotto, T. J. (2021). Some people just want to watch the world burn: The prevalence, psychology and politics of the 'Need for Chaos'. *Philosophical Transactions of the Royal Society B*, 376(1822), 20200147.
- Atran, S., & Axelrod, R. (2008). Reframing sacred values. *Negotiation Journal*, 24(3), 221-246.
- Bailey, E. R., & Iyengar, S. S. (2023). Positive—More than unbiased—Self-perceptions increase subjective authenticity. *Journal of Personality and Social Psychology*, 125(6), 1351–1372. <https://doi.org/10.1037/pspa0000359>
- Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior*, 34(3), 164-175.
- Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36(1), 59-78.
- Beer, J. (2020). The inconvenient truth about your “authentic” self. *Scientific American*. Retrieved from <https://blogs.scientificamerican.com/observations/the-inconvenient-truth-about-your-authentic-self/>
- Bellezza, S., Gino, F., & Keinan, A. (2014). The red sneakers effect: Inferring status and competence from signals of nonconformity. *Journal of Consumer Research*, 41(1), 35-54.
- Benenson, J. F. (2013). The development of human female competition: Allies and adversaries. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1631), 20130079.
- Biddle, W. W. (1931). A psychological definition of propaganda. *The Journal of Abnormal and*

## SOCIAL PARADOXES

- Social Psychology*, 26(3), 283.
- Boehm, C. (2009). *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press.
- Boehm, C. (2012). *Moral origins: The evolution of virtue, altruism, and shame*. Soft Skull Press.
- Bor, A. (2017). Spontaneous categorization along competence in partner and leader evaluations. *Evolution and Human Behavior*, 38(4), 468-473.
- Bor, A., & Petersen, M. B. (2022). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American political science review*, 116(1), 1-18.
- Brown, D. E. (1991). *Human universals*. McGraw-Hill.
- Choi, A. L., & Anderson, C. (2024). Be careful what you wish for: Individuals perceived to desire status are afforded less status. *PloS one*, 19(6), e0304727.
- Costello, T. H., Newton, C., Lin, H., & Pennycook, G. (2023). Intellectual humility questionnaires fail to predict metacognitive skill: Implications for theory and measurement. <https://osf.io/preprints/psyarxiv/gux95>
- Chen, P., Scott P., and Blais, J. (2021). The dark side of politics: Participation and the dark triad. *Political Studies* 69(3): 577-601.
- Cowlshaw, G., & Dunbar, R. I. (1991). Dominance rank and mating success in male primates. *Animal Behaviour*, 41(6), 1045-1056.
- Currid-Halkett, E. (2017). *The sum of small things: A theory of the aspirational class*. Princeton University Press.
- De Freitas, J., DeScioli, P., Thomas, K. A., & Pinker, S. (2019). Maimonides' ladder: States of mutual knowledge and the perception of charitability. *Journal of Experimental*

## SOCIAL PARADOXES

*Psychology: General*, 148(1), 158.

Dunbar, R. I. (2003). The social brain: mind, language, and society in evolutionary perspective.

*Annual Review of Anthropology*, 163-181.

Eckhardt, G. M., Belk, R. W., & Wilson, J. A. (2015). The rise of inconspicuous

consumption. *Journal of Marketing Management*, 31(7-8), 807-826.

Ellis, L. (1995). Dominance and reproductive success among nonhuman animals: A cross-

species comparison. *Ethology and sociobiology*, 16(4), 257-333.

Feltovich, N., Harbaugh, R., & To, T. (2002). Too cool for school? Signaling and

countersignaling. *RAND Journal of Economics*, 33, 630-649.

<https://doi.org/10.2307/3087478>

Fessler, D. M., Tiokhin, L. B., Holbrook, C., Gervais, M. M., & Snyder, J. K. (2014).

Foundations of the crazy bastard hypothesis: Nonviolent physical risk-taking enhances conceptualized formidability. *Evolution and Human Behavior*, 35(1), 26-33.

