Extending the Testimony Problem: Evaluating the Truth, Scope, and Source of Cultural Information

Brian Bergstrom, Bianca Moehlmann, and Pascal Boyer
Washington University at St. Louis

Children’s learning—in the domains of science and religion specifically, but in many other cultural domains as well—relies extensively on testimony and other forms of culturally transmitted information. The cognitive processes that enable such learning must also administrate the evaluation, qualification, and storage of that information, while guarding against the dangers of false or misleading information. Currently, the development of these appraisal processes is not clearly understood. Recent work, reviewed here, has begun to address three important dimensions of the problem: how children and adults evaluate truth in communication, how they gauge the inferential potential of information, and how they encode and evaluate its source.

In an important review article, Harris and Koenig draw our attention to an aspect of cognitive development that is too often neglected: the need for children to rely extensively on culturally transmitted information while simultaneously erecting safeguards against misleading or deceptive input (Harris & Koenig, 2006). In this essay, we wish to pursue their suggestive forays into the domains of science and religion, as well as extend their cognitive program to broader aspects of cultural transmission, by examining some of the cognitive challenges a developing mind must confront in order to be a successful consumer of testimony and other forms of cultural knowledge.

Remarkably, our current framework in the study of cognitive development has very little to say about the way in which young children assess the value of information acquired from cultural elders or peers. This is somewhat paradoxical because (i) children acquire vastly more information through communication than through experience and (ii) a good proportion of that information is either fragmentary or misleading. Recent advances in the study of cognitive development have not generally placed great emphasis on this particular aspect of the developing mind. So for instance, we know a great deal about how young children develop sound intuitions about living things and biological processes (see, e.g., Gelman, Coley, & Gottfried, 1994; Gelman & Hirschfeld, 1999; among many others) but much less about how they filter out, as it were, unreliable information about talking helicopters and emoting locomotives.

True, there is important research on children’s “fantasy–reality distinction” (Woolley, 1997) and their understanding of “magic” as both real and counterintuitive (Harris, 1994). But however indispensable, these research programs only focus on a very small part of the processes engaged in the acquisition of cultural knowledge. The problem of how to assess the value and validity of information is not limited to fantasy, stories, jokes, or tropes. It is pervasive in everyday communication. Most information acquired from others would be useless unless its reliability was clearly understood and its inferential potential strongly constrained, as we try to show presently.

Humans occupy a special ecological niche that physical anthropologists have called the “cognitive niche” (Tooby & DeVore, 1987). In the same way as the natural environment of whales is sea water and that of cheetahs is the savannah, the natural environment of humans is information about the natural and social environment, and to a very large degree information conveyed by conspecifics. One should therefore expect to find early developed and specific capacities for acquiring and transmitting information, as well as capacities for evaluating its reliability and occasionally deceiving others through skillful misinformation. And indeed children’s capacities for language, “theory of mind,” empathy, and social interaction are reliably developing, species-typical aptitudes.

Because of these specific capacities, humans acquire orders of magnitude more information about
their environments than members of other species. But enhanced communication abilities come at a price, specifically in vulnerability to misleading or unreliable information. As we know from the study of communication systems in other animals, there are important adaptive advantages to the capacity to deceive, which in turn make the capacity to detect deception a decisive adaptation. This kind of arms race between deceivers and detectors can reach an equilibrium when signals are too costly to fake or when detection capacities make cheating more costly than honest signaling (Bradbury & Vehrencamp, 2000). Therefore, we should expect the best experts at communication, human beings, to have considerable competence for the evaluation of information (Gintis, Smith, & Bowles, 2001).

To understand how children’s minds process the testimony they encounter, we must focus on three aspects of the encounter: how children and adults evaluate truth in communication, how they gauge the inferential potential of that information, and how they encode and evaluate its source.

Evaluating Truth

**Do We Believe First, Think Second?**

The first problem is to evaluate whether an utterance that communicates a particular piece of information is sincere, that is whether it should be taken at face value as conveying true information (or at least information the speaker holds to be true). This is not just a problem for young children or for the development of complex knowledge systems. In all everyday communication, we need to make (generally implicit) decisions about the truth value of utterances.

