JOURNAL OF COGNITION AND CULTURE 18 (2018) 358–389

BRILL

brill.com/jocc

Cross-Cultural Differences in Informal Argumentation: Norms, Inductive Biases and Evidentiality

Hatice Karaslaan[a], *Annette Hohenberger*[b], *Hilmi Demir*[b], *Simon Hall*[c], *and Mike Oaksford*[c*]

[a] nkara Yıldırım Beyazıt University, Ankara, Turkey
[b] Middle East Technical University, Ankara, Turkey
[c] Birkbeck College, University of London, London, UK
[*] *Corresponding author: mike.oaksford@bbk.ac.uk*

Abstract

Cross-cultural differences in argumentation may be explained by the use of different norms of reasoning. However, some norms derive from, presumably universal, mathematical laws. This inconsistency can be resolved, by considering that some norms of argumentation, like Bayes theorem, are mathematical functions. Systematic variation in the inputs may produce culture-dependent inductive biases although the function remains invariant. This hypothesis was tested by fitting a Bayesian model to data on informal argumentation from Turkish and English cultures, which linguistically mark evidence quality differently. The experiment varied evidential marking and informant reliability in argumentative dialogues and revealed cross-cultural differences for both independent variables. The Bayesian model fitted the data from both cultures well but there were differences in the parameters consistent with culture-specific inductive biases. These findings are related to current controversies over the universality of the norms of reasoning and the role of normative theories in the psychology of reasoning.

Keywords

cultural differences – Bayesian argumentation – source reliability – evidentiality – inductive biases – rational norms

Cross-cultural studies of reasoning and argumentation are important to two fundamental debates in the psychology of reasoning and thinking. First, what is the role, if any, of normative theories — theories of how we should or should not reason — in the psychology of reasoning (Elqayam & Evans, 2011; Elqayam & Over, 2016)? Second, do cultural differences, for example, individualistic western cultures vs collectivist Eastern cultures, lead to fundamentally different ways of thinking and reasoning (Mercier, 2011, 2013; Mercier, Zhang, Qu, Lu, & Van der Henst, 2015, Nisbett, Peng, Choi, & Norenzayan, 2001; Peng, & Nisbett, 1999)? Both debates pose the question, are there universal norms of reasoning (Mercier, 2011; Oaksford, 2014)?

However, norms come in different forms (Corner & Hahn, 2013). In particular, the norms governing informal argumentation have different origins which would be expected to bear on their universality (Corner & Hahn, 2013). Procedural norms (e.g., Van Eemeren & Grootendorst, 2004), like those governing courtroom proceedings, derive from social conventions (Corner & Hahn, 2013; Hahn & Oaksford, 2012). Mathematical theories of logic and probability theory are the source of most epistemic norms (Hahn & Oaksford, 2007, 2012). Assuming that people's behaviour respects these norms, these differing origins lead to different cross-cultural expectations. So, procedural norms, being based on social conventions, may be expected to vary across cultures.[1] In contrast, epistemic norms, being based on, presumably universal, mathematical laws, may be expected not to vary between cultures. However, research looking at the effects of levels of evidence quality on argumentation, presumably underpinned by epistemic norms, has revealed differences (e.g., Hornikx & Hoeken, 2007; Hornikx & ter Haar, 2013).

In this paper, we attempt to resolve this apparent inconsistency by considering that the norm, in this case, Bayes theorem, inheres in this theorem's functional form which dictates how evidence is combined. Consequently, while different cultures may systematically assign different values to the priors and likelihoods entering into the computation, these may be merely *inductive biases* (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010), the norm, Bayes theorem, may be invariant. In this paper, we test this conjecture by fitting a Bayesian model of argumentation to data from two quite divergent cultures, Turkish and English. If the model fits the data equally well from both cultures, we may interpret variation in the parameter values as reflecting culture-specific inductive biases rather the adoption of different norms.

1 Although, to the extent that these procedures are intended to be truth conducive, it would be expected that cultures would converge on similar procedures. However, social conventions perform many other important functions that may override getting to the truth.