Fitouchi, L., André, J. B., & Baumard, N. (2021). Moral disciplining: The cognitive and

evolutionary foundations of puritanical morality. *Behavioral and Brain Sciences*, 1-71.

Galbraith, J. K. (1998). *The affluent society*. Houghton Mifflin Harcourt.

Gkorezis, P., & Bellou, V. (2016). The relationship between leader self-deprecating humor and

perceived effectiveness: Trust in leader as a mediator. *Leadership & Organization Development Journal*, 37(7), 882-898.

Gøtzsche-Astrup, O. (2021). Dark triad, partisanship and violent intentions in the United

States. *Personality and Individual Differences*, 173, 110633.

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral

grandstanding in public discourse: Status-seeking motives as a potential explanatory

## SOCIAL PARADOXES

- mechanism in predicting conflict. *PloS one*, *14*(10), e0223749.
- Grubbs, J. B., James, A. S., Warmke, B., & Tosi, J. (2022). Moral grandstanding, narcissism, and self-reported responses to the COVID-19 crisis. *Journal of Research in Personality*, *97*, 104187.
- Guenther, C. L., Zhang, Y., & Sedikides, C. (2024). The authentic self is the self-enhancing self: A self-enhancement framework of authenticity. *Personality and Social Psychology Bulletin*, *50*(8), 1182-1196.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon.
- Hayden, B. (2018). *The power of ritual in prehistory: Secret societies and origins of social complexity*. Cambridge University Press.
- Heath, J., Potter, A. (2004). *The rebel sell: How the counterculture became consumer culture*. New York: Harper Business.
- Harcourt, A. H., & De Waal, F. (Eds.). (1992). *Coalitions and alliances in humans and other animals* (pp. 445-471). Oxford: Oxford University Press.
- Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences*, *112*(6), 1727-1732.
- Hoffman, M., & Yoeli, E. (2022). *Hidden games: The surprising power of game theory to explain irrational human behavior*. Hachette UK.
- Hoffman, M., Hilbe, C., & Nowak, M. A. (2018). The signal-burying game can explain why we obscure positive traits and good deeds. *Nature Human Behaviour*, *2*(6), 397-404.
- Horowitz, D. L. (2001). *The Deadly Ethnic Riot*. University of California Press.

## SOCIAL PARADOXES

House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebiolglu, S., Crittenden, A. N., ... &

Silk, J. B. (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour*, 4(1), 36-44.

Huchard, E., & Lukas, D. (2022). The effect of dominance rank on female reproductive success in social mammals. *Peer Community Journal*, 2.

Hughes, S. (2024). The beauty of concrete. *Works in Progress*.

<https://worksinprogress.co/issue/the-beauty-of-concrete/>

Jolley, D., Douglas, K. M., Leite, A. C., & Schrader, T. (2019). Belief in conspiracy theories and intentions to engage in everyday crime. *British Journal of Social Psychology*, 58(3), 534-549.

Kim, H. Y., & Pettit, N. C. (2015). Status is a four-letter word: Self versus other differences and concealment of status-striving. *Social Psychological and Personality Science*, 6(3), 267-275.

Krems, J. A., Kenrick, D. T., & Neel, R. (2017). Individual perceptions of self-actualization: What functional motives are linked to fulfilling one's full potential?. *Personality and Social Psychology Bulletin*, 43(9), 1337-1352.

Krems, J. A., Dunbar, R. I., & Neuberg, S. L. (2016). Something to talk about: are conversation sizes constrained by mental modeling abilities?. *Evolution and Human Behavior*, 37(6), 423-428.

Krems, J. A., Merrie, L. A., Rodriguez, N. N., & Williams, K. E. (2024). Venting makes people prefer—and preferentially support—us over those we vent about. *Evolution and Human Behavior*, 45(5), 106608.