Is truth our default value? That is, does a cognitive system start by taking all conveyed information as true, and then tag some of it as possible, probable, implausible, etc., adjusting subjective “credal states” in the light of other available information? Or, conversely, does a cognitive system “quarantine” new information until its proper truth value can be ascertained? Both assumptions receive some support from psychological research. However, they have been put forward in such different paradigms and on the basis of such different kinds of evidence that it is difficult to adjudicate between them.

From an experimental perspective in social psychology, the problem is, Does a consideration of content entail a consideration that the utterance is true? Are claims immediately accepted? Is there a phase of doubt and then acceptance, or vice versa? Could children accept information without actually believing its premises, and vice versa? Previous research by Dan Gilbert and colleagues suggests that adults generally accept new information at comprehension and later “unaccept” it (Gilbert, 1991; Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). In Gilbert’s protocols, participants were presented with statements that contained a foreign word, which made it difficult to assess the statements’ veracity. After reading each statement, participants were told whether it was true or false. But on some trials, participants were asked to complete a secondary task a few milliseconds after learning whether the statement was true. When later prompted to recall the truth status of statements, participants would erroneously claim that false statements were true if the veracity announcement had been interrupted. Ostensibly, participants accepted the initial information as true, and the subsequent interruption prevented the kind of more effortful processing of the statement and its veracity normally required to “unbelieve” it (Gilbert et al., 1990). Gilbert and colleagues argue for truth as the default assumption for most incoming pieces of information. Gilbert also emphasizes the difference between believing and merely understanding information. His results suggest that the acceptance (and comprehension) of information occurs automatically and immediately, whereas the rejection of information (and deciding whether to believe or not to believe) involves a subsequent, more effortful process (Gilbert et al., 1990, 1993). Therefore, individuals will in some circumstances believe, or think they believe, information they actually know they should not believe.

This is relevant to children’s encounters with testimony, because the same processes may operate when children receive novel information. Perhaps children automatically comprehend new information and initially accept it, but are able to “unbelieve it” in cases where they have sufficient opportunity, motivation, or previous information to evaluate the claim. In other cases, they may not be able to do so, and by default retain their initial acceptance and belief of the new information.

This may have important consequences in the domains surveyed by Harris and Koenig. Consider scientific and religious statements. Do children believe novel information about these topics because they evaluate the veracity of the information or is it because “unbelieving” is too difficult? At what age does the process become easy enough, and are there graduated developmental steps that lead there? Here is a domain of possible research that may
complement the kinds of studies summarized by Harris and Koenig.

Meta-Representations

The distinction between processing information and holding it as true is crucial to recent models in pragmatics and cognitive development. These models start from the question, How do children handle information that is neither clearly true nor clearly false? One possibility is that such information is “meta-represented.” In this view, children would not directly represent propositions like

[1] God is all-merciful

but rather entertain something like

[2] “God is all-merciful” is true

This description of some information as meta-represented (see Leslie, 1995; Sperber, 1991) allows us to solve a number of problems and explain some features of these representations. For instance, it explains why people can hold true a proposition like [2] without being committed to a specific interpretation of [1]. That is to say, a child may well acquiesce and even defend the notion that “God is all-merciful” without necessarily having a precise lexical entry for “merciful.” Similarly, adults can defend the truth of “black holes swallow all radiation in their vicinity” on the basis of a similar meta-representation (in this case, “Reputable scientists say . . .”). Many items of information, once they are acquired from an external source, may subsist in this special “meta-representational” state, which implies that they are not considered strictly true in the same sense as other beliefs such as “My name is Brian” or “It’s cold today.”

