



AKADÉMIAI KIADÓ

# Error management theory and the ability to bias belief and doubt

NATHAN J. FOX\* 

Independent Researcher, Canada

Received: September 28, 2023 • Revised manuscript received: February 10, 2024 • Accepted: March 1, 2024

DOI:

[10.1556/2055.2024.00049](https://doi.org/10.1556/2055.2024.00049)

© 2024 The Author(s)

## Culture and Evolution

## RESEARCH ARTICLE



### ABSTRACT

Error Management Theory (EMT) suggests that cognitive adaptations evolved to minimize the cost of false negative and false positive errors in detections of consequential environmental conditions. These adaptations manifest as biases tailored to specific environmental conditions. This paper proposes that the same selection pressure fostered the evolution of a self-biasing ability, allowing us to minimize such costs based on experience and culturally transmitted information. The research indicates that this ability specifically applies to productions of belief or doubt about the existence of an environmental condition that is not adequately perceptible to the senses, e.g., doubt that the man is being honest. A model of this self-biasing process, rooted in signal detection theory, is developed and its explanatory reach is demonstrated through many diverse examples from epistemological and social psychological literatures. These examples feature an evidentiary standard for belief that appears too high or low, or seems to have modulated. This study aims to correct and to significantly enhance EMT by introducing this evolved self-biasing ability.

### KEYWORDS

bias, error management theory, belief, knowing, doubt, context dependence, individual differences

## INTRODUCTION

Human survival and thriving is dependent on our ability to select or create behaviours that best advance our interests, and this in turn is dependent on our ability to make true detections and avoid false detections of environmental conditions that are consequential. So, we can incur a significant cost when we make a false negative error (assuming that a condition does not exist when it does) and when we make a false positive error (assuming that a condition does exist when it does not). Error management theory (EMT; Haselton & Buss, 2000; Haselton & Nettle, 2006; Haselton, Nettle, & Murray, 2015) recognizes that our distant ancestors were subject to selection pressure for cognitive adaptations that would help them minimize these costs. In particular, selection pressure would have arisen if there was a significant difference between the cost of a false negative and the cost of a false positive in relation to a consequential environmental condition. Selection would have favoured a biased tendency toward avoiding the more costly error. Such a “bias sometimes increases overall error rates, but by minimizing the more costly error, it minimizes overall cost” (Haselton & Nettle, 2006, p. 48).

Even so, cost asymmetries of this kind would also have produced selection pressure for an entirely different kind of adaptation, where individuals have the ability to bias themselves to achieve comparable outcomes. They would do so on the basis of an assessment of the extent and direction of an asymmetry in error costs that is informed by experience and culturally transmitted information. This would require modulating their standard of evidence for detecting a consequential environmental condition across a considerable range. While traditional EMT’s inherited biases would originate in the error costs incurred by distant ancestors, this inherited ability would generate *learned biases* that reflect the individual’s

\*Corresponding author.

E-mail: [Nathan.J.Fox@ncis.org](mailto:Nathan.J.Fox@ncis.org)



experience and learning. (Still, traditional EMT does not suggest that experience and cultural learning cannot mitigate the effect of an inherited bias.)

If such learned biases are produced, they would not be difficult to identify. Those who display one would plausibly have recognized an asymmetry in error costs, and done so on the basis of their experience, cultural learning, life circumstances, and value priorities. Moreover, many people would not display the latter bias at all, and it would be plausible that these individuals *did not recognize* an asymmetry in error costs given their particular experiences, priorities, and so on. When these elements are clear in relation to a particular bias, all other things equal, it will be far more likely that the bias was learned than that it was inherited. This will be referred to as the “Primary” indication of learning.

The particular biases that EMT proponents claim we inherit are of distinct types, and it seems that examples of two of these types *do not show* the Primary indication. They are sense perception biases and biases reflected in aversive physiological and emotional responses (examples to follow). So, this research acknowledges that we may well inherit biases of these types. However, examples of a third type *consistently* show the Primary indication. These are biases in either belief or doubt—in particular, belief or doubt about whether an environmental condition that is not adequately perceptible to the senses exists, e.g., doubt about whether the man is being honest. Biased beliefs and doubts about environmental conditions that *are* clearly perceptible to the senses seem generally to be products of sense perception biases, e.g., after climbing up the cliff and looking down, the woman believed it to be higher than what she perceived when she was looking up at it. Biases in beliefs and doubts about an environmental condition that is not adequately perceptible to the senses are distinguished from biases of the first two types in that they involve the individual consciously assessing whether the evidence is strong enough to warrant belief.

Many of these latter biases show not only the Primary indication of learning, but also another indication, viz., that it is implausible that the bias evolved by natural selection. This conclusion follows from it being implausible that the relevant error-cost-asymmetry persisted to any significant extent over a sufficiently long evolutionary period and under the varying conditions that would have been encountered. Of course, none of the examples that EMT proponents advance show this indication—though many examples of such cases are presented below.

Accordingly, the working hypothesis here shall be that biases in either belief or doubt about the existence of conditions that are not adequately perceptible to the senses are learned. Let us look at examples of the three types of bias outlined above.

An example of a *sense perception bias* pointed out by EMT proponents is reflected in the tendency to perceive objects that sound like they are approaching as being closer than are objects that sound like they are moving away, even though the objects are equidistant (Neuhoff, 2001). We also

have a tendency to perceive approaching objects that are threatening as approaching faster than are approaching objects that are not threatening. For example, approaching spiders are perceived to be approaching faster than are approaching rubber balls or ladybugs (Witt & Sugovic, 2013). These biases would tend to reduce the prevalence of false negatives for threats being dangerously close. Similarly, we perceive heights as being greater when we view them from above than from below (Jackson & Cormack, 2007; Stefanucci & Proffitt, 2009). This would tend to reduce the prevalence of false negatives for falls being badly injurious.

An example of a bias reflected in aversive physiological and emotional responses involves people with physical disabilities:

Because contagious diseases were often accompanied by anomalous physical features, humans plausibly evolved psychological mechanisms that respond heuristically to the perception of these features, triggering specific emotions (disgust, anxiety), cognitions (negative attitudes), and behaviors (avoidance). (Park, 2003, p. 65)

Let us turn to biases of the third type, examples of which consistently show the Primary indication of learning. Consider a bias in doubt that EMT proponents claim is inherited: Women tend to doubt men’s intentions to commit to a familial relationship during early stages of courtship (Haselton & Buss, 2000). EMT explains that the biological cost to women of false positives (assuming that a man intends to commit when he does not) tended to be greater than the cost of false negatives (assuming that a man does not intend to commit when he does) over the course of an evolutionary history. Being abandoned while pregnant “[i]n harsh ancestral environments ... could have been deadly to the woman’s offspring” (Haselton, 2007). Even so, the Primary indication of learning is evident. It is certainly plausible that women who display the bias learned something about the cost of a false positive (the cost of a pregnancy without support), and the cost would plausibly have weighed on them more than the cost of a false negative (losing an opportunity to form a committed relationship). Further, the bias is not at all universal: Women who are postmenopausal (Cyrus, Schwarz, & Hassebrauck, 2011, p. 13) and women who have a high level “of interest in a committed relationship” (Henningesen & Henningesen, 2010, p. 628) *do not* display the bias. It is certainly plausible that the cost of a pregnancy without support does not weigh most for these women; and clearly the cost of a false negative weighs heavily on women who are explicitly interested in a committed relationship. This Primary indication of learning makes it a good bet that the women learned the bias.

### Signal detection theory

This paper advances a model of the process by which we bias productions of either belief or doubt based of what we have learned, in order to try to minimize the cost of errors (“Learning Model”).



The foundation of the model is the same as that of traditional EMT, viz., signal detection theory (SDT; Green & Swets, 1966). This section overviews one of SDT's key principles and how it is differentially reflected in traditional EMT and the Learning Model.

SDT has been used to explain a wide range of phenomena. Most relevant here, is its use in explaining how particular species tend to minimize the cost of false positives and false negatives for environmental conditions that trigger behavioural responses. The theory identifies specific factors that determine an optimal threshold level of reliability for sensory data that must be met before a particular response can be triggered. One such factor, for example, is the benefit of avoiding a response in the absence of its associated triggering condition. The factors explain such phenomena as the level of reliability of auditory data that triggers female birds to behave as if mating signals are being produced by a high quality potential mate (Price, 2013), and explain the combination of visual and olfactory data that triggers bees to behave as if they are sensing a rewarding floral display (Leonard, Dornhaus, & Papaj, 2011).

The asymmetry in error costs to which EMT proponents refer is composed of some of these same factors; and the proponents explain that when pronounced asymmetries persist over an evolutionary period, they can lead to the evolution of a propensity “to adopt one belief on the basis of more slender evidence than would be required to believe in an alternative” (Haselton & Nettle, 2006, p. 47). So, consistent with SDT, asymmetries, composed of some of the factors, may ultimately determine a standard of evidence that tends to minimize overall error costs. In contrast, the Learning Model suggests that these factors characterize types of information that we pay attention to and internalize; and this information in turn influences the standard of evidence for belief that we set for ourselves. (The process of setting standards is mediated subconsciously. Consequently, we are unaware of the process and look for explanations for our biases.)

The Learning Model allows that evidentiary standards may be determined in this way in relation to most any type of environmental condition. So, its explanatory reach may extend generally to any case *in which the evidentiary standard for belief seems to onlookers to be too high or too low, or seems to have modulated*. Still, as suggested above, the Learning Model does not extend to conditions that are clearly perceptible to the senses, and there may be additional limitations, though this research has not yet discovered any.

*A priority of this paper is to reveal the Learning Model's explanatory reach.* It does so by illustrating its various assumptions and facets using highly varied cases. The cases that it presents were drawn from different literatures, but particularly from epistemological and social psychological literatures. Cases that receive the most attention here include DeRose's bank cases (Section “The cost of false positives”), Gettier cases (Section “The prior probability of misleading evidence”), cases involving “wishful thinking” (Section “The benefit of true positives”), “positive illusions” (Section “The benefit of true positives”), “just world” cases (Section

“Downward pressure may bias our assessment of the strength of evidence”), cases involving ideological belief (Section “Downward pressure may bias our assessment of the strength of evidence”), a lottery ticket case (Section “Resultant pressure”), and cases used to support a view called “subject sensitive invariantism” (Section “Altruistic knowledge intuitions and doubts”).

