False Beliefs and the Tethered Mind

Vinod Goel
York University, Toronto, Canada
vgoel@yorku.ca

With the advent of social media technologies--which allow for the wide dissemination and entrenchment of beliefs by the elimination of gatekeepers (editors/publishers), financial obstacles, and instant feedback and reinforcement--the issue of the propagation of false beliefs has risen to the forefront of social behaviors. Given that we are the "rational animal" why are certain false beliefs so pervasive and difficult to revise in the face of counterevidence? The most popular answer in the literature is that we are motivated reasoners. Whereas ideal reasoning involves going from data/evidence to conclusions in a disinterested manner, motivated reasoning is guided by a preexisting goal or desire that serves to filter the data in order to support the preferred conclusion. The motivated reasoning explanation does work in a number of cases, particularly in the scientific and legal domains. However, I will argue that in the most interesting cases such as gender assignment and political partisanship, it begs the question. I will offer an explanation in the form of the tethered mind hypothesis. The tethered mind hypothesis acknowledges that while we are the "reasoning animal," our system of reasoning is built on top of and integrated into (or tethered to) evolutionary older systems such as the autonomic system, the instinctive system, and the associative system. Human behavior is a blended response of all these various systems. Using this model, I will propose that false belief adherence in the face of overwhelming evidence/data can be better explained by postulating the modulation of reasoning by non-cognitive systems such as the in-group/out-group instinct.

Draft 1.0

26th SYDNEY SYMPOSIUM OF SOCIAL PSYCHOLOGY, VISEGRAD, HUNGARY, JULY 8-12, 2024

Do not be unequally yoked with unbelievers. For what partnership has righteousness with lawlessness? Or what fellowship has light with darkness?

-- Bible, 2 Corinthians 6:14

1.0 Introduction

All organisms need some way of sensing and responding to changes in the external environment. A simple organism's behavior might be a function of the *presentation* of the environment. It can only respond to stimuli while it is present. If it is not present, however momentarily, there is no awareness or response. There are more neurally complex animals whose behavior is a function of a representation of the environment. Imagine a cat visually tracking a mouse across the living room floor. The mouse disappears behind the sofa, but the cat's gaze continues to where it emerges on the other side of the sofa. In this case we are justified in saying that the cat has a representation of the mouse. In case of both presentation and representation, the organism must have a largely veridical registration of the environment, most of the time, for it to survive and thrive. This is largely the job of sensory perceptual systems.

In the case of humans, we have evolved mental representations with propositional contents. Our mental representations can be analyzed as two distinct components, the psychological state and the content, as follows:

Psychological_State (propositional content)

Beliefs are one type of psychological state. They can be true or false. They are characterized by a *mind to world* direction of fit (Searle, 1983), meaning that if a belief is false, it must change, not the facts in the world. The availability of propositional content along with the related machinery of inference or reasoning provides for another source of information about the environment. Inferences can be from direct perceptual inputs and/or existing propositional representations. In the case where beliefs are formed by direct perception there is considerable sensory machinery devoted to getting this largely right, most of the time. In the case of beliefs arrived at with the machinery of inference or reason, consistency, or more generally coherence,

will facilitate veridicality (assuming veridicality of perceptions and existing beliefs). Inconsistency will not.

The system is set up such that as new information comes in, it is vetted and justified. If it is determined to be veridical, it is integrated into our worldview (and referred to as "knowledge"); otherwise, it should be discarded to maintain mind-to-world correspondence. Integration may require confronting inconsistencies and revising existing belief networks. This is captured in the ideal belief revision model in Figure 1a. However, it is well known that existing beliefs will modulate/affect the filtering of new information coming into the system leading to more realistic model in Figure 1b.

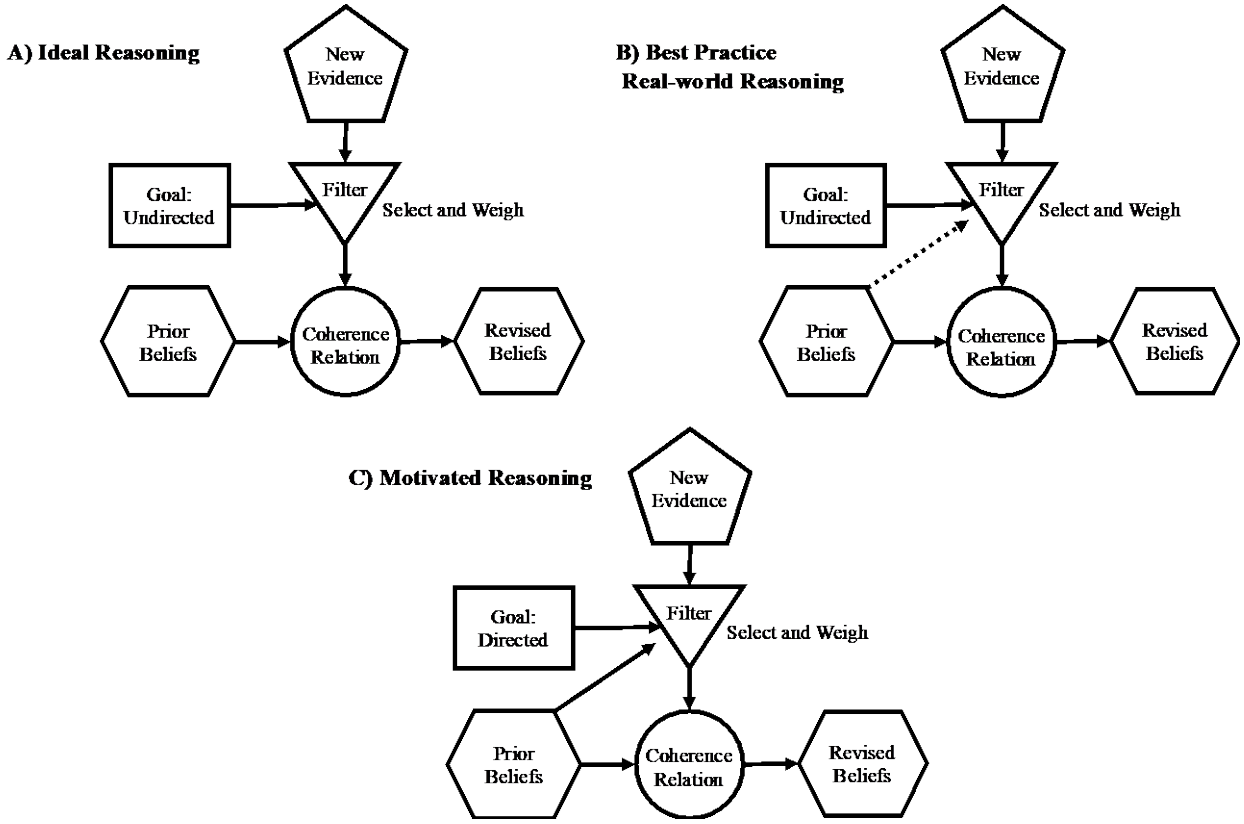


Figure 1. *Process of belief revision under different models. As new evidence comes in, prior beliefs need to be updated and revised via the coherence relation. The models vary in terms of goal directedness or vested interest of the reasoner and how these interests and prior beliefs impact the filter. The purpose of the filter is to select and weigh the evidence. (a) This is reasoning in an ideal world. The reasoner is disinterested in the outcome, the filter lets through*

all relevant information, and the coherence relation revises beliefs accordingly. While no individual human can implement this model, scientific methodology comes close. (b) This is best practice in real-world reasoning. The reasoner tries to maintain a disinterest in the outcome and minimize the impact of prior beliefs on the filter. This is indicated by the dashed line connecting prior beliefs to the filter. (c) The motivated reasoning model is characterized by goal directedness and a robust, active impact of prior beliefs on the filtering of evidence in the service of a directed goal. This can result in the shielding existing of beliefs from evidence. Reproduced from Goel (2022).