Kubin, E., von Sikorski, C., & Gray, K. (2024). Political censorship feels acceptable when ideas

## SOCIAL PARADOXES

- seem harmful and false. *Political Psychology*, 00, 1-21.
- Kurzban, R., & Aktipis, C. (2007). Modularity and the social mind: Are psychologists too selfish?. *Personality and Social Psychology Review*, 11(2), 131-149.
- Laidre, M. E., & Johnstone, R. A. (2013). Animal signals. *Current Biology*, 23(18), R829-R833.
- Leary, M. R., Gallagher, B., Fors, E., Buttermore, N., Baldwin, E., Kennedy, K., & Mills, A. (2003). The invalidity of disclaimers about the effects of social feedback on self-esteem. *Personality and Social Psychology Bulletin*, 29(5), 623-636.
- Mackin, M., & Roese, N. J. (2024). The Dark Triad predicts public display of offensive political products. *Journal of Research in Personality*, 112, 104516.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, 50(4), 370.
- Marie, A., & Petersen, M. B. (2022). Political conspiracy theories as tools for mobilization and signaling. *Current Opinion in Psychology*, 48, 101440.
- Mercier, H. (2020). *Not born yesterday*. Princeton University Press.
- Mokkonen, M., & Lindstedt, C. (2016). The evolutionary ecology of deception. *Biological Reviews*, 91(4), 1020-1035.
- Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological Science*, 25(3), 648-655.
- Nai, A., & Young, E. L. (2024). They choose violence. Dark personality traits drive support for politically motivated violence in five democracies. *Personality and Individual Differences*, 230, 112794.
- Petersen, M. B. (2020). The evolutionary psychology of mass mobilization: How disinformation and demagogues coordinate rather than manipulate. *Current Opinion in Psychology*, 35, 71-75.

## SOCIAL PARADOXES

- Petersen, M. B., Osmundsen, M., & Tooby, J. (2020). The evolutionary psychology of conflict and the functions of falsehood. *The politics of truth in polarized America*, 131-150.
- Pietraszewski, D., Curry, O. S., Petersen, M. B., Cosmides, L., & Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition*, 140, 24-39.
- Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of sciences*, 105(3), 833-838.
- Pinsof, D., Sears, D. O., & Haselton, M. G. (2023). Strange bedfellows: The alliance theory of political belief systems. *Psychological Inquiry*, 34(3), 139-160.
- Potter, A. (2011). *The authenticity hoax: How we get lost finding ourselves*. McClelland & Stewart.
- Powers, K. E., & Altman, D. (2022). The psychology of coercion failure: How reactance explains resistance to threats. *American Journal of Political Science*. 67(1), 221-238.
- Pratt, J. W. (1927). The origin of "manifest destiny." *The American Historical Review*, 32(4), 795-798.
- Rasmussen, S. H. R., Bor, A., & Petersen, M. B. (2024). The offline roots of online hostility: Adult and childhood administrative records correlate with individual-level hostility on Twitter. *Proceedings of the National Academy of Sciences*, 121(44), e2412277121.
- Reynolds, T. A., Maner, J. K., & Baumeister, R. F. (2025). Bless her heart: Gossip phrased with concern provides advantages in female intrasexual competition. *Journal of Experimental Social Psychology*, 116, 104670.
- Rogoza, M., Marchlewska, M., & Szczepańska, D. (2022). Why dark personalities participate in politics? *Personality and Individual Differences*, 186, 111319.