### The “just-so” criticism would not be fair

Both traditional EMT and the Learning Model are based on weak inferences—from a plausible selection pressure to the evolution of cognitive modules that have particular functions. As such, the theories are subject to the “just-so” criticism. This is a concern about a lack of direct evidence of the evolutionary process *per se*, which often characterizes adaptationist hypotheses (e.g., lack of direct evidence of specific genetic mechanisms or evolutionary stages). Additionally, neither EMT nor the Learning Model identify any neurological evidence for the existence of the necessary mediating cognitive structures. So, it may be that these theories are nothing more than imaginative explanations.

However, the Learning Model has tremendous explanatory and predictive power, explaining a highly diverse range of cases. (This research has not yet identified a case that is inconsistent with the model.) Moreover, the applicability of the Learning Model to any specific type of case can be falsified with even a single true counterexample of that type, e.g., a Gettier or lottery case. This would be a single true counterexample in which (a) for adequately informed onlookers, the evidentiary standard for belief seems too high or too low, or seems to have modulated, but in which (b) the Primary indication of learning is not evident. In light of this, a just-so criticism of the Learning Model would not be fair.

On the other hand, a just-so criticism of the EMT hypothesis that we *inherit* specific biases in either belief or doubt *would be* fair. This research has not found any proposed case of an inherited bias in either belief or doubt in which the Primary indication of learning is not apparent and for which a more convincing explanation is not suggested by the Learning Model. Moreover, the hypothesis is not meant to explain any cases in which it is implausible that an error-cost-asymmetry persisted to any significant extent over a sufficiently long evolutionary period. So, the explanatory power of this hypothesis is extremely weak. Additionally, the hypothesis is not falsifiable. Even if the Primary indication of learning was evident in every identified bias, and the Learning Model provided an adequate explanation of all of them, the hypothesis would not be falsified—because of the inextinguishable possibility of finding a bias in either belief or doubt that did not show the Primary indication.

### Macro and micro-level theories of individual differences and context-dependence

Many of the widely known successes in evolutionary psychology involve compelling explanations of species-typical



or sexually-differentiated phenotypical features, e.g., differences between men and women in preferred mate characteristics, altruism directed toward kin, the ability to detect cheating. This seems to explain a “widespread misunderstanding among social scientists that evolved psychological adaptations are like reflexes or ‘instincts’ – blind, inflexible, and insensitive to social and environmental circumstances” (Al-Shawaf, Lewis, Wehbe, & Buss, 2019, p. 1); and it seems to have put into sharp relief the need to closely examine individual differences and context-dependent responses.

Evolutionary psychological literature has addressed this need, particularly through macro-level theory. That theory considers average or typical differences and context-dependence responses in relation to cultural or environmental conditions—which helps us understand general trends and patterns. Tooby and Cosmides explain that the “program logic [embedded in universal adaptations] specifies how environmental inputs are operated on to become behavioural outputs” (2015, p. 34). Buss further explains that behavioural outputs result from “environmentally contingent or culturally contingent ... selection for species-typical psychological mechanisms” (2009, p. 363). For example, male generosity tends to increase when perceived as a mating signal (Iredale, van Vugt, & Dunbar, 2008), and children exhibit submissiveness in the presence of high-status individuals (Enright, Alonso, Lee, & Olson, 2020). Additionally, the social environment broadly influences personality traits (Daly, 2021).

In contrast, this paper advances micro-level theory to address individual differences and context-dependent responses. It considers individual cases, and takes into account more specific or detailed factors. Consequently, it can explain individual variations from the general trend. For example, explanations were presented above, not only of a tendency among women to doubt men’s intentions, but also of an absence of the bias among some women. Further on, explanations will be presented not only of a tendency of subjects to harbour overly optimistic self-appraisals, but also of why some individuals demonstrate overly pessimistic self-appraisals (Section “The benefit of true positives”). Moreover, such cases are only a small subset of the cases covered by the research presented here—which encompasses any sort of biased belief or doubt about a condition that is not adequately perceptible to the senses. Thus, this paper significantly expands the range of specific individual differences and context-dependent responses that can be subject to an evolutionary psychological explanation.

Finally, the macro-level explanations referred to above are generally, explicitly or implicitly, based on selection pressures, and clarify why specific adaptations evolved. They do not address how adaptations perform their evolved function. However, as Tadinac observes, “How and why questions are not alternatives; they are complementary, and to advance our understanding of human behaviour, we must tie a proximate mechanism with an ultimate explanation (Tadinac, 2020, p. 379). This paper addresses this need by

describing the computations involved in setting evidentiary standards for belief.<sup>1</sup>

\*

The structure of this paper is as follows: Section “Four assumptions” discusses four basic assumptions that the Learning Model relies on in order to accommodate SDT. Section “Mistaking behavioural biases for biases in belief or doubt” distinguishes between the process of setting evidentiary standards, which is described by the Learning Model, and the process of creating or selecting deliberate behaviours, which is not. Biases arise in both processes, and distinguishing these processes will help us perform Learning Model analyses without mistaking biases that arise in the latter process for those that arise in the former. Section “How to minimize the overall cost of false-positives and false-negatives” develops the Learning Model on the foundation of SDT. Section “Altruistic knowledge intuitions and doubts” applies the Learning Model to situations in which individuals consider the costs of false positives and false negatives, not for themselves, but for others. This aligns with biological theory of altruism, which explains why some organisms help others at their own expense. Section “Conclusion” summarizes the paper and presents a conclusion. Finally, an appendix discusses Ryan McKay and Charles Efferson’s (2010) suggestion that biases in either belief or doubt that tend to minimize the overall biological cost of errors may not exist.

## FOUR ASSUMPTIONS

In order to accommodate SDT, the Learning Model relies on the following four assumptions.

1. Evidentiary standards are modulated by a domain specific device

It seems that our evidentiary standards for belief are controlled by a domain-specific device. The suggestion here is that the range of problems that the device helps us to solve is limited to setting an evidentiary standard for belief about whether an environment condition that is not adequately perceptible to the senses exists; and it does so according to the specific factors identified by SDT and consistent with theory of altruism. Further, it appears that the device is informationally encapsulated, i.e., the flow of information into the device is restricted. While it has the inputs suggested by SDT (e.g., from processes that indicate the benefit of avoiding a response in the absence of an associated triggering condition), modulations of our standards are not dependent on conscious processes. They

<sup>1</sup>For more on the distinction between ultimate and proximate mechanisms see (Scott-Phillips, Dickins, & West, 2011, p. 38). The scientific benefits of addressing both mechanisms are examined in (Zietsch, Sidari, Murphy, Sherlock, & Lee, 2021).





do not require conscious consent and are not dependent on the individual’s conception of knowledge or their commitment to an epistemic norm. So, it does not appear as if the device is penetrable following the acquisition of relevant theory or experience, unlike certain visual input systems (McCauley & Henrich, 2006). Finally, it appears as if the primary output is directed to conscious processes that assess whether the gathered evidence meets the standard that has been set, and that create or select the deliberate behaviours that follow the production of either a belief or doubt (See Section “Mistaking behavioural biases for biases in belief or doubt”). This paper sets aside any speculation about how the processing referred to here is neurologically mediated—particularly whether this involves an anatomically localized device, as Fodor (1983) suggested. The device may be functional, not anatomical, and operationalized in interactions between brain regions (Boyer & Barrett, 2015).

2. Doubt is a unique behaviour-guiding state

Individuals often consciously realize that they can benefit or avoid harm by performing certain behaviours, but that the success of these behaviours is dependent on an environmental condition being met. (Such a condition will be referred to as a “success-enabling” condition or an “SEC”.) Consequently, performing the behaviours may well be dependent on believing that the SEC has been met. For example, Brooke will perform a sequence of behaviours to attain a particular benefit only if they believe that the SEC has been met. (This paper also addresses cases in which an individual’s plan to perform a behaviour in the future is dependent on believing that an SEC will be met by the time the behaviour is performed, even if it is not currently met. The principles discussed apply equally to such cases.)

When individuals doubt whether an SEC has been met, they may choose not to pursue the associated benefit. This paper, however, focuses on very common states of affairs in which individuals who have doubt about an SEC nonetheless perform series of behaviours in pursuit of some or all of the associated benefits. These series of behaviours can usually be distinguished from those that they would have performed if they did not have any doubt about the SEC—reflecting a recognition of the risk that a consequential environmental condition may not have been met. Therefore, doubt is properly viewed as a unique behaviour-guiding state.

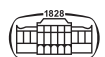
When individuals have doubts about an SEC, they have two main strategies for choosing their behaviours. They can either prioritize avoiding the harm that could result from behaving as if the SEC has been met when it has not, or prioritize gaining the benefit of behaving as if the SEC has been met when it has. The two strategies call for different series of behaviours, which are both nonetheless indicative of the presence of doubt. (Even so, individuals are sometimes not able to find worthwhile specific behaviours to perform in response to the strategy they have chosen.) These are the strategies:

<p><i>S1 (prioritizing the harm that could result from behaving as if the SEC has been met when it has not):</i></p>	<p>Do not behave as if the SEC has been met, in case it has not. Instead, perform a modified or even an entirely different series of behaviours that could produce some or all of the benefits that are at stake. Also, gather additional evidence that might decide whether the SEC has been met.</p>
<p><i>S2 (prioritizing the benefit of behaving as if the SEC has been met when it has):</i></p>	<p>Behave as if the SEC has been met, in case it has; but look for additional behaviours that would mitigate the harm that would follow if the SEC has not been met, i.e., look for a worthwhile hedge. Also, gather additional evidence that might decide whether the SEC has been met.</p>

It would seem that the women who doubt a man’s intention to commit to a familial relationship (Section “Introduction”) would adopt S1. They would prioritize avoiding the harm that could result from behaving as if the SEC (the man in fact intending to commit) has been met when it has not. They are likely to perform a modified series of behaviours that does not include having incautious sex, but may well include maintaining a courteous and engaged manner in case the SEC has been met. At the same time, they may look for new evidence that might decide whether the SEC has been met.