Veridicality of beliefs is critical because beliefs in conjunction with desires lead to actions. That is, human volitional behavior is thought to be determined by our beliefs, desires and a principle of coherence (or "making sense"). We have already introduced beliefs. Desires (or goals) are also psychological states with propositional content and have a mind to world direction of fit (i.e. a desire is satisfied not by changing your mind but changing the world). Coherence for our purpose is an intuitive notion meaning roughly "making sense." Given certain beliefs and desires, the principle of coherence determines whether a particular action makes sense or is reasonable. This is our basic model of rational behavior.

Consider a judge adjudicating a parole application (Figure 2). She might begin with a desire to uphold the law and protect society from harm and then, based on the presented information form certain beliefs about the applicant's original crime and sentence, the number of years served, behavior during incarceration, and potential to reoffend, etc. She would apply the principle of coherence (to the relationship between actions, desires and beliefs) to determine whether to grant or deny parole. In the example in Figure 2, it may be reasonable to grant parole. If her beliefs about the applicant were otherwise, a very different action might be considered reasonable. So, veridicality of beliefs matters. Yet despite the existence of this machinery, we all harbor false beliefs that we are reluctant to change despite the evidence.

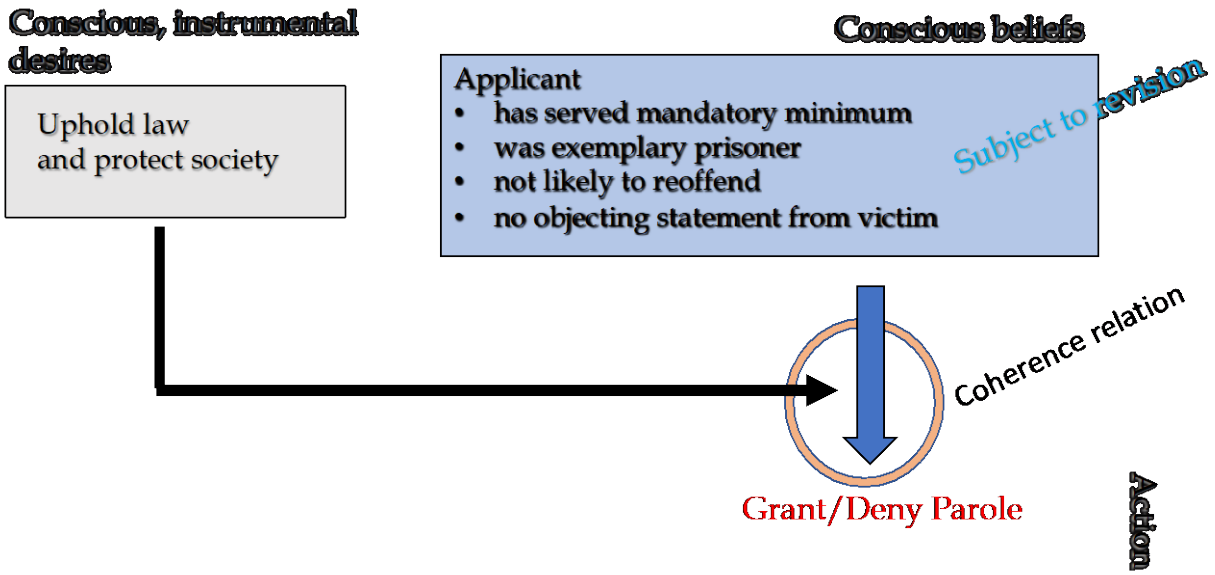


Figure 2. *The Machinery of Rationality. Notice that you change behavior by changing beliefs.*

Explaining Failure of Belief Revision as Motivated Reasoning

The most common explanation for failure to change beliefs in the face of counter evidence is motivated reasoning (Kunda, 1990; Kunda & Sinclair, 1999). Motivated reasoning is "reasoning with vested interests." Reasoning and argumentation in the courtroom provide a succinct illustration. The presiding judge enters the courtroom without any prior beliefs as to the guilt or innocence of the defendant and has no personal interest in the outcome of the case. She listens to the evidence, as impartially as possible, and afterward revises her beliefs to either guilty or not guilty. This reasoning should be captured by the ideal reasoning model (figure 1a) but is more realistically an instance of the best practice real-world reasoning model (figure 1b). The prosecutor, on the other hand, enters the courtroom believing that the defendant is guilty and has the very specific goal of convincing the judge of this. These prior beliefs and goal directedness will result in a calculated, judicious selection and weighted presentation of the evidence: "Your Honor, the defendant lunged across the counter to threaten the cashier." The defense lawyer, on the other hand, enters the courtroom believing that his client is innocent. His goal is to convince the judge of the same. His goal directedness and prior beliefs also result in a calculated selection and weighting of the evidence: "Your Honor, the defendant leaned across the counter to speak to the cashier." This is motivated reasoning, indicated in figure 1c.

The motivated reasoning explanation for false beliefs is that people reason like attorneys rather than like judges. They have certain beliefs and desires and filter, twist, and select the evidence to advance them. In the courtroom, there are certain constraints imposed by the legal system (enforced by the judge) on the level of filtering and quality of information allowed to be presented by the attorneys. The presence of these constraints and an impartial, disinterested judge usually allows for evidence-based decisions.[1] In the absence of these constraints, arriving at evidence-based decisions can be more challenging. Either way, it should be apparent that motivated reasoning is not flawed reasoning. It is rational. The reasoning mind is being utilized to achieve a particular goal. This is the purpose of the reasoning mind.

Let us examine the explanatory scope of this model using the following three beliefs, two from science and one from politics:

1. Peptic ulcers are caused by stomach acidity and stress.

2. Gender is a purely social construct.

3. Russia's invasion of Ukraine is an unprovoked war of aggression started by a cartoonishly evil madman…

These are all legitimate beliefs in so far as they advance propositions that can be evaluated against states of affairs in the world and determined to be either true or false (even if only to degrees).