Investigating Turkish-English differences is also of interest because these different linguistic cultures appear to mark evidence quality in different ways. One difference in evidence quality, which features strongly in courtroom procedures, is between eye-witness and hearsay evidence. In the formal setting of the courtroom, eye-witness testimony is regarded as of higher quality and as more persuasive than hearsay evidence. In Turkish, it is obligatory to mark this distinction in the morpho-syntactic structure of verbs in the past tense (Aikhenvald, 2004). In contrast, in English, marking this distinction (e.g., I saw that *p* vs I was told that *p*), is voluntary and lexical. That is, these linguistic cultures have different evidential systems (Aikhenvald, 2004), which have been argued to explain differences in cognitive processes between these cultures. For example, Turkish participants seem to be more trusting of reliable sources of information (Lucas, Lewis, Cansu Pala, Wong, & Berridge, 2013) and more dismissive of indirect, hearsay evidence (Tosun, Vaid, & Geraci, 2013) than English participants (but see, Papafragou, Li, Choi, & Han, 2007; Ünal, Pinto, Bunger, & Papafragou, 2016). The experiment we report here is the first to investigate whether differences in the way linguistic cultures mark evidence quality differentially affects informal argument evaluation.

We first introduce the procedural and epistemic accounts of argumentation and in particular how Bayes theorem has been used to explain how people evaluate informal arguments. We then review cross-cultural research in informal argumentation where evidence quality, for example, differences in expert opinion, has been varied. We conclude that the observed differences may be explained as inductive biases rather than by the employment of different norms. We then introduce the Turkish evidential system. We present the differences in the perceived strength of arguments that it would predict for an experiment that varies both evidentiality and levels of expert opinion. In the *Discussion*, we trace out the relevance of this research to the recent debate about the role of normative theories in the psychology of reasoning (Elqayam & Evans, 2011; Elqayam & Over, 2016).

1 Argumentation

Argumentation is the social activity whereby someone attempts to persuade an audience of a, possibly controversial, position by providing a series of propositions in its support. In critical discussions, the caveat "before a rational judge" (van Eemeren, Grootendorst, & Snoeck Henkemans, 1996, p. 5) is usually added to emphasise that the argument should be persuasive with respect to some normative standard. Informal arguments are those that people often

find convincing but for which there is no formal, logical treatment. Many of the informal argumentation schemes identified since Aristotle have therefore earned the label "fallacies," for example, the argument ad hominem (against the person) or the argument ad ignorantium (the argument from ignorance). However, many instances of these fallacies appear to be perfectly good arguments (Hamblin, 1970). For example, while we may find the argument from ignorance in (1) persuasive, we may not be persuaded by the argument from ignorance in (2):

This drug is safe because there is no evidence that it is not (1)
Ghosts exist because no one has proved that they do not (2)

Consequently, the task has been to explain why some instances (2) are fallacious while others seem reasonable (1).

Two complementary normative approaches have been proposed to explain this difference, the *procedural* and *epistemic* approaches (Hahn & Oaksford, 2012). Procedural approaches, for example, *pragmadialectical* theory (Van Eemeren & Grootendorst, 2004), provide procedural rules of engagement in a critical discussion like those that govern courtroom proceedings. Epistemic approaches use Bayes' theorem to evaluate the degree to which arguments should change people's subjective degree of belief in a conclusion (Hahn & Oaksford, 2007, 2012; Korb, 2004; Oaksford & Hahn, 2004; Zenker, 2013). In the procedural approach, fallacies arise because an interlocutor has used a discourse rule in the wrong context or the incorrect phase of an argument. However, even if both (1) and (2) occurred in the same phase of a critical discussion (i.e., the argumentative context), (1) would still be deemed stronger than (2) (Hahn & Oaksford, 2007).

The epistemic approach explains this difference using a content dependent measure of argument strength based on Bayes theorem (Hahn & Oaksford, 2006, 2007). On this view, the reason that (1) is deemed stronger than (2) is due to the factors that influence the computation of the posterior probability using Bayes rule. So, people's quantitative change in the degree of belief in a conclusion, C, brought about by an argument, a, is given by Equation 1 (where "¬" = not):

$$\Pr(C|a) = \frac{\Pr(a|C)\Pr(C)}{\Pr(a|C)\Pr(C) + \Pr(a|\neg C)\Pr(\neg C)} \quad \text{(Eq. 1)}$$

That is, the posterior degree of belief in the conclusion C given the argument a, $\Pr(C|a)$, is a function of the likelihoods, $\Pr(a|C)$ and $\Pr(a|\neg C)$, and the priors

$\Pr(C)$. The specific content of the argument fixes these quantities. The posterior ($\Pr(C|a)$), provides a measure of argument strength. Regarding our examples, (1) may be considered stronger than (2) because people's prior degree of belief that the drug is safe ($\Pr(C)$) may be higher than their prior degree of belief in the existence of Ghosts. Moreover, people may believe that although there are highly sensitive ($\Pr(a|C)$ is high) and specific ($1 - \Pr(a|\neg C)$ is high) tests for whether a drug is safe, similarly sensitive and specific tests for the existence of Ghosts are not available (Hahn & Oaksford, 2007; Oaksford & Hahn, 2004).