Representational status is important for the issues discussed by Harris and Koenig, because the number of such representations is probably very high among young children, who lack the conceptual means to make literal sense of most of what they hear. This description raises a difficult question, however, concerning the computational operations associated with such meta-representations. One might argue that meta-represented content is simply quoted or “mentioned” as philosophers put it. In this description, people for instance do not believe

[1] “God is all-merciful”

What they believe is

[2] “‘God is all-merciful’ is true”

But this is problematic because, under this description, representing [2] would be exactly the same as representing

[3] “‘Gott ist gnädig’ is true”

which is certainly not equivalent. Meta-represented content differs from pure quotations in several ways. First, mental content that is meta-represented allows operations similar to those engaged by content that is actually understood (it is an “iconic” representation of literal content [Recanati, 1997]). That is, even children with no grasp of theology will understand that if [1] is true, then people who say that God is not merciful are wrong. Second, meta-represented content very often supports inferential processes. Even if the child does not have a proper interpretation of [1], perhaps because she does not know what “merciful” means, she can still infer that God has a quality that others do not have, that this is a behavioral quality, that God surpasses humans in this particular respect, etc. Third, meta-represented content, even if stored in a fairly literal manner (so that the child repeats [1] in the exact manner she heard it), is stored in such a way that it can be retrieved in the service of further inferences, once the rest of the knowledge base has changed in a relevant way. The child who is told one day what “mercy” means will retrieve statement [1] and assign a new interpretation to it.

In sum, a consideration of recent work in both social psychology and pragmatics suggests that variables of attention, processing effort, and the representational format of incoming testimony will all play a role in a better understanding of how truth value is ascertained. These different empirical paradigms generate divergent predictions, yet no imminent resolution to these issues is discernable from the literature, and as such we may expect that future research will shed important light on this aspect of children’s trust in testimony.

Evaluating Inferential Potential

Inferences and Scope

A consideration of mental representation leads us into a more general description of the kinds of mental “tags” that get attached to communicated content. So far, we have discussed truth-related tags like “proposition ‘p’ is true.” But communicated
content is also associated with a number of other tags, to do with the person who made the statement, the physical location and social context in which the statement was made, information about other such statements, information about the ways in which a vague statement might become more precise, and so on.

A general model for the processes involved in these “tagging” operations is what evolutionary psychologists have called “scope syntax” (Cosmides & Tooby, 2000). Without some way for the mind to attach boundary conditions to encountered testimony, there is no way to specify the legitimate inferential potential of a given piece of information, and therefore no way to prevent suspicious or erroneous information from infiltrating knowledge stores and inference systems. A single error in knowledge could potentially generate cascades of spurious conclusions, which may be fed as input into other mental systems, compounding and proliferating error throughout the mind and engendering mal-adaptive behavior. Qualifications make information truly informative. That this bush is rich in berries may be true, but only in certain seasons; that this animal is dangerous may be true, but only at certain proximities and when its young are near; that this person is a reliable cooperator may be valid in the context of work, but tragically ill-founded when faced with serious danger. The concern is that a developing mind may fail to perceive the limits or conditions under which a given piece of testimony is valid or applicable. Without cognitive safeguards, a child with a capacity for language is vulnerable to misunderstanding or misremembering, being misled or outright deceived—an enormous liability.

This allows us to reconsider Gilbert’s results in a new way. Participants who are not allowed to process a piece of information may not be demonstrating a bias to prejudge information as true so much as they are demonstrating a failure to create a memory trace that includes all the appropriate tags to limit the scope of that information. In this framework, one could make further predictions for the results of such experiments. Participants should not just believe the communicated information, they should also lose track of its scope. So experimental subjects would tend to find the information relevant to more situations than control participants who created the relevant tags. Such research with both children and adults may help to clarify the precise causal pathways by which our cognitive systems process and evaluate the truth value of testimony and specify its inferential potential.

**Plausibility and Domain-Specific Assumptions**

As mentioned above, one crucial aspect of meta-represented or “tagged” information is that it can be retrieved and combined with appropriate items from the extant knowledge base. Thus, even entirely novel information is bound to create some inferential effects.