Belief 1: Excess Stomach Acid Causes Peptic Ulcers

For decades the medical community was certain that peptic ulcers were caused by excess stomach acidity. There was widespread agreement among scientists and practitioners, and it was considered a closed issue. This conclusion was based on data, and reasoning captured by the best practice real-world reasoning model (Figure 1b). In the 1980s two Australian scientists, Barry Marshall and Robin Warren, resurrected an old, discarded idea that ulcers were caused by

[1] There are of course many documented failures of the legal system attributable to judges (and juries) also being goal-directed, motivated reasoners.

bacteria, specifically, Helicobacter pylori. The bacteria theory of ulcers had been proposed multiple times over the previous 100 years but was repeatedly set aside due to lack of conclusive evidence. Scientists were unconvinced that bacteria could survive in the acidic environment of the stomach. A particularly influential study in the 1950s had failed to find any bacteria in the stomach. So many in the medical community initially ignored, even ridiculed Marshall and Warren and refused to believe their claims. The reputations and careers of individual scientists were at stake. The individuals who had developed, nursed, and defended the acid theory of ulcers were highly motivated to believe and advocate for it, as per the motivated reasoning model (Figure 1c). Identities become associated with theoretical contributions and people resist falsification by questioning and rationalizing the evidence, just like attorneys. There is always some evidence for any claim and some evidence against any claim.

But the saving grace of science is that it doesn't care what individual scientists believe. Data are data, and in the end data always wins. Among any group of scientists there will be many different beliefs and goals. There will always be some individuals who are not hampered by strong prior beliefs and vested interests in any particular theory and will be guided by the data (or commitment to an alternative theory). This plurality of beliefs and interests overcomes individual shortcomings and allows science to approximate the ideal reasoning model (Figure 1a). Once the evidence is overwhelming, scientists must (and do) accept it and change beliefs or are passed over. In 2005, Marshall and Warren shared the Nobel Prize in medicine for their conclusive demonstration that bacteria can survive in the stomach and cause ulcers. This changed the treatment of ulcers and improved the lives of millions of patients. This is an example of the successful revision of a false belief in the face of motivated reasoning. This is how vetting, justification, and belief revision are supposed to work.

Belief 2: Gender Identity is a Social Construct

Society has long believed that gender identity is biologically determined by external sexual characteristics (and, more recently, chromosomes) and that sex and gender are one and the same. If societal norms are relevant in determining gender differences, it is only because boys and girls exhibit innate behavioral differences, which trigger differential treatment by parents.

Recently, a number of groups in Western societies (mostly from non-science disciplines) have begun arguing that gender identity is socially constructed, independent of biological sexual characteristics (Lorber, 2008). For example, the American Psychological Association (2014) notes that "gender refers to the socially constructed roles, behaviors, activities, and attributes that a given society considers appropriate for boys and men or girls and women." Any emerging behavioral differences between boys and girls are explained by differential treatment by parents who hold societally enforced gender stereotypes; for example, giving boys trucks and soldiers and girls dolls to play with. This differential treatment results in stereotypical gender roles such as women are emotional, caring, nurturing, dependent, and physically weak, while men are more competitive, aggressive, physically strong, and less emotional than women. The only correct way to assign gender is based on one's "internal sense of being male, female, or something else" (American Psychological Association, 2014) and presumably change it as necessary.

Gender refers to behavioral characteristics associated with being a male or female. Sex refers to one's endowment of reproductive organs. The issues are whether sex and gender are the same and whether gender is determined by biological factors or is a social construct or choice. Can one simply choose to be male or female? This would seem to be a question of biology (if not common sense).

What does biology say? All mammalian fetuses begin as females. Fetuses with a Y chromosome undergo the process of masculinization outlined in figure 3. The presence of the Y chromosome triggers two separate independent processes (one for masculinizing the body (sex organs) and the other for masculinizing the brain (behavior). The process of masculinizing the mammalian brain involves the conversion of testosterone to estradiol by the enzyme aromatase. The critical window for this process is the second half of pregnancy in humans. These hormonal processes serve to permanently sculpt developing neural systems by, among other things, either inhibiting or facilitating neural apoptosis (cell death) and modulating the formation and elimination of synaptic connections, in certain key subcortical brain regions such as the sexually dimorphic nucleus of the preoptic area (SDN-POA) in the anterior hypothalamus (Lenz et al., 2012; MacLusky et al., 1986; Matsuda et al., 2011; Morris et al., 2004; Nugent et al., 2015; Sato et al., 2004; Wu & Shah, 2011; Zuloaga et al., 2008). Thus, the normal unfolding of the gestation process results in a fetus with a male brain and a male body or a female brain and a female body

(figure 3).

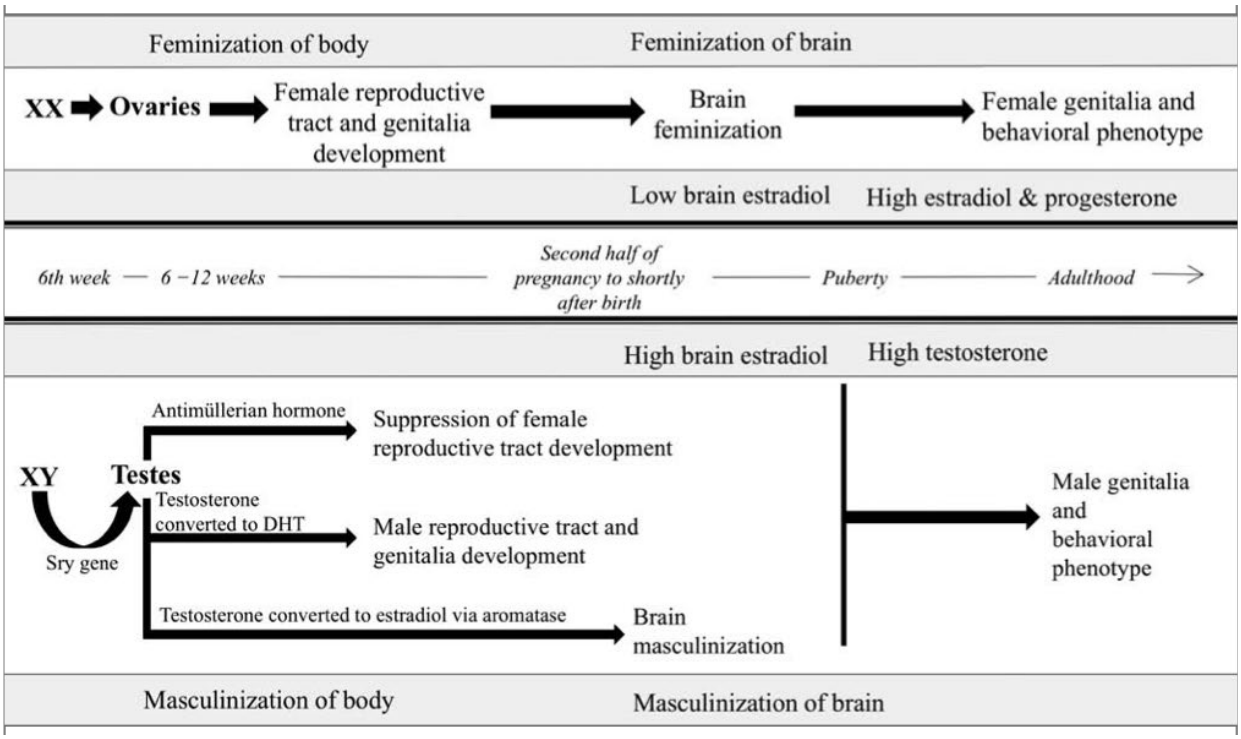


Figure 3. Differentiation of sex and gender in mammals.