The Bayesian account has also been extended to incorporate source reliability (Hahn, Harris, & Corner, 2009; Oaksford & Hahn, 2013). This extension can account for variations in expert opinion (Hornikx & ter Haar, 2013), which we also varied in the study we report here. The odds version of Bayes theorem, factoring in the probability that a source is reliable, ($\Pr(R)$), is:

$$O(C|a) = \left(\frac{\Pr(a|C, R)\Pr(R) + \Pr(a|C, \neg R)(1 - \Pr(R))}{\Pr(a|\neg C, R)\Pr(R) + \Pr(a|\neg C, \neg R)(1 - \Pr(R))} \right) . O(C) \qquad \text{(Eq. 2)}$$

When someone is an unreliable source of information, we assume they are equally likely to deploy the argument whether the conclusion is true or false, that is, $\Pr(a|C, \neg R) = \Pr(a|\neg C, \neg R) = .5$ (Bovens & Hartmann, 2003). So, when the source is completely unreliable ($\Pr(R) = 0$), the likelihood ratio (LR) = 1, that is, there is no change in degree of belief, $O(C|a) = O(C)$. On the other hand, when the source is completely reliable ($\Pr(R) = 1$), then LR = $\Pr(a|C, R)/\Pr(a|\neg C, R)$.

2 Culture and Argumentation

Recent work on cross-cultural differences in argumentation has focused on evidence quality (Hornikx & Hoeken, 2007; Hornikx & ter Haar, 2013). The dimensions examined include statistical (high quality) vs anecdotal (low quality) evidence and evidence from experts with domain-relevant knowledge (high quality) vs experts with no domain-relevant knowledge (low quality). This research has shown that superficially quite similar, and geographically close cultures show marked differences in how high and low-quality evidence effects how persuaded they are of a conclusion. For example, French participants show less sensitivity to domain-relevant expert opinion than Dutch participants (Hornikx & Hoeken, 2007). This finding was explained by French students' greater obedience to authority figures (Hornikx, 2011) and it has

been replicated for Dutch vs Indian participants (Hornikx, & De Best, 2011). Moreover, German participants are less sensitive to the quality of statistical evidence than Dutch participants (Hornikx & ter Haar, 2013). This finding was consistent with German culture's apparently higher levels of uncertainty avoidance (Hofstede, 2001). However, individual measures of uncertainty avoidance did not correlate with the difference in the acceptability of high and low-quality arguments (Hornikx & ter Haar, 2013).

Hornikx and ter Haar (2013), argue that their findings may arise because different cultures adhere to different norms. They suggest that "It is still an open question as to whether the norms are universal and people's reactions to them are culture-dependent, or as to whether norms themselves may be culture dependent." (Hornikx & ter Haar, 2013, p. 498). As we have suggested, there are different normative accounts of argumentation, and these tend to depend on how the norms they postulate are justified (Corner & Hahn 2013). Procedural norms, like those governing courtroom proceedings, are social conventions to which societies, over the course of their development, have assented to obey. It is difficult to conceive of such norms as universal standards obeyed by members of all cultures (Corner & Hahn 2013). In contrast, Bayesian argumentation has two routes to normativity. First, the self-evidence of the Kolmogorov axioms of probability theory which are used to derive Bayes theorem. Second, the Dutch book argument, which shows that following the rules of probability will prevent a rational agent making gambles that would lead to sure losses (see, Corner & Hahn, 2013). It would seem hard to conceive of a culture in which avoiding sure losses is not something one ought to do (Oaksford, 2014).