We often adopt S2 when we make financial or economic choices under the shadow of doubt. For example, an SEC for a man buying and living in a house is the house not catching fire. Say that the man is almost certain that this SEC will be met while he owns and lives in the house. Even so, he has a slight doubt. So, he prioritizes the benefit of behaving as if the SEC will be met when it will. He buys and lives in the house, but also hedges by buying fire insurance. Similarly, a woman buys stocks but hedges by buying options, and she holds American dollar assets but hedges by buying foreign assets. She may well also gather additional evidence that might decide whether hedging is worthwhile.

The S2 strategy is also reflected in our presumptions. Edna Ullmann-Margalit (1983) explains that a presumption “comes to the aid of a deliberating agent when ... the choice of the course of action to be performed hinges in a material way on whether a certain state of affairs obtains [i.e., whether an SEC will be met], and when the agent is in a state of ... doubt” (p. 154). The law often calls on us to presume, e.g., to presume innocence or to presume death after an unexplained absence of seven years. An SEC for a juror who presumes the accused’s innocence is the juror actually being innocent. The juror who has doubt about whether the SEC has been met is likely to prioritize the benefit of behaving as if it has when it has: They



will adopt S2, and presume that the accused is innocent as required by the law. Even so, they will hedge by listening with particular care to the prosecutor, by having a particularly positive attitude about the prosecutor's honesty and forthrightness, and by being fully prepared to vote "guilty".

S2 is apparent in "rational acceptance" cases. Consider an example from Michael Bratman (1999).

My close friend has been accused of a terrible crime, the evidence of this guilt is strong, but my friend insists on his innocence ... [My] close friendship may argue for [rationally accepting] that he is innocent of the charge. In making plans for a dinner party, for example, such considerations of loyalty might make it reasonable ... [not to] use this issue to preclude inviting him. (p. 25)

Here, Bratman's friend has been accused of a terrible crime and inviting the friend to a party would provide Bratman the benefit of offering them his loyalty. However, the overall benefit of performing this behaviour depends on whether his friend is in fact innocent, and Bratman has some doubt about that. Nonetheless, Bratman may well adopt S2 and invite him. (Bratman calls the mental state that would precedes that invitation, "rational acceptance".) Even so, Bratman would likely also hedge. For example, if a guest at his party brought up the crime, Bratman might well demur or offer a half-hearted defence. In contradistinction, if he had no doubt at all, he would likely offer a committed and emphatic defence of his friend, and evince moral indignation about the accusation. Moreover, Bratman would also likely look for evidence: Has his close friend said something that suggests either innocence or guilt?

Finally, this account of doubt is broadly consistent with certain neuropsychological findings. It seems that doubt has a special role in executive functioning and decision making, particularly in relation to the basic need to represent social and physical situations. The "process of cognitive representation involves an initial belief, and if there are discrepancies between the initial belief and other mental representation, doubt can be retroactively affixed to this belief" (Asp, Manzel, Koestner, Denburg, & Tranel, 2013, p. 1). Moreover, scientists seem to have located the device that mediates productions of doubt. Damage to the ventromedial prefrontal cortex can produce "credulity and a tendency to believe inaccurate information" (Asp et al., 2012, p. 2), a state referred to as "doubt-deficit".

### 3. Proper analyses of bias cases identify the slightest doubt

Belief may be accompanied by a small doubt (Moon, 2017, pp. 1830–1831). This is reflected in the use of the word "belief". To the question, "Is Nashville the capital of Tennessee?", Blake responded, with a small doubt, "I believe it is" (emphasizing "believe"). Critically, even the slightest doubt can affect the series of behaviours that individuals perform. Courts cannot impose the capital punishment unless an accused is found guilty beyond the shadow of a doubt. So, proper analyses of bias cases identify the slightest doubt that accompanies belief.

In those moments in which an individual believes without any doubt at all, they have *accepted* that they have the truth about an environmental condition being met and that their evidence is fully sufficient. The epistemological literature suggests that we ordinarily flag such moments using the word "know" or one of its cognates. Going back to Blake's response, their friend may respond back, "Do you *believe* it is the capital, or do you *know* that it is? Is there any doubt in your mind at all?" Hawthorne and Stanley (2008) find that consequential behaviours are susceptible to being appraised by others to be improper or inappropriate if associated success-enabling conditions were not believed to be met in the moment without any doubt. In fact, "ordinary folk appraisals of the behaviour of others suggest that the concept of knowledge is intimately intertwined with the rationality of action" (p. 571). Moreover, when there is no doubt at all, belief does not vary in strength. Accordingly, the word "know" does not come in degrees, it is not a "gradable expression" (Stanley, 2005, pp. 35–46). Accordingly, a belief that is not accompanied by any doubt will be referred to as an "intuition of knowing" or a "knowledge intuition" throughout this paper.

*So henceforth, every reference to an evidentiary standard will refer to an evidentiary standard for the production of a knowledge intuition; and the deliberate behaviour that follows from a knowledge intuition that an SEC has been met will be referred to as "behaving-as-if-the-SEC-has-been-met-without-hedging".*

\*

It is evident that we use "know" differently in ordinary practical contexts (which are the focus of this paper) than we do in some epistemological contexts. In the former contexts, the principal concern is whether one would do well to take it that an environmental condition has been met because this will determine the series of behaviours that one will perform. Here, evidentiary standards for the production of a knowledge intuition are not influenced by our conceptions or theories of knowledge. Conversely, in some epistemological contexts, our conceptions of knowledge, knowledge theory or commitments to an epistemic norm come to the fore, e.g., classic invariantists hold the view, roughly, that the standard of evidence for knowledge is fixed across contexts and extremely high. The concern is often whether requirements for knowledge indicated by these conceptions or theories have been met; and, in these contexts, "know" flags that they have.

Consequently, the use of "know" in practical contexts is often inconsistent with its use in epistemological contexts. Many epistemologists have held, incorrectly from the perspective adopted here, that these inconsistencies arise out of what Earl Conee (2005) calls "loose talk", a lax use of "know":

In ordinary contexts ... people will claim knowledge, and attribute knowledge to themselves and others, in belief and in speech ... [However,] if asked whether some proposition to which knowledge is ascribed on some such basis is really



known, or truly known, or really and truly known, fluent speakers have a strong inclination to doubt or deny that it is. (p. 52)

However, Conee's question about whether a proposition is "really known ... truly known ... really and truly known" changes the context for people. They set aside their practical reactions to an ordinary proposition in order to answer a theoretical question, viz., does their evidence meet a standard that is proposed by a particular conception or theory. This is not a scenario in which people use "know" in a lax fashion, but instead one in which the word is used to flag two different states of affairs. This complements Edward Craig's (1990) recognition that we have, on one hand, "reactions to examples both real and imaginary [that lead to] intuitive ascriptions and withholdings of the title of knowledge [while we have, on the other hand,] intuitions about why certain cases do, and others do not, qualify as knowledge" (p. 1).

#### 4. Values and value priorities tend to support inclusive fitness

The evolution of an ability to bias belief and doubt in accordance with SDT is dependent on individuals being able to evaluate the inclusive fitness consequences of false-positives and false-negatives.<sup>2</sup> It is assumed here that we can do this. These evaluations follow from the pursuit of our values and value priorities; and while that pursuit sometimes fails to support inclusive fitness, it broadly tends to do so. This picture is consistent with (a) social psychological literature indicating that values play an important role in determining behaviour (Schwartz, 1992), (b) the high level of fitness of our species indicated by the overarching human population growth, and (c) cross-cultural investigations showing that broad value categories are universal (Schwartz, 1994).

## MISTAKING BEHAVIOURAL BIASES FOR BIASES IN BELIEF OR DOUBT

The Learning Model describes a process that culminates in setting the evidentiary standard for producing a knowledge intuition. It does not address how we create or select the deliberate behaviours that follow, i.e., the behaviours that will constitute behaving-as-if-the-SEC-has-been-met-without-hedging and constitute behaving-under-doubt. However, biases arise in both processes. Clearly distinguishing the two will help us perform Learning Model analyses without mistaking biases that arise in the latter for biases that arise in the former.

Whereas evidentiary standards are ordinarily modulated by an entirely subconscious process, creating or selecting

deliberate behaviours ordinarily involves conscious cognition, regardless of whether the process is facilitated by a behavioural bias or heuristic. Think of Bratman deliberately inviting his friend to the dinner party or deliberately being reserved when he defends his friend.

To the extent that one creates or selects behaviours without the aid of a behavioural bias or heuristic, one will (a) methodically consider alternative behaviours, (b) aim at maximizing expected utility based on value priorities, and (c) rely only on relevant and reliable information. The use of biases and heuristics can make this often overly complicated or time-consuming task more manageable. For example, the availability heuristic can lead us to take it that a particular behaviour is the best choice because it was quickly recalled. The behavioural "herding bias" can lead us to take it that a specific behaviour is the best choice because it was performed by others, and a behavioural "loss aversion bias" can lead us to take it that a specific behaviour is not the best choice because it entails incurring a loss even though an equivalent or greater gain would subsequently arise.

EMT proponents leave themselves open to mistaking a behavioural bias for a bias in either belief or doubt when they identify a bias only on the basis of behaviour. For example, Haselton (2003) claims to identify a bias in belief among men that involves overperceiving women's sexual interest, and does so by looking only at women's reports of men appearing to overperceive this. However, such reports would likely have been primarily based on the men's overt behaviour. The aim is to avoid such a mistake by looking for signs that the subject explicitly questions and assesses an SEC, and signs that the behaviours that they perform depend of that assessment. Has the man said that he knows that the woman is interested? Is it clear that the man carefully considered the fact that the woman very briefly smiled at him? If answers to questions like these are in the negative, then it should not be concluded that he has overperceived her sexual interest; and it should be allowed that his behaviours may be a consequence of a behavioural bias.