In the vast majority of cases, these basic processes unfold normally.[2] However, notice the role of the Y chromosome; it determines the release of androgens at two different time points for two different purposes: masculinizing the body and masculinizing the brain. The independence of these processes means that there are two ways in which the modulation of the timing and/or quantity of hormone release can result in transsexuality (Swaab, 2007). Similarly, even in the "default" female developmental processes, hormonal imbalances during critical periods can disrupt normal sex and gender development (Nordenström et al., 2002). Furthermore, the process can also break down at the chromosomal level, as in the case of Klinefelter syndrome (Smyth & Bremner, 1998).

These data suggest that those claiming that gender identity is black and white are wrong

[2] Deviation rates are estimated at 1:10,000 for male to female and 1:30,000 for female to male (Swaab, 2007).

because the biology is saying that gender ambivalence is very real, but it is biological and rare.

Those claiming that gender is a purely social construct are wrong because the studies show that the sex-dimorphic preferences under consideration can be detected very early in human infants prior to any social gender category formation (Alexander et al., 2009), are present in other species (Alexander & Hines, 2002), and vary as a function of fetal androgen exposure (Nordenström et al., 2002). These data are speaking to the importance of biological factors, but they are *not* precluding the role of environmental factors. In fact, many mediating environmental factors are actually biological, including prenatal exposure to hormones, medications, environmental chemicals, and stress on the mother during pregnancy (Coolidge et al., 2002; Dessens et al., 1999; Zucker et al., 1996). Environmentally mediated factors in the unfolding of the two separate processes triggered by the presence of the Y chromosome can result in gender ambivalence.

What should be at issue is the extent to which postnatal social and cultural environmental factors interact with the biological factors. We understand much less about this. The most charitable counterinterpretation of the data is that even if prenatal biological and environmental factors do not mandate specific gender roles, sexual dimorphism surely predisposes humans to be more receptive to certain socially presented gender-specific cues than to others (McCarthy, 2016).

But neither side is willing to accept the evidence and for the same reasons. They wish to enact very different social policies and have a vested or motivated interest in using (or misusing) science to justify their preferred positions. Can we explain this as motivated reasoning? These positions are certainly motivated, but it is harder to argue that they are “reasoned.”

Belief 3: "Russia’s invasion of Ukraine is an unprovoked war of aggression…

On February 24, 2022, Russia invaded Ukraine. The official position of Western governments, media, and general population is that " Russia’s invasion of Ukraine is an unprovoked war of aggression and expansion to reestablish the "Russian Empire" or the Soviet

Union and eventually enslave Europe. Putin is an evil, deranged war criminal who must be stopped." It is an obligatory mantra to be repeated at every mention of the topic. Some academics such as John Mearsheimer of the University of Chicago and Geoffrey Sachs of Columbia University (to say nothing of the Russians) have a different view. They argue that "the Ukraine war is a predictable (and avoidable) last resort effort by Russia to prevent aggressive NATO expansion to their doorstep and support the Russian speaking population of eastern Ukraine from attacks by Western Ukrainian fascist-aligned military units. NATO expansion into Ukraine is an existential threat for Russia and they have said so on multiple occasions. Putin is a measured, rational agent acting in the best interest of Russia. Americans do the same thing as per the Monroe Doctrine (e.g. Cuban missile crisis)."

These are two very different views/beliefs of the same event. Which one corresponds more closely to the facts in the world and can the marshaling of evidence change minds on the other side? Unlike the previous 2 examples, data here does not come from science. But nonetheless, there should be some data/evidence in support of each belief. Upon examining the evidence, if we find our belief unsupported, or greater evidence for the other belief, we should be able to revise our beliefs.

Here are some data points in support of the 1st position. (i) Putin has often said that Ukraine and Russia have a long-shared history and are brothers. (This presumably indicates he wants to incorporate it into a greater Russia.) (ii) Russia did not invade the Baltic states when they joined NATO. (iii) There were no specific talks of Ukraine joining NATO at the time of the invasion. (iv) Russia started interfering militarily in eastern Ukraine in 2014 and annexed Crimea. (v) Putin and Russians speak very openly about defying American world hegemony and bringing about a multipolar world where multiple countries have a seat at the table. [This is of course unacceptable to the west/America.] Furthermore, if Russians are allowed to challenge Western supremacy, then other countries will be emboldened to do the same. (v) The invasion of Ukraine violates international law. This position is also supported by the following existing beliefs: (a) the west/Americans are the good guys and the only ones to be trusted with world hegemony (Boris Johnson said it out loud in a recent op-ed). This belief requires no evidence. We will see below why it is self-evident. (b) Putin longs to reestablish the Soviet empire. If successful in Ukraine he will march into Poland, Berlin, and end up in Portugal. (The actual

evidence here seems to be nonexistent in terms of military preparation and resources. It may be based on half a quote from Putin: "anyone who does not miss the Soviet Union does not have a heart," but the quote went on to conclude "anyone who wants to see it return does not have a head".) (c) Ukrainians should be able to choose their own destiny.

Here are some data points in support of the 2nd position as articulated by John Mearsheimer and Geoffrey Sachs, all pointing towards aggressive actions by the Americans to militarily and economically cripple Russia, using Ukraine as a proxy:

- 1991 Soviet Union dismantles itself. There are promises from the Americans that NATO will not move "one inch eastward".
- 1999 & 2004 Warsaw Pact nations are incorporated into NATO. The Russians object but are too weak to act.
- 2000 Putin asks to join NATO but is rebuffed by the Americans.
- 2008 Ukraine is invited to join NATO. Russians indicate that this is their backyard and an extensional red line for them.
- 2014: A popular Ukrainian elected government is deemed too pro-Russian and overthrown in a coup for a USA selected government (substantial credible evidence of US involvement). A dozen CIA bases are established Ukraine along the Russian border. A 2014 Gallup poll indicates that majority of Ukrainians are not interested in joining NATO with Western Ukraine leaning to Europe and eastern Ukraine leaning to Russia. There are popular uprisings in Russian-speaking eastern Ukraine and Crimea seeking a measure of autonomy but are opposed by the new Kyiv government with military force. Russia intervenes to support the Russian speaking population in Donbass and Crimea. Crimea votes to join Russia and is annexed. The Donbass region also asks Russia to annex them but Russia stops short. The USA starts sanctioning Russia.
- 2014-15: Minsk 1 and Minsk 2 Accords are brokered by the Europeans to stop the internal fighting in Ukraine and undertake constitutional reforms granting self-government to certain areas of Donbass and restoring control of the state border to the Ukrainian government. The Russians agree but the agreements are never implemented by Ukraine. (Angela Merkel later admits the discussions were just a delay tactic for NATO to arm and train the Ukrainian Army.)