However, the evidence we have just reviewed showed cross-cultural differences in how persuasive members of different cultures find high and low-quality evidence. What constitutes high and low-quality evidence is based on epistemic norms (Hornikx & ter Haar, 2013), which are derivable from probability theory. For example, there is an obvious Bayesian account of statistical (50 trials) vs anecdotal (1 trial) evidence (Hahn & Oaksford, 2007; Oaksford & Hahn, 2004). Bayes theorem is a model of learning such that as each piece of positive evidence is used iteratively to update the posterior, it converges to 1. Moreover, as we have seen, source reliability can be incorporated into a Bayesian analysis (Hahn, Oaksford & Bayindir, 2005; Hahn, Harris, & Corner, 2009; Hahn & Oaksford, 2007; Oaksford & Hahn, 2013). Experts with domain-relevant knowledge are more reliable sources of evidence than experts with no domain-relevant knowledge. If argument quality, as investigated by Hornikx and ter Haar (2013), is underpinned by a universal Bayesian norm, then why were there clear cross-cultural differences?

As we have observed, Bayes theorem is a mathematical function which prescribes how probabilistic sources of information should combine. It is normative for argument quality because (i) the more evidence that accrues in favour of a hypothesis the higher the posterior probability, and (ii) the higher the probability that a source is reliable, then, again, the higher the posterior probability. As a mathematical function, Bayes theorem only determines the relationship between the input and the output. So if either the prior, the likelihood, or the probability that the source is reliable go up, while the others stay the same, so does the posterior probability. Consequently, for any experimental manipulation that a participant interprets as increasing, say, reliability while not affecting the prior or the likelihood ratio, should, according to Bayes theorem, lead to an increase in the posterior and so to an increase in how persuasive they find the argument. However, Bayes theorem does not determine the value of any of these input variables. So, some cultures may value expert opinion more than others, which we can capture as differences in Pr(R). But as long as they combine these values according to Bayes rule, each culture would still be respecting the appropriate epistemic norm. So, on this view, any cultural difference would be explained by culture-dependent inductive biases (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010) rather than by the adoption of different epistemic norms.

3 Evidentiality and Argumentation

As we have indicated, we chose to test this account of cross-cultural differences between Turkish and English cultures because they appear to differ in how they linguistically mark evidence quality. Two caveats are worth making at the outset.

First, there has been much discussion of the difference between obligatory/morphological and voluntary/lexical marking of evidence in the literature on linguistic relativity (Gerrig & Banaji, 1994; Lucy, 1992; Robinson, 2009; Slobin, 2003). However, it is difficult to determine whether a language has developed obligatory marking of evidentiality because it is important in that culture or it is important in that culture because the language has obligatory marking. We remain neutral on whether any cognitive differences are because of language differences or cultural differences.

Second, linguistic marking of evidence is not necessarily evaluative; it may just be informative. The reason natural languages have ways of marking evidence may be just to inform about how the evidence was acquired. People do

not interpret this information as conveying value judgements about the quality of the evidence. We resolve this question experimentally. If direct (eye-witness) marking of evidence leads participants to assign higher argument strength to a conclusion, then one would have to conclude that such linguistic marking has an evaluative role. But we also note that it is conceivable that, in ordinary usage, the evidential system in one linguistic culture is purely informative, but in another it is evaluative. We say "in ordinary usage" because in the Anglo-Saxon legal system (in the USA and UK) the rules of evidence strongly favour eye-witness over hearsay evidence. But this does not mean that, outside the formal context of the courtroom, people ordinarily treat this distinction as consequential.

The Turkish Evidential System. In Turkish, four verb suffixes are used to mark evidential distinctions (Aikhenvald, 2004). "–DI"[2] marks evidence acquired by direct perceptual experience. "–mIş" marks evidence acquired by inference based on available evidence. "–(I)mIş" marks evidence acquired from someone else's testimony. "–DIr" marks evidence acquired by deduction without immediate evidence, that is, using prior knowledge (Lucas et al., 2013). An example from Lucas et al. (2013, pp. 579–580) makes these distinctions clear:

> By way of example, in English, it is acceptable to make unqualified assertions such as, "There was a storm." In Turkish, however, it is obligatory also to mark whether the storm was directly witnessed (–DI), whether it was inferred from the aftermath (–mIş), whether one was told there was a storm (–(I)mIş), or whether general climate knowledge leads one to believe there should have been a storm (–DIr).

Work on linguistic relativity focuses on the possible effects of language differences on non-linguistic processes. Turkish speaking children appear to successfully identify the correct source of a message at an earlier age than English-speaking children (Ögel-Balaban, Aksu-Koç, & Ercan, 2012). In adults, this effect seems to be due to the specific marking of sources in the evidential system (Tosun et al., 2013). For Turkish speakers, propositions encoded with

2 In "–DI" / "–DIr", the capital "D" indicates that it can take different values based on the consonant harmony rule of grammatical suffixes: the consonant of the suffix changes to harmonize with the syllable-final consonant of the stem, "çarptı" ('it crashed') vs. "sevdi" ('s/he loved'). Similarly, the capital "I" indicates that it can take different values based on the vowel harmony rule of grammatical suffixes: the vowel of the suffix changes to harmonize with the vowel of the preceding syllable (front/back or rounding harmony), "faydalıydı" ('it was beneficial') vs. "yeşildi" ('it was green'), or "faydalıdır" ('it is beneficial) vs. "yeşildir" ('it is green').