(This section responds to a 2010 paper by Ryan McKay and Charles Efferson. An [appendix](#) looks more closely at it.)

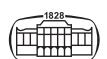
## HOW TO MINIMIZE THE OVERALL COST OF FALSE-POSITIVES AND FALSE-NEGATIVES

Signal detection theory is based in large part on the principle that the cost of false positives should put proportional upward pressure on standards of evidence, while the cost of false negatives should simultaneously put proportional downward pressure on them. This section discusses these pressures separately, and concludes by discussing their interaction.

### Upward pressure

**The cost of false positives.** Individuals can decrease the likelihood of incurring the cost of a false positive by raising their evidentiary standard, e.g., the woman can decrease the

<sup>2</sup>The term 'inclusive fitness' refers to an organism's ability to pass its genes on to subsequent generations, together with its ability to help its relatives (aside from offspring) pass down their genes.





likelihood of incurring the cost of a pregnancy without support by raising her evidentiary standard for taking it that the man intends to commit to a familial relationship. Accordingly, SDT suggests that as the anticipated cost of a false positive increases, so too should upward pressure on the associated evidentiary standard. This notion is expressed in terms of values and behaviours in the following prediction: The extent to which we disvalue the consequences of behaving-as-if-the-SEC-has-been-met-without-hedging when it has not corresponds with the extent of the upward pressure on the associated evidentiary standard (“Upward Pressure”). Additionally, SDT suggests that the extent to which we disvalue the consequences of behaving-as-if-the-SEC-has-been-met-without-hedging when it has not corresponds with the extent to which those consequences are anticipated to be worse than what the consequences would have been if instead we behaved-with-doubt (see Section “Doubt is a unique behaviour-guiding state” and Section “Proper analyses of bias cases identify the slightest doubt”). So finally, *Upward Pressure correlates with the anticipated extent to which we would disvalue the consequences of behaving-as-if-the-SEC-has-been-met-without-hedging when it has not more than we would the consequences of behaving-under-doubt (when the SEC has not been met). This will be referred to as the ‘Cost-of-a-False-Positive’.*

This prediction explains why a knowledge intuition is produced in Keith DeRose’s (1992) Bank Case A and not in his Bank Case B.<sup>3</sup>

*Bank Case A.* My wife and I are driving home on a Friday afternoon. We plan to stop at the bank on the way home to deposit our paycheques. But as we drive past the bank, we notice that the lines inside are very long ... Although we generally like to deposit our paycheques as soon as possible, it is not especially important in this case that they be deposited right away, so I suggest that we drive straight home and deposit our paycheques on Saturday morning. My wife says, “Maybe the bank won’t be open tomorrow. Lots of banks are closed on Saturdays.” I reply, “No, I know it’ll be open. I was just there two weeks ago on Saturday ...”

*Bank Case B.* My wife and I drive past the bank on a Friday afternoon, as in Case A, and notice the long lines. I again suggest that we deposit our paycheques on Saturday morning, explaining that I was at the bank on Saturday morning only two weeks ago and discovered that it was open until noon. But in this case, we have just written a very large and very important cheque. If our paycheques are not deposited into our chequing account before Monday morning, the important cheque we wrote will bounce, leaving us in a very bad situation. And, of course, the bank is not open on Sunday. My wife reminds me of these facts. She then says, “Banks do change their hours. Do you know the bank will be open tomorrow?” Remaining as confident as I was before

that the bank will be open then, still, I reply, “Well, no. I’d better go in and make sure.” (1992, p. 913)

In both cases, DeRose assesses whether his success-enabling condition, which is that the bank will be open on Saturday, will be met; and he does so on the basis of his recollection of the bank being open on a Saturday two weeks earlier. Yet, while DeRose takes himself as knowing in the first case, he has some doubt in the second. This is a result of the Cost-of-a-False-Positive being much higher in the second case than it is in the first, which in turn produces higher Upward Pressure on DeRose’s evidentiary standard. While his evidence is strong enough to meet the standard in Case A, it is not in Case B. Notice too, the Primary indication of learning: DeRose plausibly anticipated a significant difference between the Costs-of-a-False-Positive in the two cases as a result of what he learned about the consequences of having an important cheque fail to transact due to insufficient funds, and his value priorities. Moreover, some individuals may well not have demonstrated DeRose’s bias. They would plausibly have not anticipated as great a difference between the costs in the two cases—because of their specific experiences, cultural learning, life circumstances, and value priorities. They may have taken themselves as knowing in both cases.

**A more detailed look.** The first step in getting a sense of what the Cost-of-a-False-Positive would have been in these cases is to identify the series of behaviours that would have constituted behaving-as-if-the-SEC-will-be-met-without-hedging and behaving-under-doubt.

DeRose describes what behaving-as-if-the-SEC-will-be-met-without-hedging consists of in Case A: He will go straight home and return to the bank on Saturday. Behaving-under-doubt in Case A would likely have been based on S2. This is because the benefit of behaving as if the SEC will be met when it will (the benefit of avoiding the long lines and getting home early on the Friday afternoon) seems to be more important to DeRose than is the harm that could result from behaving as if the SEC will be met when it will not (the harm that could result from returning to the bank on Saturday only to find that it is closed). So, he would have prioritized the benefit of behaving as if the SEC will be met; and behaving-under-doubt would have involved going straight home and returning Saturday. However, it seems that DeRose would not have had a worthwhile hedge against the harm he would face if he behaved as if the SEC would be met and it was not. It seems that the only harm here would have been returning to the bank Saturday only to find it closed; and DeRose had the opportunity to eliminate that harm by going into the bank on Friday to enquire about the Saturday hours (like he did in Case B), but chose not to. So, it seems that DeRose disvalued the inconvenience of checking the hours on Friday (the only available hedge) more than the harm of returning to the bank Saturday to find it closed.

The foregoing suggests that the Cost-of-a-False-Positive for DeRose in Case A would have been inappreciably low. If DeRose did not have a worthwhile hedge, then there would have been no difference between behaving-as-if-the-SEC-will-

<sup>3</sup>There are several other similar paired cases in the epistemological literature, e.g., Stewart Cohen (1999, p. 58), Jeremy Fantl and Matthew McGrath (2002, pp. 67–68), and Chandra Sripada and Jason Stanley (2012, pp. 11–12).





be-met-without-hedging and behaving-under-doubt (under S2). Whether he doubted or not, he would have gone straight home and returned Saturday. So, if his SEC was not going to be met, he could not have disvalued the consequences of the former more than the consequences of the latter; and there would not have been any Upward Pressure on his evidentiary standard for taking it that the SEC would be met. He had an unelevated standard that his evidence was able to meet.

Behaving-as-if-the-SEC-will-be-met-without-hedging in Case B would have been the same as it was in Case A. He would have gone straight home and returned Saturday. DeRose tells us, in Case B, what behaving-under-doubt consists of. He will “go in and make sure” that the bank will be open on Saturday. This is consistent with S1, which he chose because the level of harm that could result from behaving as if the SEC will be met when it will not is very high. DeRose will not behave as if the SEC will be met unless and until he has gathered additional evidence that confirms that the bank will be open on Saturday.

This suggests that the Cost-of-a-False-Positive in Case B is high. If the SEC will not be met (and the bank will be closed on Saturday), DeRose would have disvalued the consequences of behaving-as-if-the-SEC-will-be-met-without-hedging (going straight home and returning Saturday to find the bank closed) much more than behaving-under-doubt (going into the bank and learning that it will be closed). As a result, there is substantial Upward Pressure that raises DeRose’s evidentiary standard to a level that his evidence does not meet.

**The prior probability of misleading evidence.** SDT recognizes that the predictive value of inductive evidence can be dependent on a range of environmental conditions. For example, the predictive value of DeRose seeing the bank to be open two weeks prior on a Saturday, as evidence that it will be open the coming Saturday, is dependent on the likelihood that something could cause the bank to change its operating hours, e.g., a currency shortage, power outage, labour disruption, and so on. If that likelihood is great, then it will be likely too that that evidence is misleading. Accordingly, evidentiary standards need to be calibrated to take this into account. SDT suggests that the prior probability of the inductive evidence being misleading,  $PP(E_{\text{misleading}})$ , is an additional factor in upward pressure on evidentiary standards; and the Learning Model holds that the individual’s sense of  $PP(E_{\text{misleading}})$  is a factor in Upper Pressure.<sup>4 5</sup>

<sup>4</sup>“ $PP(E_{\text{misleading}})$ ” tokens the same concept as the SDT term “prior probability of noise”.

<sup>5</sup>The thesis that evidentiary standards correlate with the individual’s sense of a prior probability is inconsistent with social psychology experiments that reveal that we are *not* good at reasoning with probabilities. However, Jason S. McCarley and Aaron Benjamin (2013) find that we *are* good at reasoning with natural frequencies. Roughly, a natural frequency is the number of occurrences of a state or event in a meaningful unit of time or out of a meaningful set of occurrences. For example, the natural frequency with which the dog’s barking disturbs the neighbours is three or four times a day. Accordingly, it is plausible that our sense of  $PP(E_{\text{misleading}})$  in cases is provided by a sense of a natural frequency.

DeRose takes this additional factor into account when he reflects on his Bank Case A, “If very many nearby banks have discontinued their Saturday hours in the last two weeks, then it seems that my original claim to know may well have been false” (1992, p. 921). Here, new indications of a likelihood that the bank has discontinued Saturday hours causes DeRose to have the sense that  $PP(E_{\text{misleading}})$  is higher than he first thought. This increases Upward Pressure and raises his standard to a level that his evidence cannot meet.

However, the likelihood that evidence is misleading is dependent, not only on the likelihood of an environmental disturbance *making* it misleading (e.g., a power outage), but also on the prevalence of intrinsically misleading evidence. For example, say that the evidence for the success-enabling condition, the object being an X, consists of seeing an object that looks like an X; and say that there are both X’s and fake X’s in the environment (there are non-X objects that look like X’s). If fake X’s are prevalent, then  $PP(E_{\text{misleading}})$  will be high, and both SDT and the Learning Model will suggest that the evidentiary standard should be raised.