Now a crucial issue here for the issues raised by Harris and Koenig is the extent to which these inferential effects are influenced by which domain of knowledge is activated. Human semantic knowledge is not a seamless network of interconnected bits of information. Rather, it is organized into specific domains, for which one can find (a) a specific database, (b) specific targets for learning (what input is relevant), and (c) specific inferential rules. A wide array of evidence, from neuropsychology to neuroimaging and behavioral experiments to developmental trends, seems to confirm this organization of semantic knowledge. The notion of an intuitive ontology as a motley of different domains informed by different principles was first popularized by developmental psychologists who proposed distinctions between physical–mechanical, biological, social, and numerical competencies as based on different learning principles (Hirschfeld & Gelman, 1994). These empirical findings have led developmental psychologists to cast doubt on the notion of a general “learning logic” that would govern cognitive development across domains (Hirschfeld & Gelman, 1994).

If there are early developed inferential principles, specific to knowledge domains, one would expect them to have an effect on the treatment of incoming cultural input. In particular, some pieces of culturally available information should be easier to acquire than others, given their compatibility with previous, spontaneous assumptions in the child’s mind. Some elements of cultural input may just “fill in” the place holders in children’s ontological expectations. For instance, acquiring culturally specific notions about personality and motivation may provide richer causal contexts for elaborating very general and spontaneous assumptions of theory of mind (Boyer, 1998). Conversely, it is clear that early dispositions for numerical competence not only provide little help but actually hinder children’s acquisition of cultural information about negative numbers or fractions (Gelman & Meck, 1992).

With regard to testimony, this is likely to have a strong effect on acquisition processes and the evaluation of sources. To the extent that a certain piece of cultural knowledge is compatible with spontaneous
assumptions, we would expect the child to pay little attention to source reliability. By contrast, information that violates early developed principles should trigger at least some search for additional evidence, including evaluation of the sources.

Consider for instance the domain of illness and its transmission. In many different cultures, people see illness as a state that can be transmitted from person to person, in a way that is generally inscrutable (the vectors are not visible), and independently of the mode of contact (one can acquire another’s illness through sharing food, having sex, or simple touch or proximity). What makes the avoidance sound symbolic or mystical are the explicit notions (“bad air,” “miasma”) people invoke to explain intuitions they already had. For evolutionary reasons humans may be rather good at detecting sources and processes of contamination, yet remain very vague in their explicit rationale for avoiding them (Rozin, Millman, & Nemeroff, 1986). Note, also, that intuitions about illness appear early—and differentiate between disease transmission and other kinds of causal processes (Keil, Levin, Richman, & Gutheil, 1999; Siegal, 1988).

Children seem strongly predisposed to acquire local theories of disease transmission based on the following principles: (1) the source of illness (pathogen) is invisible; (2) any mode of bodily contact is equally dangerous; (3) there is no dose effect (hence even very small amounts of a contaminant are dangerous). This would explain why most “folkbiological” theories include some theory of transmission along these lines (Atran, 1998). This would also predict that the aspects of scientific biology that dovetail with this intuitive picture are easily acquired, and crucially require no special evaluation of the source.

By contrast, consider a domain in which explicit cultural input seems to go against entrenched intuitions. Religious concepts generally include some salient violation of intuitive assumptions (Boyer, 1994). Supernatural agents for instance are construed as physically (and often biologically) counterintuitive (Barrett & Nyhof, 2001). At the same time, supernatural agents are tacitly construed as standard intentional agents, whose perceptual and thought processes are essentially similar to those of humans (Barrett & Keil, 1996). In such a domain, the child is invited to acquire beliefs that emphatically do not match spontaneous assumptions—and are often described precisely in those counterintuitive terms. That is, ancestors or gods are explicitly described as different from agents with whom we usually interact. In such a domain it may not be a coincidence that we also find extremely strong effects of source. The diffusion of religious knowledge, norms, and concepts generally involves particularly authoritative figures (shamans, priests, sages, saints) whose statements are supposedly more reliable or closer to truth than those of average believers (Boyer, 1990, 2001).