— DI (i.e., direct evidence) showed better recognition and source memory than those encoded with — mIş (i.e., indirect evidence). English speakers showed no such differences in recognition or source memory when provided with lexically marked direct or indirect evidence. Moreover, Turkish children (pre-schoolers) seem to have an advantage over Chinese and English children in selective trust and false belief tasks (Lucas et al., 2013). Turkish children were more likely to trust a reliable informant than Chinese and English children, and they were more likely to pass false belief tasks. The habitual use of obligatory evidential marking for Turkish speakers was hypothesised to sensitise Turkish speakers to attend to trustworthy sources of information (Lucas et al., 2013). Turkish children also appear to resist misinformation better in a suggestibility task (Gudjonsson, 1984) because they ignore information from indirect sources and attend to information from direct sources more than English children (Aydin & Ceci, 2013). However, there have been failures to show source memory effects for Korean, which also has an obligatory evidential system (Papafragou et al., 2007), and for Turkish (Ünal et al., 2016) compared to English speakers.

In this study, as we have said, we are neutral on the question of linguistic relativity. We treat this linguistic difference in how evidence is marked in the same way as other researchers have treated the greater respect for authority shown by the French (Hornikx & Hoeken, 2007) and the higher uncertainty avoidance showed by Germans (Hornikx & ter Haar, 2013), compared to the Dutch. That is, we treat these differences as strongly suggestive that these cultures will show differences in how they treat high and low-quality evidence in evaluating arguments. However, in the particular case of the different evidential systems in Turkish and English, we can also linguistically manipulate whether evidence is direct or indirect and see if there are any differences in argument evaluation.

Anticipating some of our results, a pilot experiment, reported in the *Method* section, which concentrated solely on expert opinion, revealed cross-cultural English-Turkish differences. Turkish participants found arguments from experts more persuasive and arguments from lay people less persuasive than their English counterparts. This result rules out a possible pragmatic explanation of our findings such that the English are simply more polite than the Turkish (Bhatia & Oaksford, 2015; Brown & Levinson, 1987; Leech, 1983; van Eemeren, Garssen, & Meuffels, 2009). This hypothesis can only push in one direction; it must predict that English participants will be more accepting of arguments from any source. So, it can explain why English participants were more accepting of arguments from lay people but not why they were less accepting of arguments from experts.

4 Introduction to the Current Experiment

In this experiment, we used argumentative dialogues like the following (Hahn et al., 2005):

(3) Margaret: Do you think clone technology is a threat to human beings?
Anton: I sort of believe that clone technology is not a threat to human beings.
Margaret: You can do more than sort of believe it; you can be certain that it is not a threat.
Anton: Why do you say that?
Margaret: Because they *apparently* had an interview with a professor who is an expert in his field in an academic journal including scientific studies. I *gather* the professor expressed his opinion considering that the long-term effects of cloning are not known to a great extent. According to him, *it seems* cloning was dangerous for human beings.

After the dialogue, participants were asked to rate how convinced Anton should now be that clone technology is a threat to human beings. We used a third-person argument evaluation paradigm (see, also Bhatia & Oaksford, 2015; Oaksford & Hahn, 2004, 2013) because we did not want to confound participants' own prior beliefs with those of Anton. Using these third-person judgments should dissociate participants' assessments of the beliefs of an interlocutor in the experimental dialogues from their own prior beliefs. We hoped to ensure that participants would attend to the prior belief manipulation (see next paragraph) rather than attribute their own prior beliefs to the relevant interlocutor in the dialogues (Anton in (3)). We tested this assumption in a pre-test for the pilot experiment we report in the *Method* section.