- 2019: RAND Corporation prepares a report commissioned by the American military entitled “Overextending and unbalancing Russia” outlining various economic and military options for weakening Russia and their cost/benefit and risk assessments.
- 2021: In December Putin sends a letter to Biden asking to negotiate a settlement in Ukraine sensitive to Russian security interests. Biden rebuffs him.
- 2022: On February 24, Russia invades Ukraine. In April a peace deal (Istanbul Communiqué) is tentatively initialed by Ukraine and Russia. It calls for neutrality for Ukraine, UN supervised elections in Donbass region, and Crimea stays with Russia (Financial Times, March 2022). The Americans send Boris Johnson to scuttle the deal and convince Ukrainians to fight on because Russia is weak and with the support of the West they can win (confirmed by recent admission of a member of the Ukrainian negotiating team).
- America views the Ukraine war in terms of a geopolitical strategy to weaken/destroy Russia (as per the RAND report) at Ukraine's expense (Senators Mitt Romney and Lindsey Graham: “without expending American lives”). At the beginning of the conflict there is open talk in the American State Department of undertaking regime change in Russia and even breaking it up into smaller parts.
- Mearsheimer points out that Russia is not powerful enough to be a regional hegemon in Europe. They are not even interested in taking all of Ukraine much less the Baltic states and Berlin. The same assessment was reached in the 2019 RAND report. They have been there, done that, and understand it did not work out well. It bankrupted the USSR. They are not currently interested in repeating that mistake.

As rational agents we should be able to sit down, consider the evidence/data support both sides, and reach some sort of consensus underlying the most plausible reasons for the invasion and solutions to end the war. (Notice that different beliefs about the causes of the war point to different resolutions.) Is this really an unprovoked war of aggression or a rational move in the face of a very real threat from American hubris and hegemony? We are not capable of doing this. The majority of Westerners hold the first view and regard anyone holding the second view

to be "Putin puppets parroting Russian talking points." The second group regards the first group as naïve about great power politics and ignorant of the facts.

Motivated Reasoning as Question Begging

Motivated reasoning is the dominant framework for explaining adherence to false beliefs in the face of counterevidence. It does work in certain cases but in others it is largely unsatisfying. In our three examples, it offers a reasonable account of the first but falls short of explaining the second and third examples. It may be guilty of question begging in these latter cases. The question is, in the face of counterevidence, why can't we use the reasoning machinery to change beliefs about gender being a social construct and the Ukraine war? That is, why can't we apply reasoning recursively to see whether it is rational to hold these beliefs in the first place in light of the evidence? This is presumably prevented by the motivated goal or desire (e.g. social policy preference, geopolitical dominance). But why can't the reasoning machinery be turned onto the desire itself to see whether it is rational?

There is a common view that rationality is a tool or instrument for achieving some goal or desire. It is not for the tool to question the end. Hume ([1739] 1888, p. 416) famously put it thus:

> "Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. 'Tis not contrary to reason to prefer the destruction of the world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater and have a more ardent affection for the former than the latter."

Notice that Hume does qualify the unquestioned supremacy of desires, that they not be "founded on false suppositions." Hume is recognizing that some desires are cognitive in that they are arrived at through reason. If so, we should be able to change them by reason if they are founded on false beliefs.

Let us return to our prosecuting district attorney who, based on evidence is convinced of the guilt of a defendant and formulates the goal or desire to prosecute the individual. New compelling evidence arises indicating the individual is not guilty. This evidence forces the prosecutor to change his beliefs about the defendant's guilt, and this in turn changes his desire to prosecute the individual to a desire not to prosecute. In this classic paradigm of motivated reasoning, the machinery of reason can indeed be turned onto itself to modify cognitive desires, and false beliefs supported by those desires.[3] This is largely what is happening in the peptic ulcer example. The desire to support the acid theory is a cognitive desire (unless you are the originator of the theory) and is overturned when (and only when) the evidence becomes overwhelming.

But not all desires are cognitive. Many desires arise from lower-level autonomic, instinctual, and associative learning systems in the brain stem, diencephalon, and subcortical brain systems. They will lack propositional content and as such not be susceptible to reason (coherence relations). One such example is my desire to add sweetener to my coffee, even though I'm diabetic. I'm capable of understanding the basic biology underlying diabetes, the need to limit carbohydrate intake, and I certainly don't have a death wish. Yet I do add sweetener to my coffee because it tastes good. This desire is generated by primordial systems anchored in old brain neural circuitry and is not particularly amenable to reason (though I can choose to use artificial sweetener).

I want to suggest that some false beliefs are resistant to rational evidence because the motivating desire is not a cognitive desire. It is a desire propped up by instinctual (and other) low-level subcortical systems. I believe that some researchers are reaching for such an explanation when they invoke emotions, particularly anger (Mullen & Skitka, 2006), and "identity protection" to buttress models of motivated reasoning (Kahan, 2016, p. 3): "The truth independent goal of 'politically motivated reasoning' is identity protection: the formation of beliefs that maintain a person's status and affinity in a group united by shared values." But there is no straightforward way of accommodating these intuitions in standard models of rationality

[3] If the desire is noncognitive, e.g. racial bias, the prosecution may proceed despite the evidence as explained below by the tethered rationality model.

(figure 2). I propose a model of tethered rationality that provides a much more natural way of encompassing and expressing these insights.

Introduce Tethered Mind Model

While I accept that we are the reasoning animal, the untethered reasoning mind in figure 2 is a myth. Reason does not hover in midair above the biology; nor is it powered by angel dust. That is, while we are not bats and baboons, our system of reasoning evolved on top of— and is tethered to— the neurobiology we inherit from our common ancestors with bats and baboons. More specifically it is tethered to evolutionary older systems such as the autonomic system, the instinctive system, and the associative system. Unless the basic laws of biology changed with the emergence of Homo sapiens and reason, the tethered mind must be a truism.

Each of these systems are evolved solutions (i.e. adaptations) to different types of problems. We have already introduced the reasoning system. The autonomic system controls various bodily functions such as blood sugar levels, digestion, breathing, etc. Common examples of instincts in humans range from the suckle response in infants to our ability to recognize cheaters (Cosmides et al., 2005). The associative system underlies much of our learned behaviors, such as driving a car or behaviors driven by "likes" on social media (Skinner, 1984). Each system is different from the other and from reasoning.

But despite these multiple systems, behavior consists of a single response. This implies a mechanism that takes input from each system and combines or blends it into a single behavior (Figure 3). That is, our behavior is affected not just by reason, but by reason interacting with these various systems. This allows for the possibility of not just rational and irrational behaviors but also *arational* behaviors. Unlike irrational behaviors which are defined as a failure of coherence, I define arational behaviors as those driven by, or at least modulated by one or more of the nonrational systems. The structures and mechanisms of propositional content and coherence relations do not arise in these systems.

This is a simple, commonsense idea but it raises two challenges: (1) how do these various systems communicate and interact, and (2) what is the control structure? Details aside, I have proposed that (causally efficacious) *feelings* of pleasure and displeasure are associated with each

system and they provide the common communication language. For example, unpleasant hunger pangs are generated by the autonomic system to trigger food foraging/eating behavior. Intense pleasure associated with sexual arousal leads to the search for a mate with the aim of copulation. Parental instincts generate pain/distress when our children suffer and lead us to help them at all costs. The pleasure of reward and pain of punishment are instrumental for reinforcing associations. More controversially, I have argued that even coherence relations are associated with pleasurable feelings of rightness, because they or fitness enhancing, while incoherence/inconsistency is associated with unpleasant feelings because it is harmful (Goel, 2022).