To sum up, the processes whereby children evaluate information and its sources may be strongly influenced by prior assumptions about the domain of knowledge concerned. Because different domains carry different kinds of assumptions and different expectations about possible new information, it might be profitable to consider testimony evaluation domain by domain rather than just the outcome of a central estimation process.

“Speculative” Versus “Applied” Knowledge?

A consideration of domain specificity also brings to our attention connections between testimony and behavior. Because the systems and processes of the human mind rally around a common purpose (to orchestrate and implement useful, appropriate behavior), an important distinction may be drawn between testimonial content that is behaviorally salient versus behaviorally neutral. Not all testimony is equally relevant to behavior, making some types of acquired information more “speculative,” as it were, versus more “applied.” If testimony is fed into mental decision rules that are prepared to give specialized treatment to particular kinds of information, we might expect different arenas of testimony to register different patterns of effect (e.g., on attention, memory, and manifest behavior).

The particular examples of testimony described by Harris and Koenig (e.g., the shape of the earth, the cognitive attributes of God) may be construed as “speculative” in this sense (although crucial to the child’s knowledge acquisition). Whether the earth is round or flat, or whether God really is everywhere at once, may have few discernible consequences for most children’s day-to-day behavioral choices. By contrast, some domains of testimony carry significance that makes the information content especially salient and memorable, for example information regarding germs and contagion, dangerous animals, or what foods are safe to eat.

Furthermore, in domains where knowledge directly affects behavior, we often observe that source effects (e.g., authoritativeness as a cue of reliability) are very weak compared with early acquired, probably evolved dispositions. Consider the relationship between fear and danger. Children readily learn to
fear snakes and spiders, heights and water, storms and darkness, abandonment and social failure with little or no effort, and with little testimonial input. But other dangers—such as electricity, hand guns, crossing busy streets, talking to strangers, smoking cigarettes, or unprotected sex—fail to compel the same level of apprehension, even when adult testimony is persistent, coherent, and evidence based, and comes from people who have the child’s welfare at heart (Maurer, 1965).

Such discrepancies pose interesting challenges for an understanding of the cognitive basis of children’s trust in testimony. In this particular case, evolutionary explanations show great promise. Dangers that posed a recurrent threat to fitness over human evolutionary history may now carry indelible stamps of prepotent salience, whereas sources of danger like power sockets and fast cars carry no intuitive valence because they are evolutionarily novel (Marks, 1987; Ohman & Mineka, 2001). In this view, the child’s mind may be less receptive to those aspects of adult testimony that, by the nature of the content, are not behaviorally relevant, or are behaviorally relevant but not evolutionarily privileged.

In sum, the inferential potential—salience, relevance, resilience—of testimonial input appears to be shaped and constrained by fundamental features of children’s cognitive architecture, such as domain-specific expectations, processing boundaries (“scope syntax”) attached to incoming information, and the behavioral significance of testimony as registered by mental decision-making algorithms. An enhanced understanding of how children develop these key cognitive capacities may take us a long way toward understanding how, as Harris and Koenig document, children manage to negotiate complex informational environments with such remarkable competence.

Evaluating the Source

Let us now return to the issue of source—who the teller is—as a special case of the general “tagging” process. Surprisingly, there is no tradition of empirical cognitive research into the process whereby the identity of a speaker results in an evaluation of reliability. An exception is the recent research into “source memory” and particularly “source or reality monitoring.” The latter terms were first introduced by Johnson and colleagues (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981) and has since been studied extensively, especially in the field of eyewitness memory (for a review, see Mitchell & Johnson, 2000). In addition, research in this tradition is often concerned with the ability to distinguish one’s own actions from actions observed or imagined (for a review, see Garry & Polaschek, 2000; Loftus, 1997).