This state of affairs explains how we respond to “Gettier cases” (which garnered tremendous interest among epistemologists in the latter half of the last century). Consider Alvin Goldman’s (1976) fake barns case. Unbeknownst to Henry, there are many excellent papier-mâché facsimiles of barns in the area he is driving through. He happens to see one of the only real barns in the area. The sight of the structure is sufficient to cause Henry to take himself as knowing that the structure is a barn. However, readers of this case generally deny that Henry knows that the structure is a barn, even though they recognize that Henry has justified true belief for the SEC. The Learning Model explains that readers have the sense, unlike Henry, that  $PP(E_{\text{misleading}})$  is high—due to a high prevalence of fake barns. This increases readers’ Upward Pressure and raises their evidentiary standard to a level that Henry’s evidence cannot meet.<sup>6</sup>

(Still, the reader of the fake barns case knows that the structure is in fact a barn, and consequently that Henry would do well to behave as if it is one, e.g., he may do well to inquire about stabling his horse there. This reminds us that a trait needs only to *tend* to promote inclusive fitness in order for selection pressure to be produced. A trait may very well not be helpful under a range of circumstances. The trait at issue here—not producing a knowledge intuition unless the evidence meets a standard that is subject to certain Upward and Downward Pressures—plausibly tends to be advantageous even if it may not be advantageous to readers of Gettier cases.)

\*

<sup>6</sup>Like Gettier cases, ordinary second and third-person knowledge attributions and denials (e.g., Brooke knows that the SEC has been met) seem generally to be based, not on the subject’s sense of the  $PP(E_{\text{misleading}})$  (e.g., not on Brooke’s sense) but on the speaker’s—except when the speaker is convinced that the subject has better information than they do.



In summary, under the Learning Model, total Upward Pressure is the notional product of the two sources of pressure described above: Cost-of-a-False-Positive  $\times$   $PP(E_{\text{misleading}})$ .<sup>7</sup>

### Downward pressure

Let us turn our attention to the cost of false negatives (assuming that a condition does not exist when it does). Unfortunately, detailed discussions of these costs are inherently hard to follow because they inevitably require double or triple negations. The following might have been written: “Brooke increases the likelihood of not incurring the cost of not taking it that the lump could be cancerous when it is so by lowering the relevant evidentiary standard.” There is a way to avoid this. The costs that arise in these cases *are the costs of failing to benefit from true positives* (failing to benefit from assuming that a condition exists when it does). When a biased knowledge intuition or doubt is produced, the magnitude of the cost of a false negative is the same as that of the benefit of a true positive, and they have the same effect on Downward Pressure. Brooke’s cost of not accepting that the lump could be cancerous when it could be so has the same magnitude as the benefit of a true positive for the lump possibly being cancerous, viz., the benefit of taking it that the lump could be cancerous when it is so and getting early medical attention. Discussions of these benefits do not require multiple negations. So, the orientation here will be changed towards the benefit of true positives.

**The benefit of true positives.** Individuals can increase the likelihood of benefiting from a true positive by lowering their evidentiary standard, e.g., Brooke can increase the likelihood of benefiting from taking it that the lump could be cancerous when it is so (the benefit of getting early medical attention) by lowering their evidentiary standard for taking it that the lump could be cancerous. Accordingly, SDT suggests that as the anticipated benefit of a true positive increases, so too should downward pressure on the associated evidentiary standard. This notion is expressed in terms of values and behaviours in the following prediction: The extent to which we value the consequences of behaving-as-if-the-SEC-has-been-met-without-hedging when it has corresponds with the extent of the downward pressure on the associated evidentiary standard (“Downward Pressure”). Additionally, SDT suggests that the extent to which we value the consequences of behaving-as-if-the-SEC-has-been-met-without-hedging when it has corresponds with the extent to which those consequences are anticipated to be better than what the consequences would have been, if instead we behaved-with-doubt when the SEC has been met (see Section “Doubt is a unique behaviour-guiding state” and Section “Proper analyses of bias cases identify the slightest doubt”). So finally, *Downward Pressure correlates with the*

*anticipated extent to which we would value the consequences of behaving-as-if-the-SEC-has-been-met-without-hedging when it has more than we would the consequences of behaving-under-doubt (when the SEC has been met). This will be referred to as the ‘Benefit-of-a-True-Positive’.*

This component of the Learning Model is consistent with Dion Scott-Kakures (2000) theory of motivated believing. The author explains that “the goal of ... believing is [the] realization of the agent’s goals and values” (p. 362); and, more particularly, that “familiar cases of wishful believing ... can be understood as motivated ... by reflection upon the cost incurred by failing to believe the relevant proposition” (p. 366). In one of Scott-Kakures’ cases, a parent takes it that their daughter is a talented dancer based on weak evidence. Scott-Kakures explains that the parent’s poorly supported belief motivates them to help the child “excel in dance—through encouragement ...” (p. 366), and that this helps the parent promote their values. The Learning Model explanation is more detailed. First, the parent has recognized that helping their child would promote their values, if the child is in fact talented. The SEC is the child being talented, and behaving-as-if-the-SEC-has-been-met-without-hedging consists of helping the child, through encouragement, etc. Second, the Benefit-of-a-True-Positive for the SEC is substantial. It could result in the child becoming a professional dancer. Finally, this creates significant Downward Pressure on the evidentiary standard for taking it that the SEC has been met, which in turn lowers that standard to a level that even weak evidence can meet.

Scott-Kakures also discusses cases in which “terminal cancer patients, and those near to them ... to the very end, believe that a recovery is possible, that a treatment might be effective” (p. 366). The author explains again that such a belief could lead to helpful behaviours (such as getting the patient into a clinical trial) and that those behaviours may help the individual realize their goal of having the patient survive. According to the Learning Model, these individuals have recognized that behaving as if the patient has a chance of surviving would promote their values, if in fact the patient does have such a chance. The SEC is the patient having a chance of surviving, and behaving-as-if-the-SEC-has-been-met-without-hedging might involve trying to get the patient into a clinical trial. Finally, the Benefit-of-a-True-Positive for the SEC is substantial. It could save the patient. This creates significant Downward Pressure on the evidentiary standard for taking it that the SEC has been met, which in turn lowers that standard to a level that even weak evidence can meet.

**Positive illusions.** Social psychological studies have shown that subjects tend to be overly optimistic about their abilities (Taylor & Brown, 1988) and their positive qualities (Alicke, 1985). Researchers refer to instantiations of this tendency as “positive illusions”. EMT proponents suggest that positive illusions are produced by an inherited bias:

trying and failing may not matter very much, whereas failing to try could be very costly, especially in competitive contexts. Thus, evolution can be expected to produce mechanisms biased toward positive illusion in domains where there is

<sup>7</sup>The hypothesis that the simultaneous effect of these two factors is appropriately represented as a product is inherent to the mathematics of SDT. See for example Godfrey-Smith (1991, 1998, pp. 232–254).



uncertainty about outcomes, and the cost of trying and failing is reliably less than that of not trying where success was possible. (Haselton et al., 2015, pp. 980–981)

However, the Primary indication of learning is evident. Those who display this bias have plausibly learned that a positive self-appraisal in a particular domain can lead to trying and succeeding. This anticipated Benefit-of-a-True-Positive would produce Downward Pressure on the evidentiary standard for taking it that the positive self-appraisal is correct. Moreover, many of us are prone to being, not overly optimistic about our abilities, but overly pessimistic, prone to having negative illusions—at least in certain domains. This may motivate one to devote “considerable energy to mentally playing through or reflecting on all the possible outcomes [that they] can imagine for a given situation” (Norem & Chang, 2002, p. 996). For example, surgeons and trial lawyers may repeatedly rehearse in preparation for important surgeries and trials respectively. Consequently, negative illusions tend to cause these individuals to overprepare and perform very well (Norem & Illingworth, 1993; Spencer & Norem, 2016); while optimistic self-appraisals may sabotage their preparation process. So, it seems that, for these individuals, it is not a Benefit-of-a-True-Positive (e.g., the benefit of taking oneself as knowing that a positive self-appraisal is correct when it is) that is prominent, but rather a Cost-of-a-False-Positive (e.g., the cost of taking oneself as knowing that a positive self-appraisal is correct when it is) that is prominent.

**The prior probability of reliable evidence.** SDT identifies a second source of Downward Pressure. It is the need to calibrate evidentiary standards so that they take into account *the prior probability that the inductive evidence is reliable*,  $PP(E_{\text{reliable}})$ . (This mirrors the discussion about  $PP(E_{\text{misleading}})$  in Section “The prior probability of misleading evidence”) Consider again the case of the young dancer’s parent (Section “The benefit of true positives”). Say that the parent’s evidence for the child being talented was repeated praise from teachers and other parents; and say that the parent later discovers that teachers and parents often praise untalented dancers. The parent gets the sense that reliable evaluations of young dancers are not prevalent, i.e., that the  $PP(E_{\text{reliable}})$  is markedly lower than previously thought. This will cause Downward Pressure to drop and consequently the evidentiary standard to rise. The parent may well begin to doubt that their daughter is as talented as they thought.<sup>8</sup>

\*

In summary, under the Learning Model, total Downward Pressure is the notional product of the two sources of pressure described above: Benefit-of-a-True-Positive X  $PP(E_{\text{reliable}})$ .

<sup>8</sup> $PP(E_{\text{reliable}})$  tokens the same concept as the SDT term “prior probability of signal”.