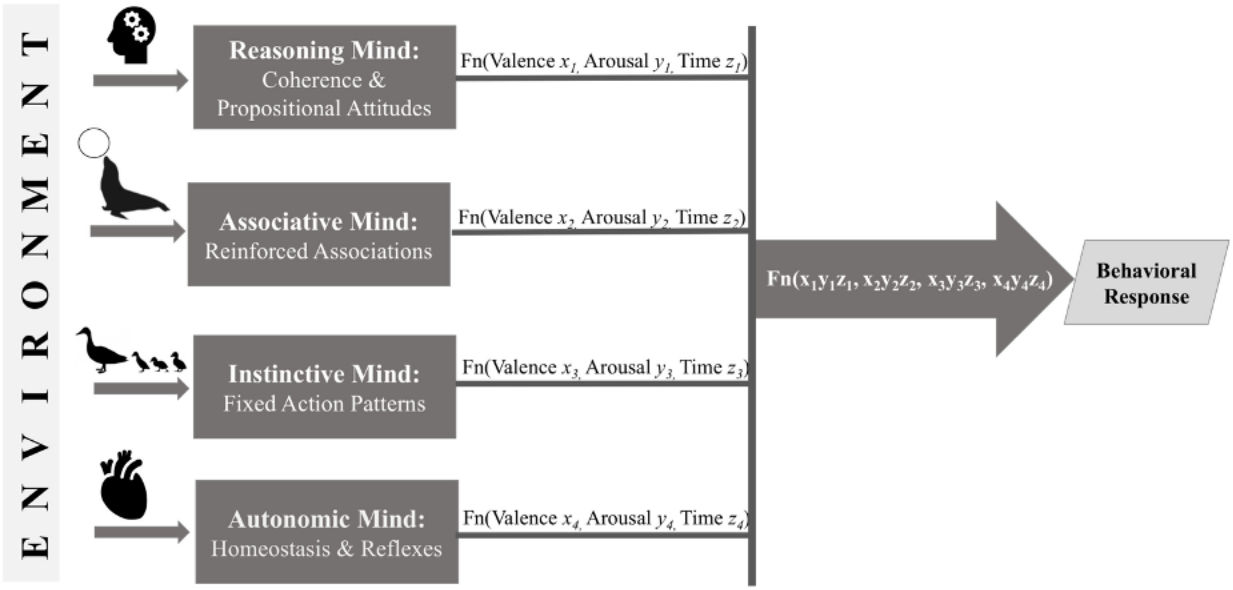


Figure 4. *The tethered mind. Human behavior is a function of the responses of the autonomic system, the instinctive system, the associative system, and the reasoning system to any given situation. While the mechanism underlying each system is different, all utilize the common currency of feelings. The response generated by each system is also in the currency of feelings, with valence, arousal, and duration components. The system is set up to maximize pleasure and minimize pain or displeasure. The selected behavior will usually be a blended response based on*

the output of all systems. There is no central executive in charge. The reasoning system has an input into the response, but so do the other systems. Individual differences in behavior are explained not just in terms of individual differences at the level of reasoning but also individual differences at the level of the autonomic, instinctive, and associative systems. A notion of variable individual effort may be captured in this model by variability in levels of arousal associated with different systems. Reproduced from Goel (2022).

The second issue is the control structure. Who is in charge of the tethered mind? All or several systems may be triggered in any particular situation. In some cases the same action may be initiated by multiple systems, while in other cases, different, even contradictory actions may be triggered. Either way, the overall response is determined by combining the sensory feelings generated by individual systems, guided by the *principle of maximizing overall pleasure and minimizing displeasure*. There is no central executive in charge. The model is motivated and described in some detail in Goel (2022).

This model of tethered rationality is not restricted to coherence relations among propositional contents. It has a larger repertoire of mechanisms, consisting of the autonomic mind, the instinctive mind, and the associative mind, in addition to the reasoning mind, to appeal to in explaining human behavior. I will now introduce the instinct of ingroup/outgroup and modulates the reasoning system and show how it affects belief revision in our three examples.

Ingroup/outgroup Instinct

It is widely held that humans naturally/automatically group themselves into "us and them" (Sherif, 1988). Groupings can range from evolutionarily and socially/culturally salient divisions such as kinship, sex, race language, nationality or territory, diet, political ideology, income/resources, etc. to what are known as "minimal groups," such as forming an ingroup with strangers standing in the same line at the grocery store versus those standing in another line (outgroup) (Tajfel, 1970; Tajfel & Turner, 2004). Whatever the basis of the distinction, our group is good and pure of heart; the other group deceitful and dangerous. The most widely accepted explanation of the phenomenon is offered by Social Identity Theory (Tajfel & Turner, 2004)

(though see (Pratto et al., 2006; Rubin & Hewstone, 2004)). It states that group categorization leads to the assignment of (socially constructed) properties and behavioral dispositions of the group to the individual. The group we belong to is the ingroup; the other group is the outgroup. Our self-esteem is attached to our ingroup identity. Furthermore, we have a need to maintain high self-esteem. Therefore, our group must compare favorably to other groups, leading to the promotion of our ingroup and discrimination and hostility against the outgroup (Hamlin et al., 2013). Importantly, the animosity is directed at the group, not necessarily the individual (Tajfel & Turner, 2004). We can, in certain circumstances pullout individual members from an outgroup to an ingroup if we categorize them as friends or make exceptions for other reasons.

I think there is a case to be made that the ingroup/outgroup phenomenon is not a social construct but rather an instinct. Instincts are innate, species- (and sex/gender-) specific stereotyped behaviors (available from birth or within a developmental window) that are triggered by specific stimuli. Instinctive behaviour is compulsive but is not a reflex arc or robot-like. It contains "an element of appetite, or aversion, or both" (Craig, 1918). 'Appetite' refers to a state of agitation in the absence of the stimuli (e.g. hunger pangs in the absence of food or craving for group membership) and receiving the stimuli (food; membership) is "consummatory" or satisfying/pleasurable. It relieves the agitation and returns the animal to rest. An "aversion" is a state of agitation resulting from the presence of certain stimuli (e.g. a snake or threat presented by outgroup) and ceases when the stimulus is withdrawn/destroyed. Grouping contains both an appetitive and aversive element. There is great pleasure and belonging to the ingroup, and perhaps even greater pleasure in hating the outgroup (Hamlin et al., 2013).

One can differentiate between social constructs and instincts in terms of the answers to the following five questions (Goel (2022)): (1) Is the trait universally present in human societies and not culture-specific. (2) Is it available on other branches of the phylogenetic tree. (3) Does it emerge early in human infants, prior to any opportunity for extensive socialization. (4) Is it underwritten by implicit, automatic, low-level mechanisms. (5) Is it possible to trace specific neural circuitry devoted to it (and find homologous behavior and circuitry in other species). One can give affirmative answers to at least the first four of these questions (Amodio, 2014; Hamlin et al., 2013; Kawakami et al., 2020; LeVine & Campbell, 1972; Mahajan & Wynn, 2012) providing support for an instinct interpretation of grouping.