This tradition has not yet delivered a clear and consistent picture of the developmental aspects of source memory, even though we know that such processes occur in children. This is perhaps due to the concern with reality monitoring (was this particular representation imagined, experienced suggested?), a different question from the issue of who said what, and how that makes a particular statement more or less reliable. On this precise issue, research by Koenig and colleagues has shown that children are able to remember sources of information as well as which sources to trust (Koenig, Clément, & Harris, 2004). In one of these studies, young children were presented with two speakers who provided nominal labels for objects familiar to the child. One speaker labeled the objects correctly; the other did so incorrectly. Children were able to identify which speaker was unreliable, and when presented with a novel object children showed selective trust. In other words, children displayed decisive trust for the labeling decisions made by the more reliable speaker.

However, previously established trustworthiness becomes irrelevant when children have firsthand knowledge of the information in question. When two speakers made statements about the color of a hidden item, children learned to trust the reliable, but not the unreliable, speaker. But when children had the opportunity to observe in advance the color of the item, and both the reliable and the unreliable speaker made incorrect statements, children ignored both speakers and used their firsthand knowledge to report the color of the item (Koenig et al., 2004).

Therefore it would seem that as children’s knowledge about a topic increases, the acceptance of novel, contradictory, or unexpected information on that topic becomes less likely. In contrast, information is accepted quite readily if the child has little or no previous knowledge, or if the information does not seem to contradict previously existing knowledge. Future research may need to consider new information gained from authorities. Children are often presented with novel information multiple times a day, received from parents, teachers, friends, and the media. In most cases, children have learned to trust these sources, and such sources could therefore be considered authorities. In addition to trusting reliable sources, children are also likely to trust sources that are authorities.

In the experiments reported by Harris and Koenig (2006), children received information about topics
with which they were only somewhat familiar. In addition, the information did not contradict what they already knew about the topic. The information was provided by sources that, most likely, had been proven to be either reliable or authoritative. Therefore it might be of great interest to pursue this experimental program and address the next questions: Would children believe completely novel religious or scientific information if it was provided by unfamiliar sources? Would they believe information that contradicts their existing beliefs or knowledge?

**What Is To Be Done?**

As human beings, we survive by acquiring information—and by judging its reliability and value to our own goals. Strikingly, there is no very precise cognitive psychology of credal states and their dynamics (Gigerenzer & Murray, 1987; Gigerenzer & Selten, 2001), although there are many normative theories of how subjective probability should change as a function of external information (Gärdenfors & Sahlin, 1988). There is even less research on credal states in young children, even though the dynamics of belief fixation are clearly crucial to cognitive development.

Perhaps one crucial limitation of our current studies is that we only use explicit or direct methods in understanding belief fixation, where implicit or indirect tests would be more appropriate. The protocols used in the study of testimony generally consist of explicit questions that probe the child’s belief states (“do you think it’s true that . . .”), inferences (“so we know that p, does this mean that q?”), or counterfactuals (“if p was the case, would q be the case too?”). But the processes involved in creating tags and other scope information may operate in computational modes that do not allow easy explicit access in adults, a fortiori in children. The fact that a new piece of information (So-and-so was wrong about x) can trigger cascades of re-evaluation of other information (communicated or inferred from what So-and-so said) is not easy to express for children. Indeed, the epistemic lexicon of children (“true,” “false,” “lie”) is extremely narrow, yet their evaluation of plausible and possible information is often remarkably subtle. Implicit tests might permit finer resolution in our studies of child cognition. Moreover, parallels exist in other domains. In lexical meaning for instance, we know that children are often initially extremely bad at providing explicit definitions for words that they use appropriately on the basis of an implicit definition, a finding that also applies to adults.

Harris himself showed the way for such implicit testing in his experiments on children’s magical thinking. Children’s explicit descriptions of what “magic” is, how it may work, whether it is real, etc., seem to lag behind the subtlety of their thoughts about fantasy processes (Harris, 1991, 1994). The same children who assert that an imaginary animal or ice-cream cone is only imaginary are also very likely to inspect a box where they placed that imaginary object (Johnson & Harris, 1994). In other words, there are some aspects of their own credal states that their explicit concepts simply do not capture. Unfortunately, our analytical vocabulary does not capture them either, at least so far. We may need to develop new techniques and protocols to address the issues opened by Harris and Koenig.

**References**