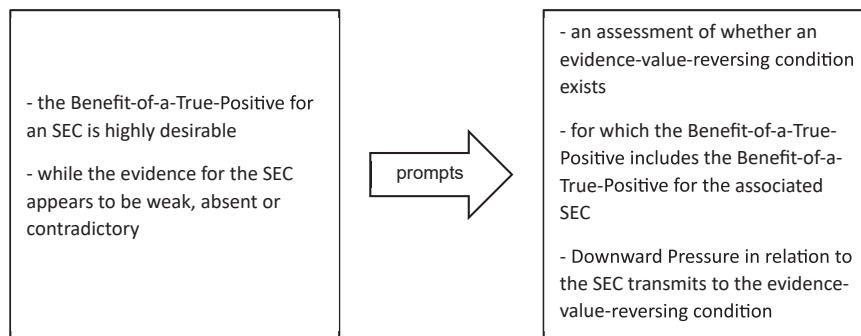
**Downward pressure may bias our assessment of the strength of evidence.** Knowledge intuitions can be produced, not only when the affirmative evidence is weak, but also when there is an absence of good affirmative evidence or even when there is substantial contradictory evidence. In every such case included in the present research, it seems that the Benefit-of-a-True-Positive for the success-enabling condition is highly desirable, and that this prompts the individual to assess whether an entirely different environmental condition is distorting the true evidentiary support for the SEC. Such a condition would either make what is in fact acceptable affirmative evidence appear to be faulty, or make what is in fact faulty contradictory evidence appear to be acceptable. (Call these “evidence-value-reversing” conditions.) Further, the Benefit-of-a-True-Positive for an evidence-value-reversing condition consists primarily of increasing the likelihood that the evidence for the SEC will meet its standard—which increases the likelihood of gaining the Benefit-of-a-True-Positive for the SEC. Consequently, the Benefit-of-a-True-Positive for the evidence-value-reversing condition is taken as including the Benefit-of-a-True-Positive for the SEC, which is highly desirable. This state of affairs creates Downward Pressure on the standard for the evidence-value-reversing condition such that even weak evidence may be sufficient to take it that the evidence-value-reversing condition exists. In effect, the Downward Pressure on the evidentiary standard for taking it that the SEC has been met transmits through to the evidentiary standard for taking it that the evidence-value-reversing condition exists. This dynamic is captured in Fig. 1. Let us consider two examples.

**Just world belief.** There is a widely-held belief that we live in a “just world”—a world in which ‘people generally get what they deserve’ (Lerner & Miller, 1978, p. 1030). However, the “evidence for the importance of the theme of justice in our society can be strikingly juxtaposed against the equally vivid signs of institutionalized injustice and widespread indifference to the fate of innocent victims” (Lerner, Miller, & Holmes, 1976, p. 134). A learned bias may well cause many of us to find these vivid signs, this contradictory evidence, to be faulty.

Living in a just world would be a success-enabling condition for individuals who are able and inclined to perform behaviours that a just world would reward and to avoid behaviours that a just world would punish. Such behaviours would constitute behaving-as-if-the-SEC-has-been-met-without-hedging. So, the desirability of the Benefit-of-a-True-Positive for the SEC correlates with an individual’s capacity to perform such behaviours; and Downward Pressure on the standard for taking it that the SEC has been met in turn correlates with these capacities.

When the Benefit-of-a-True-Positive is highly desirable, individuals are prompted to assess an evidence-value-reversing condition, viz., a condition that would cause what is in fact faulty evidence to appear acceptable. For example, individuals may assess whether in fact sins of the victim have been obscured:





**Fig. 1.** Cases in which knowledge intuitions are produced though the evidence is weak, absent or contrary

[Perhaps] no injustice has occurred, after all [insofar as] the victim had done something for which he merited his fate ...

[Perhaps] the seemingly innocent victim was the kind of person for whom this suffering was not an inappropriate fate [and perhaps] the victim was a less than desirable person, at least in some respects—selfish, unintelligent, crude, etc. (Lerner et al., 1976, p. 138)

Great Downward Pressure in relation to the SEC effectively transmits to such evidence-value-reversing conditions, and may well cause individuals to take themselves as knowing that the latter conditions exist—even with poor affirmative evidence. This would in turn allow them to dismiss acceptable contradictory evidence in relation to the SEC.

**Ideological belief.** Ideologies give life-purpose to their adherents:

An image of the world becomes an ideology if it creates in the mind of the person holding it a role for himself which he values highly ... To create a role, however, an ideology must create a drama. The first essential characteristic of an ideology is then an interpretation of history sufficiently dramatic and convincing so that the individual feels that he can identify with it and which in turn can give the individual a role in the drama it portrays. (Boulding, 1964, pp. 161–162)

So, the Benefit-of-True-Positives for these images of the world that are foundational to an ideology includes the promise of being involved in an historic enterprise and living a life with great purpose. Here, the Italian nationalist Giuseppe Mazzini suggests how deeply some individuals might value this:

Life is a mission. Every other definition of life is false, and leads all who accept it astray. Religion, science, philosophy, though still at variance upon many points, all agree in this, that every existence is an aim ... Young brothers, when once you have conceived and determined your mission without your soul, let naught arrest your steps. (1946, pp. 111–13)

However, adherents frequently encounter evidence that contradicts their ideology; and they frequently take it that evidence-value-reversing conditions are at play. For example, genocide deniers find faults in “the motives of witnesses, ...

[they pounce] on errors or exaggerations in the arguments of their opponents, [quibble] over terminology, and so on” (Goertzel, 1992, p. 323). In fact, adherents may accept the existence of evidence-value-reversing conditions even if doing so undermines deeply held principles. For example,

the opponents of Galileo maintained that his cosmological theory might be ‘mathematically true’ but that it was ‘philosophically false’ [i.e., it failed to capture a deeper and truer reality] ... Nazi ideologists [held] that the ‘abstract’ mathematical laws of the theory of relativity lacked the ‘concrete truth-’ possessed by German science ...

The ideological conception of truth thus becomes a doctrine of ‘anti-truth’. By one means or another, it tries to overcome or weaken the scientific conception of reality, and to cause the individual to lose confidence in his own perceptions, hypotheses, and judgements. (Feuer, 2010, p. 106)

Again, the Benefit-of-True-Positives for the SECs that underlay an ideology produces substantial Downward Pressure that effectively transmits to evidence-value-reversing conditions; and this allows adherents to take themselves as knowing that the evidence-value-reversing conditions exist—even with exceptionally weak evidence. Having done so, they can dismiss acceptable evidence that contradicts the foundations of their ideology.

Finally, notice the Primary indication of learning. Ideologues plausibly learned or discovered a significant Benefit-of-a-True-Positive in adhering to an ideology and they value that benefit. Moreover, whatever the ideology, there will be many who do not adhere to it, and these people would plausibly not have found the Benefit-of-a-True-Positive to be as significant, as a result of their specific experiences, cultural learning, life circumstances, and value priorities.

## Resultant pressure

The ratio of Upward Pressure to Downward Pressure represents the simultaneous effect of the forces.<sup>9</sup>

<sup>9</sup>The hypothesis that these simultaneous forces are appropriately represented as a ratio is inherent to the mathematics of SDT. See for example Godfrey-Smith (1991, 1998, pp. 232–254).





$$\text{Resultant Pressure} : \frac{\text{Upward Pressure}}{\text{Downward Pressure}}$$

$$\frac{(\text{Cost-of-a-False-Positive}) \times \text{PP}(E_{\text{misleading}})}{(\text{Benefit-of-a-True-Positive}) \times \text{PP}(E_{\text{reliable}})}$$

We may view a Resultant Pressure as an evidentiary standard multiplier that operates on an initial standard of evidence. (This research sets aside consideration of the level of the initial standard and its aetiology.) For example, when Upward and Downward Pressures are in balance, the evidentiary standard is not modulated, either up or down, and the standard does not seem to be either too high or too low. DeRose’s low-stakes Bank Case A (Section “The cost of false positives”) which presents the success-enabling condition of the bank being open on Saturday can serve as an example. As discussed, the Upward Pressure on DeRose’s evidentiary standard would have been inappreciably low. For similar reasons, the Downward Pressure would also have been inappreciably low.<sup>10</sup> The two pressures would have been in balance, and as a consequence DeRose would not have modulated an initial evidentiary standard for taking it that the SEC will be met. Conversely, when there is a great imbalance between Upward and Downward Pressures, evidentiary standards may be extremely high or extremely low. Consider the much-discussed lottery case from the epistemological literature.

Say that Alex addresses the possibility that their recently purchased lottery ticket will lose. As will become apparent momentarily, the lottery ticket losing constitutes an SEC for Alex. Their evidence for the SEC consists only of reports of the total number of tickets in the lottery, *n*. That number is so large that commentators round it to the nearest hundred thousand. Alex realizes that there is a soaringly high probability that the SEC will be met, viz.,  $(n-1)/n$ . What is puzzling is that, like most lottery ticket holders, Alex has some doubts about whether the SEC will be met. They might just win. Apparently, they have set an absurdly high evidentiary standard for taking it that the SEC will be met. Consider the Learning Model explanation.

To start,  $\text{PP}(E_{\text{misleading}})$  and  $\text{PP}(E_{\text{reliable}})$  are *not* factors in such cases simply because prior probabilities are not attributable to purely probabilistic evidence, such as reports of the total number of tickets in a lottery. So, the Resultant Pressure in lottery cases reduces to the following:

$$\text{Lottery Resultant Pressure} : \frac{\text{Cost-of-a-False-Positive}}{\text{Benefit-of-a-True-Positive}}$$

Further, behaving-as-if-the-SEC-will-be-met-without-hedging (as if the ticket will lose) would have consisted of throwing away the ticket, while behaving-under-doubt (behaving as if it may or may not lose) consists of keeping

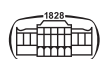
the ticket in a safe place and waiting for the winner to be announced. Accordingly, the Cost-of-a-False-Positive, the numerator of the pressure ratio, would have correlated with the extent to which Alex disvalued throwing away a ticket that would turn out to be a winner (behaving-as-if-the-SEC-will-be-met-without-hedging when it will not be) more than keeping the ticket and later discovering that it won (behaving-under-doubt when the SEC will not be met). That would have been devastating. The denominator, the Benefit-of-a-True-Positive, would have correlated with the extent to which Alex valued throwing away the ticket that would turn out to be a loser (behaving-as-if-the-SEC-will-be-met-without-hedging when it will be) more than keeping the ticket and later discovering that it lost (behaving-under-doubt when the SEC will be met). Alex does not value the former more the latter at all, and may even disvalue it. So, the denominator for Alex would have been inappreciably low. Thus, the Resultant Pressure, along with Alex’s evidentiary standard, would have been soaringly high.