## SOCIAL PARADOXES

- Ruttan, R. L., & Nordgren, L. F. (2021). Instrumental use erodes sacred values. *Journal of Personality and Social Psychology*, 121(6), 1223.
- Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Bloomsbury Publishing.
- Sharpe, B. M., Hyatt, C. S., Lynam, D. R., & Miller, J. D. (2022). “They are such an asshole”: describing the targets of a common insult among english-speakers in the United States. *Collabra: Psychology*, 8(1), 32552.
- Smaldino, P. E., Flamson, T. J., & McElreath, R. (2018). The evolution of covert signaling. *Scientific Reports*, 8(1), 1-10.
- Smith, J. M., & Harper, D. (2003). *Animal signals*. Oxford University Press.
- Smith, M. K., Trivers, R., & Von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology*, 63, 93-101.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity, and mind-reading. *Mind & Language*, 17(1-2), 3-23.
- Sperber, D. (2000). Metarepresentations in an Evolutionary Perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 117-137). OUP.
- Sznycer, D., Al-Shawaf, L., Bereby-Meyer, Y., Curry, O. S., De Smet, D., Ermer, E., ... & Tooby, J. (2017). Cross-cultural regularities in the cognitive architecture of pride. *Proceedings of the National Academy of Sciences*, 114(8), 1874-1879.
- Sznycer, D., Xygalatas, D., Agey, E., Alami, S., An, X. F., Ananyeva, K. I., ... & Tooby, J. (2018). Cross-cultural invariances in the architecture of shame. *Proceedings of the National Academy of Sciences*, 115(39), 9702-9707.
- Sznycer, D., Delton, A. W., Robertson, T. E., Cosmides, L., & Tooby, J. (2019). The ecological

## SOCIAL PARADOXES

- rationality of helping others: Potential helpers integrate cues of recipients' need and willingness to sacrifice. *Evolution and Human Behavior*, 40(1), 34-45.
- Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, 29(1), 93-104.
- Storr, W. (2021). *The status game: On social position and how we use it*. William Collins.
- 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.
- Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. *Human morality and sociality: Evolutionary and comparative perspectives*, 191-234.
- Trebay, G. (2023). 'Quiet luxury' becomes a flex for the ultra-rich. *The New York Times*.  
<https://www.nytimes.com/2023/07/22/style/quiet-luxury-wealth-status.html>
- Urban Dictionary. (2006). Dick move. Retrieved from  
<https://www.urbandictionary.com/define.php?term=dick+move>
- Urban Dictionary. (2013). Douche. Retrieved from  
<https://www.urbandictionary.com/define.php?term=Douche>
- von Rueden, C. R., & Jaeggi, A. V. (2016). Men's status and reproductive success in 33 nonindustrial societies: Effects of subsistence, marriage system, and reproductive strategy. *Proceedings of the National Academy of Sciences*, 113(39), 10824-10829.
- von Rueden, C. R. (2024). Status. In J. Koster, B.A. Scelza & M.K. Shenk (Eds), *Human behavioral ecology* (pp. 153-179). Cambridge University Press.
- Watts, J., Sheehan, O., Atkinson, Q. D., Bulbulia, J., & Gray, R. D. (2016). Ritual human sacrifice promoted and sustained the evolution of stratified societies. *Nature*, 532(7598),

Wrangham, R. W., & Glowacki, L. (2012). Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: Evaluating the chimpanzee model. *Human Nature*, 23, 5-29.

Williams, D. (2021). Socially adaptive belief. *Mind & Language*, 36(3), 333-354.

Williams, D. (2022). Signaling, commitment, and strategic absurdities. *Mind & Language*, 37(5), 1011-1029.

Williams, K. E., Krems, J. A., Ayers, J. D., & Rankin, A. M. (2022). Sex differences in friendship preferences. *Evolution and Human Behavior*, 43(1), 44-52.

Wollast, R., Phillips, J. B., Yahiaiev, I., Malysheva, K., Klein, O., & Sengupta, N. K. (2024). Russians' attitudes towards the war in Ukraine. *European Journal of Social Psychology*.

Yandell, B. (1979). Those who protest too much are seen as guilty. *Personality and Social Psychology Bulletin*, 5(1), 44-47.

Zefferman, M. R., & Mathew, S. (2020). An evolutionary theory of moral injury with insight from Turkana warriors. *Evolution and Human Behavior*, 41(5), 341-353.