Think of the grouping instinct in terms of an analogy with the imprinting instinct found in some birds. The newly hatched chick will imprint upon the first moving object that it encounters, be it the mother bird or Konrad Lorenz, or even Konrad Lorenz's yellow boots! The imprinting itself is an instinct. What is imprinted upon is an accidental feature of the environment. I think the ingroup/outgroup phenomenon works in the same way. In the case of grouping the system is sensitive to familiarity/unfamiliarity. Familiarity leads to preference (Goel, 2022). What is familiar/and familiar is an accidental feature of our environment. The detection of differences activates the instinct resulting in the formation of groups and being favorably disposed to the ingroup and suspicious of the outgroup.

Tethered Rationality and Failure of Belief Revision

Let us now return to our three beliefs with the model of tethered rationality and the instinct of ingroup/outgroup to explain behavior. What is critical about the tethered rationality model is that our behavior is a function not just of reason, but reason modulated by noncognitive (i.e. autonomic, instinctive, and associative) mechanisms. Behavior is a blended response of these various systems. The output of each of these systems, including reason, is in terms of affect or feelings (with valence, arousal, and temporal duration dimensions). The control strategy is to maximize pleasure and minimize displeasure.

In the case of the peptic ulcer example, the desire to prefer the acid theory was a cognitive desire based on evidence. Even so, if you are committed to the acid theory, the evidence for the bacteria theory will be inconsistent with your existing beliefs resulting in unpleasant discordant feelings. However, as additional evidence is accumulated feelings of discord will be replaced by positive feelings of coherence. Some noncognitive factors will also intervene including grouping preferences. Here the grouping is triggered by a cognitive difference in terms of the most coherent relationship between the two theories and data. Those committed to the acid theory feel a sense of ATTACHMENT to their ingroup and are suspicious of the new evidence for the bacteria theory presented by outgroup members. [4] The pleasure

[4] The terms in all-caps are instinctive systems identified by Panksepp (2011)

associated with the coherence provided by the new evidence, will be pushed back upon by the pleasure of ingroup ATTACHMENT and outgroup suspicion. This will require greater preponderance of evidence to revise beliefs. When this evidence was provided the beliefs were revised. If you are a member of the outgroup looking to overthrow the acid theory, then your group membership and coherence relations push in the same direction and you will be satisfied with a lower degree of evidence. This can be interpreted within the motivated reasoning framework in untethered models of reasoning.

In the case of the gender example, the difference on which group formation is based is not a cognitive judgment about which theory the data support, but rather a social policy preference of how we wish to live and organize society. In this case, the issues of veridicality and coherence are secondary. Instinctual factors such as group membership come to the forefront. As group bias itself is not a social construct based on beliefs, it cannot be reasoned away. In-group members are friends; out-group members are evil and wish to do us harm and destroy society. If the scientific evidence supports the policy preferences of the other group (or neither group), then science itself gets relegated to the out-group and subject to attack and ridicule.

There are powerful feelings of pleasure associated with membership in the in-group (perhaps through activation of the ATTACHMENT system) and equally powerful feelings associated with FEAR and RAGE, directed at the out-group. Striving to discredit the out-group activates the SEEKING system. These feelings override any pleasurable feelings associated with coherence relations between evidence and conclusion. They can be intoxicating. We seek them out and maintain them by forming bubbles or echo chambers that reinforce the false beliefs that allow for the contemptuous dismissal of the out-group members, along with their evidence. There need not be a great deal of reasoning involved in these processes. They are automatic. They can be rationalized by the cognitive mind after the fact and articulated in terms of beliefs and desires and reason, but these are not the driving mechanisms. The machinery of instincts is driving the tethered mind in these situations. It is not inevitable, but unless people can derive greater pleasure from coherence relations than from activation of In-Group/Out-Group systems, dismissal of the outgroup is an unfortunate consequence of the tethered mind.

The third belief is about international politics, where no scientific evidence enters the picture. Homo sapiens are driven to maximize resources. Being dominant and higher in the social hierarchy means access to more resources. Politics is about negotiating and maintaining power alliances within and across social groups as a means to this end, both at individual and nation-state levels. The grouping criteria here are very raw and fundamental, a combination of shared worldviews and territoriality. Western media reminds us daily that we are fair, benevolent, honest, upright, peaceful, beloved of God, the essential nation/peoples, and always looking out for and assisting our fellow man, while Russian are less than human (orcs), liars, cheaters, murderers and rapists, show no value for human life, and want to conquer and enslave us.

The source of the specific belief in the example may be a statement such as "Russia has invaded Ukraine." As Russia is an outgroup member with the above properties this is sufficient to trigger suspicion and FEAR and RAGE, along with the associated feelings and actions. This particular example also involves POWER/DOMINANCE, resource maximization and territoriality instinctive systems. The interaction of In-Group/Out-Group, FEAR, RAGE, DOMINANCE, resource maximization and territoriality is a toxic mix. There is also the pleasure associated with the ATTACHMENT system as a function of bonding with the in-group (reveling in our goodness/godliness), and the pleasure in the activation of the SEEKING system to crush the out-group. Once these instinctive systems are triggered, they are going to activate the relevant action tendencies to (in this case literally) destroy the outgroup. Notice also the role of the reasoning system here. It can be consciously, intentionally used to amplify the rhetoric against the outgroup, which will further accentuate the instincts, which will further amplify the rhetoric, resulting in a dangerous escalating cycle leading to the cry that "Russia must be destroyed" (even if it means destroying Ukraine).

A purely reason-based analysis in terms of motivated reasoning cannot explain these phenomena. The motivated reasoning strategy can explain some of scientific and legal reasoning. It can explain why someone may consciously press someone else's instinctive buttons to further their agenda, such as raising alarm that the "Russians are coming for Europe once they are done with Ukraine" to garner support for raising taxes and increasing military expenditures, but it cannot explain why the rational mind falls into the trap given the counter evidence. The model

of tethered rationality along with particular instincts provide a much more compelling explanation. Once the buttons activating the In-Group/Out-Group system and other instincts are triggered, the pleasure associated with the activation of these systems may be much greater than any pleasure (or indeed displeasure) associated with any evidence and coherency associated with the counterargument.

Notice the same model explains the behavior of the Russians with the ingroup/outgroup membership reversed. There are no satisfactory purely rational explanations of these behaviors.