## ALTRUISTIC KNOWLEDGE INTUITIONS AND DOUBTS

Evolutionary biologists have identified specific types of altruistic behaviour that would in theory tend to promote inclusive fitness and that are in fact practised (Delton, Krasnow, Cosmides, & Tooby, 2011; Nowak & Sigmund, 2005; West-Eberhard, 1975, 1983). Two of these are basic and uncontroversial. According to kin-selection theory, we would do well to incur a biological cost in the process of helping kin—provided that the maximum cost is proportional to (a) the degree of relatedness between benefactor and recipient, and (b) the anticipated biological benefit for the recipient. According to direct reciprocity theory, we may do well to help those who are reasonably likely to reciprocate. In paradigm cases, these interactions promote the inclusive fitness of both interactants. Benefactors seem to adopt the motto, “I’ll scratch your back, you later scratch mine”.

These findings suggest that the explanatory reach of the Learning Model can be extended. Let us assume that it sometimes occurs to us that we may be able to help someone by performing a certain series of behaviours, but that success in producing a benefit for the prospective recipient is dependent on an SEC being met. Consequently, performing the series of behaviours is dependent on whether we take ourselves as knowing that the SEC has been or will be met. Further, let us assume that we are concerned about the cost of false positive and false negative errors in relation to the SEC under these circumstances, as we ordinarily are. However, here, the concern would largely relate to the cost incurred by prospective recipients. The Learning Model assumes that, under these circumstances, we modulate our evidentiary standard for taking it that the SEC has been met in order to minimize these overall error costs, which are largely incurred by recipients.

<sup>10</sup>If there was no available hedge, then there would have been no difference between behaving-as-if-the-SEC-has-been-met-without-hedging and behaving-under-doubt (under S2). They both would have involved going straight home. So, DeRose could not have valued the consequences of the former more than the consequences of the latter; and both the Benefit-of-a-True-Positive and Downward Pressure would have been inappreciably low.



So, benefactors evaluate the Cost-of-a-False-Positive and the Benefit-of-a-True-Positive, not for themselves, but for their recipients. This suggests a reconciliation of two well-known positions in epistemology, which are standardly viewed as being incompatible competitors. In fact, each aligns with a different component of the Learning Model and their core tenets are complementary.

*Epistemic contextualism* deals with the evidentiary standard that we apply when we judge the truth of ordinary knowledge attributions and knowledge denials, e.g., the standard that we apply when we judge whether “I know that p”, or “you do not know that p” is true. The view holds that evidentiary standards vary across contexts in accordance with the speaker’s practical interests. Keith DeRose argues for this view, and uses the bank cases (discussed in Section “Upward pressure”) to support his position. In these cases, the speaker’s egoistic evaluations of costs and benefits affect his standard of evidence.

*Subject sensitive invariantism* is also concerned about the evidentiary standard that is applied to ordinary knowledge attributions and denials, and also holds that it varies across contexts. However, its proponents argue that standards vary in accordance with the practical interests, not of the speaker of the attributions and denials, but rather of the *subjects* (Fantl & McGrath, 2002; Stanley, 2005). On this view, the standard that someone applies when they tell Mary, “you know that p” is affected by their sense of Mary’s practical interests.

The Learning Model explains that the basic tenets of subject sensitive invariantism hold only if the speaker has altruistically assessed the Cost-of-a-False-Positive and the Benefit-of-a-True-Positive in relation to the subject of attributions and denials. (However, altruistic evaluations cannot underlay first-person statements, such as “I know that p”, because the subject is the speaker. So, the tenets hold in relation only to second-person and third-person attributions and denials.) Say that in both the bank cases an older brother B of DeRose is in the car, along for the ride, and concerned in the moment only that things go well for DeRose; and in both cases, when B sees the long lines at the bank, it strikes him that it might be helpful to advise his brother to go straight home and return to the bank on Saturday. However, the benefit of doing so in both cases is dependent on whether the bank will be open on Saturday. The SEC is the bank being open Saturday. Further, factors that determine B’s evidentiary standard for taking it that the SEC will be met include his evaluation of the Cost-of-a-False-Positive and the Benefit-of-a-True-Positive, not for himself, but for his brother. Consequently, B has a knowledge intuition that the SEC will be met in the low-stakes case, but not in the high-stakes case.

So, epistemic contextualism and subject sensitive invariantism are complementary in that while the former addresses cases that involve egoistic evaluations of the Cost-of-a-False-Positive and the Benefit-of-a-True-Positive, the latter addresses cases that involve altruistic evaluations of these.

## CONCLUSION

The essence of the Learning Model is that evidentiary standards for taking oneself as knowing are subject to simultaneous pressures, consistent with signal detection theory. The resulting pressure may be viewed as an evidentiary standard multiplier, and is appropriately represented as a ratio as follows (Section “Resultant pressure”):

$$\text{Resultant Pressure} : \frac{\text{Upward Pressure}}{\text{Downward Pressure}}$$

$$\frac{(\text{Cost-of-a-False-Positive}) \times \text{PP}(E_{\text{misleading}})}{(\text{Benefit-of-a-True-Positive}) \times \text{PP}(E_{\text{reliable}})}$$

Further, the model holds that such costs and benefits may be evaluated either egoistically or altruistically, in relation to one’s own practical interests or to those for whom one has an altruistic concern (Section “Altruistic knowledge intuitions and doubts”).

This theory explains a very wide range of cases in which the evidentiary standard for producing a knowledge intuition seems too high or too low, or seems to have modulated. For example, when all else is equal and the Cost-of-a-False-Positive is increased, the evidentiary standard rises. When a false positive for DeRose will result in a very important cheque failing to transact due to insufficient funds, his evidentiary standard for taking it that the bank will be open on Saturday seems to rise to the point that his previously sufficient evidence is found to be insufficient (Section “The cost of false positives”). Conversely, when the Benefit-of-a-True-Positive is prominent, evidentiary standards seem to drop. When the parent considers whether their child is a talented dancer and plausibly has the sense that there is a significant Benefit-of-a-True-Positive, even weak evidence is taken as being sufficient (Section “The benefit of true positives”). However, when the Cost-of-a-False-Positive and the Benefit-of-a-True-Positive are more or less balanced, evidentiary standards do not seem to be either too high or too low, such as in DeRose’s low-stakes Bank Case A (Section “The cost of false positives”); and when the Cost-of-a-False-Positive and the Benefit-of-a-True-Positive are extremely unbalanced, the evidentiary standard can be soaringly high, as it is in lottery cases (Section “Resultant pressure”).

Further, the Downward Pressure on the evidentiary standard for taking it that a success-enabling condition has been met can be transmitted through to the evidentiary standard for taking it that an associated evidence-value-reversing condition has been met—if the Benefit-of-a-True-Positive for the latter is taken as including the Benefit-of-a-True-Positive for the former. For example, the Downward Pressure on the evidentiary standard for taking it that an ideological image of the world is correct may transmit through to the evidentiary standard for taking it that seemingly acceptable contradictory evidence is in fact faulty (Section “Downward pressure may bias our assessment of the strength of evidence”).

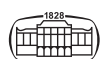
These findings, together with a consistent Primary indication of learning, suggest that the Learning Model should be adopted as a major component of error



management theory, to replace its previous posits of specific inherited biases in belief or doubt.