References

Alexander, G. M., & Hines, M. (2002). Sex differences in response to children's toys in nonhuman primates (Cercopithecus aethiops sabaeus). *Evolution and Human Behavior*, *23*(6), 467–479. https://doi.org/10.1016/S1090-5138(02)00107-1

Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex Differences in Infants' Visual Interest in Toys. *Archives of Sexual Behavior*, *38*(3), 427–433. https://doi.org/10.1007/s10508-008-9430-1

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), Article 10. https://doi.org/10.1038/nrn3800

Coolidge, F. L., Thede, L. L., & Young, S. E. (2002). The Heritability of Gender Identity Disorder in a Child and Adolescent Twin Sample. *Behavior Genetics*, *32*(4), 251–257. https://doi.org/10.1023/A:1019724712983

Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005). Detecting cheaters. *Trends Cogn Sci*, *9*(11), 505–506; author reply 508-10. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16198616

Dessens, A. B., Cohen-Kettenis, P. T., Mellenbergh, G. J., Poll, N. V. D., Koppe, J. G., & Boer, K. (1999). Prenatal Exposure to Anticonvulsants and Psychosexual Development. *Archives of Sexual Behavior*, *28*(1), 31–44. https://doi.org/10.1023/A:1018789521375

Goel, V. (2022). *Reason and Less: Pursuing Food, Sex, and Politics*. The MIT Press.

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not Like Me = Bad: Infants Prefer Those Who Harm Dissimilar Others. *Psychological Science*, *24*(4), 589–594. https://doi.org/10.1177/0956797612457785

Hume, D. (1888). *Treatise of Human Nature/Book 1: Of the understanding*. Clarendon Press; Wikisource.

Kahan, D. M. (2016). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. In *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–16). American Cancer Society. https://doi.org/10.1002/9781118900772.etrds0417

Kawakami, K., Hugenberg, K., & Dunham, Y. (2020). Perceiving others as group members: Basic principles of social categorization processes. In P. van Lange & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles, third edition*. Guilford Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kunda, Z., & Sinclair, L. (1999). Motivated Reasoning With Stereotypes: Activation, Application, and Inhibition. *Psychological Inquiry*, *10*(1), 12–22. https://doi.org/10.1207/s15327965pli1001_2

Lenz, K. M., Nugent, B. M., & McCarthy, M. M. (2012). Sexual Differentiation of the Rodent Brain: Dogma and Beyond. *Frontiers in Neuroscience*, *6*, 1–13. https://doi.org/10.3389/fnins.2012.00026

LeVine, R. A., & Campbell, D. T. (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior* (pp. ix, 310). John Wiley & Sons.

MacLusky, N. J., Naftolin, F., & Goldman-Rakic, P. S. (1986). Estrogen formation and binding in the cerebral cortex of the developing rhesus monkey. *Proceedings of the National Academy of Sciences*, *83*(2), 513–516. https://doi.org/10.1073/pnas.83.2.513

Mahajan, N., & Wynn, K. (2012). Origins of "Us" versus "Them": Prelinguistic infants prefer similar others. *Cognition*, *124*(2), 227–233. https://doi.org/10.1016/j.cognition.2012.05.003

Matsuda, K. I., Mori, H., Nugent, B. M., Pfaff, D. W., McCarthy, M. M., & Kawata, M. (2011). Histone Deacetylation during Brain Development Is Essential for Permanent Masculinization of Sexual Behavior. *Endocrinology*, *152*(7), 2760–2767. https://doi.org/10.1210/en.2011-0193

McCarthy, M. M. (2016). Multifaceted origins of sex differences in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1688), 20150106. https://doi.org/10.1098/rstb.2015.0106

Morris, J. A., Jordan, C. L., & Breedlove, S. M. (2004). Sexual differentiation of the vertebrate nervous system. *Nature Neuroscience*, *7*(10), 1034–1039. https://doi.org/10.1038/nn1325

Mullen, E., & Skitka, L. J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger? *Journal of*

Personality and Social Psychology, *90*(4), 629–643. https://doi.org/10.1037/0022-3514.90.4.629

Nordenström, A., Servin, A., Bohlin, G., Larsson, A., & Wedell, A. (2002). Sex-Typed Toy Play Behavior Correlates with the Degree of Prenatal Androgen Exposure Assessed by CYP21 Genotype in Girls with Congenital Adrenal Hyperplasia. *The Journal of Clinical Endocrinology & Metabolism*, *87*(11), 5119–5124. https://doi.org/10.1210/jc.2001-011531

Nugent, B. M., Wright, C. L., Shetty, A. C., Hodes, G. E., Lenz, K. M., Mahurkar, A., Russo, S. J., Devine, S. E., & McCarthy, M. M. (2015). Brain feminization requires active repression of masculinization via DNA methylation. *Nature Neuroscience*, *18*(5), Article 5. https://doi.org/10.1038/nn.3988

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? *Neuroscience & Biobehavioral Reviews*, *35*(9), 1791–1804. https://doi.org/10.1016/j.neubiorev.2011.08.003

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, *17*(1), 271–320. https://doi.org/10.1080/10463280601055772

Rubin, M., & Hewstone, M. (2004). Social Identity, System Justification, and Social Dominance: Commentary on Reicher, Jost et al., and Sidanius et al. *Political Psychology*, *25*(6), 823–844. https://www.jstor.org/stable/3792280

Sato, T., Matsumoto, T., Kawano, H., Watanabe, T., Uematsu, Y., Sekine, K., Fukuda, T., Aihara, K., Krust, A., Yamada, T., Nakamichi, Y., Yamamoto, Y., Nakamura, T., Yoshimura, K., Yoshizawa, T., Metzger, D., Chambon, P., & Kato, S. (2004). Brain masculinization

requires androgen receptor function. *Proceedings of the National Academy of Sciences*, *101*(6), 1673–1678. https://doi.org/10.1073/pnas.0305303101

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Sherif, M. (1988). *The Robbers Cave Experiment: Intergroup Conflict and Cooperation. [Orig. Pub. as Intergroup Conflict and Group Relations]*. Wesleyan University Press.

Skinner, B. F. (1984). Behaviorism at Fifty. *Behavioral and Brain Sciences*, *7*(4), 615–667. https://doi.org/10.1017/S0140525X00027618

Smyth, C. M., & Bremner, W. J. (1998). Klinefelter Syndrome. *Archives of Internal Medicine*, *158*(12), 1309–1314. https://doi.org/10.1001/archinte.158.12.1309

Swaab, D. F. (2007). Sexual differentiation of the brain and behavior. *Best Practice & Research Clinical Endocrinology & Metabolism*, *21*(3), 431–444. https://doi.org/10.1016/j.beem.2007.04.003

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, *223*(5), 96–103. https://www.jstor.org/stable/24927662

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In *Political Psychology*. Psychology Press.

Wu, M. V., & Shah, N. M. (2011). Control of masculinization of the brain and behavior. *Current Opinion in Neurobiology*, *21*(1), 116–123. https://doi.org/10.1016/j.conb.2010.09.014

Zucker, K. J., Bradley, S. J., Oliver, G., Blake, J., Fleming, S., & Hood, J. (1996). Psychosexual Development of Women with Congenital Adrenal Hyperplasia. *Hormones and Behavior*, *30*(4), 300–318. https://doi.org/10.1006/hbeh.1996.0038

Zuloaga, D. G., Puts, D. A., Jordan, C. L., & Breedlove, S. M. (2008). The role of androgen receptors in the masculinization of brain and behavior: What we've learned from the testicular feminization mutation. *Hormones and Behavior*, *53*(5), 613–626. https://doi.org/10.1016/j.yhbeh.2008.01.013