## REFERENCES

- Al-Shawaf, L., Lewis, D. M. G., Wehbe, Y. S., & Buss, D. M. (2019). Context, environment, and learning in evolutionary psychology. In T. K. Shackelford, & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 1–12). Springer International Publishing. [https://doi.org/10.1007/978-3-319-16999-6\\_227-1](https://doi.org/10.1007/978-3-319-16999-6_227-1).
- Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, 49(6), 1621.
- Asp, E., Manzel, K., Koestner, B., Cole, C. A., Denburg, N. L., & Tranel, D. (2012). A neuropsychological test of belief and doubt: Damage to ventromedial prefrontal cortex increases credulity for misleading advertising. *Frontiers in Neuroscience*, 6, 100. <https://doi.org/10.3389/fnins.2012.00100>.
- Asp, E., Manzel, K., Koestner, B., Denburg, N. L., & Tranel, D. (2013). Benefit of the doubt: A new view of the role of the prefrontal cortex in executive functioning and decision making. *Frontiers in Neuroscience*, 7, 86. <https://doi.org/10.3389/fnins.2013.00086>.
- Boulding, K. (1964). *The meaning of the twentieth century*. Harper & Row.
- Boyer, P., & Barrett, H. C. (2015). Intuitive ontologies and domain specificity. In: *Handbook of evolutionary psychology* (2nd ed., pp. 168–180). Hoboken, NJ: Wiley.
- Bratman, M. (1999). *Faces of intention*. Cambridge University Press.
- Buss, D. M. (2009). How can evolutionary psychology successfully explain personality and individual differences? *Perspectives on Psychological Science*, 4(4), 359–366. <https://doi.org/10.1111/j.1745-6924.2009.01138.x>.
- Cohen, S. (1999). Contextualism, skepticism, and the structure of reasons. *Nous*, 33, 57–89. <https://doi.org/10.1111/0029-4624.33.s13.3>.
- Conee, E. (2005). Contextualism contested. In M. Steup, & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 47–56). Blackwell Publishing.
- Craig, E. (1990). *Knowledge and the state of nature: An essay in conceptual synthesis*. Clarendon Press.
- Cyrus, K., Schwarz, S., & Hassebrauck, M. (2011). Systematic cognitive biases in courtship context: women's commitment-skepticism as a life-history strategy? *Evolution and Human Behavior*, 32(1), 13–20. <https://doi.org/10.1016/j.evolhumbehav.2010.07.006>.
- Daly, M. (2021). Personality traits and social structure. In M. MacLachlan, & J. McVeigh (Eds.), *Macropsychology: A population science for sustainable development goals* (pp. 63–85). Springer International Publishing. [https://doi.org/10.1007/978-3-030-50176-1\\_4](https://doi.org/10.1007/978-3-030-50176-1_4).
- Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), 13335–13340. <https://doi.org/10.1073/pnas.1102131108>.
- DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, 52(4), 913–929.
- Enright, E. A., Alonso, D. J., Lee, B. M., & Olson, K. R. (2020). Children's understanding and use of four dimensions of social status. *Journal of Cognition and Development*, 21(4), 573–602. <https://doi.org/10.1080/15248372.2020.1797745>.
- Fantl, J., & McGrath, M. (2002). Evidence, pragmatics, and justification. *The Philosophical Review*, 111(1), 67–94.
- Feuer, L. S. (2010). *Ideology and the ideologists*. Transaction Publishers (1975).
- Fodor, J. A. (1983). *The modularity of mind*. MIT Press.
- Godfrey-Smith, P. (1991). Signal, decision, action. *The Journal of Philosophy*, 88(12), 709–722. <https://doi.org/10.2307/2027008>.
- Godfrey-Smith, P. (1998). *Complexity and the function of mind in nature*. Cambridge University Press.
- Goertzel, T. (1992). *Turncoats and true believers: the dynamics of political belief and disillusionment*. Prometheus Books.
- Goldman, A. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, 18, 771–791. <https://doi.org/10.2307/2025679>.
- Green, D. M., & Swets, J. A. (1966). *Signal detection and psychophysics*. Wiley.
- Haselton, M. G. (2003). The sexual overperception bias: Evidence of a systematic bias in men from a survey of naturally occurring events. *Journal of Research in Personality*, 37(1), 34–47. [https://doi.org/10.1016/s0092-6566\(02\)00529-9](https://doi.org/10.1016/s0092-6566(02)00529-9).
- Haselton, M. G. (2007). Error management theory. In R. F. Baumeister, & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (Vol. 1, pp. 312–313). SAGE Publications, Inc. <https://www.doi.org/10.4135/9781412956253.n189>.
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78(1), 81–91. <https://doi.org/10.1037//0022-3514.78.1.81>.
- Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, 10(1), 47–66. [https://doi.org/10.1207/s15327957pspr1001\\_3](https://doi.org/10.1207/s15327957pspr1001_3).
- Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (2nd ed., pp. 1–20). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119125563.evpsych241>.
- Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *The Journal of Philosophy*, 105(10), 571–590. <http://www.jstor.org/stable/20620129>.
- Henningsen, D. D., & Henningsen, M. L. M. (2010). Testing error management theory: Exploring the commitment skepticism bias and the sexual overperception bias. *Human Communication Research*, 36(4), 618–634. <https://doi.org/10.1111/j.1468-2958.2010.01391.x>.
- Iredale, W., van Vugt, M., & Dunbar, R. (2008). Showing off in humans: Male generosity as a mating signal. *Evolutionary Psychology*, 6(3), 386–392. <https://doi.org/10.1177/147470490800600302>.
- Jackson, R. E., & Cormack, L. K. (2007). Evolved navigation theory and the descent illusion. *Perception & Psychophysics*, 69(3), 353–362. <https://doi.org/10.3758/bf03193756>.
- Leonard, A. S., Dornhaus, A., & Papaj, D. R. (2011). Flowers help bees cope with uncertainty: Signal detection and the function of





- floral complexity [Article]. *Journal of Experimental Biology*, 214(1), 113–121. <https://doi.org/10.1242/jeb.047407>.
- Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, 85(5), 1030–1051. <http://libezproxy.open.ac.uk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=1979-26009-001&site=eds-live&scope=site>.
- Lerner, M. J., Miller, D. T., & Holmes, J. G. (1976). Deserving and the emergence of forms of justice. *Advances in Experimental Social Psychology*, 9, 133–162. [https://doi.org/10.1016/S0065-2601\(08\)60060-X](https://doi.org/10.1016/S0065-2601(08)60060-X).
- Mazzini, J. (1946). *The living thoughts of Mazzini* (2nd ed.). Cassell & Co.
- McCarley, J. S., & Benjamin, A. S. (2013). Bayesian and signal detection models. In J. D. Lee, & A. Kirlik (Eds.), *The oxford handbook of cognitive engineering* (pp. 465–475). Oxford University Press.
- McCauley, R. N., & Henrich, J. (2006). Susceptibility to the Müller-Lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology*, 19(1), 79–101. <https://doi.org/10.1080/09515080500462347>.
- McKay, R., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, 31(5), 309–319. <https://doi.org/10.1016/j.evolhumbehav.2010.04.005>.
- Moon, A. (2017). The nature of doubt and a new puzzle about belief, doubt, and confidence. *Synthese*, 195(4), 1827–1848. <https://doi.org/10.1007/s11229-016-1310-y>.
- Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, 13(2), 87–110. [https://doi.org/10.1207/s15326969eco1302\\_2](https://doi.org/10.1207/s15326969eco1302_2).
- Norem, J. K., & Chang, E. C. (2002). The positive psychology of negative thinking. *Journal of Clinical Psychology*, 58(9), 993–1001.
- Norem, J. K., & Illingworth, K. S. S. (1993). Strategy-dependent effects of reflecting on self and tasks: Some implications of optimism and defensive pessimism. *Journal of Personality and Social Psychology*, 65(4), 822–835. <https://doi.org/10.1037/0022-3514.65.4.822>.
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291–1298.
- Park, J. H. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior*, 27(2), 65–87. <https://doi.org/10.1023/a:1023910408854>.
- Price, J. J. (2013). Why is birdsong so repetitive? Signal detection and the evolution of avian singing modes [Article]. *Behaviour*, 150(9/10), 995–1013. <https://doi.org/10.1163/1568539x-00003051>.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60281-6](https://doi.org/10.1016/S0065-2601(08)60281-6).
- Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, 50(4), 19–45. <https://doi.org/10.1111/j.1540-4560.1994.tb01196.x>.
- Scott-Kakures, D. (2000). Motivated believing: Wishful and unwelcome. *Nous*, 34(3), 348–375.
- Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, 6(1), 38–47. <https://doi.org/10.1177/1745691610393528>.
- Spencer, S. M., & Norem, J. K. (2016). Reflection and distraction defensive pessimism, strategic optimism, and performance. *Personality & Social Psychology Bulletin*, 22(4), 354–365. <https://doi.org/10.1177/0146167296224003>.
- Sripada, C. S., & Stanley, J. (2012). Empirical tests of interest-relative invariantism. *Episteme*, 9(1), 3–26. <https://doi.org/10.1017/epi.2011.2>.
- Stanley, J. (2005). *Knowledge and practical interests*. Oxford University Press.
- Stefanucci, J. K., & Proffitt, D. R. (2009). The roles of altitude and fear in the perception of heights. *Journal of Experimental Psychology: Human Perception & Performance*, 35(5), 424–438.
- Tadinac, M. (2020). There and back again: On the need to put the proximate back into causation. *Evolutionary Behavioral Sciences*, 14(4), 379–383. <https://doi.org/10.1037/ebs0000208>.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210. <https://doi.org/10.1037/0033-2909.103.2.193>.
- Tooby, J., & Cosmides, L. (2015). The theoretical foundations of evolutionary psychology. *The Handbook of Evolutionary Psychology*, 2.
- Ullman-Margalit, E. (1983). On presumption. *The Journal of Philosophy*, 80(3), 143–163.
- West-Eberhard, M. J. (1975). The evolution of social behavior by kin selection. *The Quarterly Review of Biology*, 50(1), 1–33. <https://doi.org/10.1086/408298>.
- West-Eberhard, M. J. (1983). Sexual selection, social competition, and speciation. *The Quarterly Review of Biology*, 58(2), 155–183. <https://doi.org/10.1086/413215>.
- Witt, J. K., & Sugovic, M. (2013). Spiders appear to move faster than non-threatening objects regardless of one’s ability to block them. *Acta Psychologica*, 143(3), 284–291.
- Zietsch, B. P., Sidari, M. J., Murphy, S. C., Sherlock, J. M., & Lee, A. J. (2021). For the good of evolutionary psychology, let’s reunite proximate and ultimate explanations. *Evolution and Human Behavior*, 42(1), 76–78. <https://doi.org/10.1016/j.evolhumbehav.2020.06.009>.

## Appendix: Brief Comment on McKay and Efferson (2010)

Ryan McKay and Charles Efferson (2010) suggest that the behaviours that EMT proponents tell us are attributable to inherited biases in belief or doubt may consistently be attributable to behavioural biases, and, accordingly, that inherited biases in belief or doubt may not exist. This last conclusion is consistent with the findings of the present research. However, the picture they present misrepresents the situation insofar as they fail to consider the possibility that learned biases in belief or doubt can achieve comparable outcomes. Consider their well-known example.





A man may be nearly convinced that a particular woman is not interested in him. Nonetheless, if he admits any hope, however slim, he will behave as if she is interested. He will do so precisely because the fitness costs of a missed encounter are so much larger than the costs of brief embarrassment (2010, p. 314).

The suggestion is that behaviour that EMT proponents may attribute to an inherited biased belief about the woman being interested might well be properly attributed to an inherited behavioural bias to behave as if she is. However, they do not examine the man's substantial doubt about the woman being interested, telling us only that he is "nearly convinced that [the] woman is not interested". The doubt is a given, and the example focuses on the creation

or selection of behaviours. However, the behaviours that are performed are better explained by the Learning Model, and by recognizing the behaviour-guiding role of doubt. The success-enabling condition for the man is the woman being interested, he has doubt about the SEC, and adopts behaviour strategy S2 (Section "Doubt is a unique behaviour-guiding state"). He behaves as if the SEC has been met. He might approach and engage the woman in conversation; but he will also likely look for opportunities to hedge, to mitigate the harm that could follow from behaving as if the SEC has been met when it has not. Such hedging might well include not being overly presumptuous. He would also likely study her demeanour for signs of interest or disinterest.