The argument in (3) is a negative argument like the argument from ignorance in (1) (substituting "not toxic" for "safe"). We presented both negative and positive versions of these arguments. For positive versions, the strength of the argument, according to the Bayesian epistemic approach, $\Pr(C|a)$, is given in Eq. 1. Using the same parameters, $\Pr(a|C)$, $\Pr(a|\neg C)$, and $\Pr(C)$, we can also compute the strength of a corresponding negative argument, $\Pr(\neg C|\neg a)$. We also manipulated one of the interlocuter's, in this case, Anton's, prior degree of belief in the conclusion. In (3), Anton has a weak prior degree of belief, he only "sort of believes" the conclusion initially. We also used the expression, "fairly convinced" to express a strong prior degree of belief.

We introduced further manipulations in the final paragraph of this argumentative dialogue from Margaret. This paragraph concerns a highly reliable source of information, a Professor in this field of research (expert opinion). Participants also evaluated arguments from sources with low, and medium-reliability. For example, a TV street interview of a passing couple (lay opinion) would be expected to have low reliability. We can also mark the information provided by Margaret in the final paragraph as direct or indirect, morpho-syntactically in Turkish or lexically in English. The example in (3) is marked as indirect. In the Turkish dialogues, the suffix –(I)mIş was used three times in the relevant verbs. In the English version in (3), following the advice in Göksel and Kerslake (2005), a comprehensive grammar of Turkish, the terms "apparently", "I gather", and "it seems" were lexically incorporated in the text. We also included a neutral condition which was not marked for evidentiality. In the final statement in each dialogue, the speaker, here Margaret, appeals to a third person source of evidence. As the speaker is not asserting that they are the source of the evidence, it is not obligatory for them to mark how it was obtained as they simply may not know. The neutral condition provides a control revealing the effect of source reliability independent of evidentiality markers. Having this control would not have been possible with a first-person report as the neutral case would have been ungrammatical.

5 Experimental Hypotheses

Previous research (e.g., Hahn & Oaksford, 2007; Hahn et al., 2005; Oaksford & Hahn, 2004) has shown that people are sensitive to the factors predicted by Bayes theorem to affect how persuaded they are by arguments like those in (3). The first three predictions relate to replicating previous results. Bayes rule predicts that the degree of belief assigned to Anton after hearing the argument should be higher if he is already fairly convinced of the conclusion. So the higher the prior, the higher the argument strength. Moreover, as long as specificity is higher than sensitivity, positive arguments should be more persuasive than negative arguments (Hahn & Oaksford, 2007; Oaksford & Hahn, 2004). Finally, an argument should be perceived as stronger the more reliable the source. Consequently, the first three experimental hypotheses were:

(H1) There will be a main effect of prior belief such that strong prior beliefs lead to higher argument strength than weak prior beliefs.
(H2) There will be a man effect of polarity such that positive arguments are perceived as stronger than negative arguments.

(H3) There will be a main effect of reliability such that high > medium > low-reliability arguments (where ">" = has higher argument strength than)

If these three hypotheses are confirmed, then the experiment will have replicated previous findings using similar argument forms. The next set of predictions concern the new culture and evidentiality manipulations introduced in this experiment.

First, the existing evidence shows that Turkish people are more trusting of reliable sources and less trusting of unreliable sources than English people (Aydin & Ceci, 2013; Lucas et al., 2013; Tosun et al., 2013). However, some of this evidence derives from children, and it is a big jump to inferring adult performance from such results. We therefore also conducted a pilot experiment, which, as we will see in the *Method* section, supported the following hypothesis based on the prior observations with children:

(H4) Turkish participants will treat arguments from high-reliability sources as stronger than English participants, but they will treat arguments from low-reliability sources as weaker than English participants.

We pretested the third-party sources, appealed to the final paragraph of our dialogues, for reliability (see, *Method* section). For the medium-reliability group, we made no direct predictions.

Second, varying evidentiality, morpho-syntactically in Turkish and lexically in English, in a variety of paradigms (Aydin & Ceci, 2013; Tosun et al., 2013), appears to be ignored by English participants but leads to differential performance for Turkish participants. We, therefore, made the following prediction.

(H5) While Turkish participants will treat arguments marked as from a direct source as stronger than from an indirect source, English participants will not.

The purpose of the neutral control was to ensure that we could separate out the contributions of evidentiality and reliability on judgements of argument strength. We predicted that Turkish participants will treat the unmarked, neutral case as the least persuasive.

Our final hypothesis concerns our original research question. Do these two diverse cultures employ the same epistemic Bayesian norm in evaluating these arguments? We addressed this question by fitting the Bayesian model in Equation 2 to these data.

(H6) If the same epistemic norm is applied, we would expect to find equally good fits to the experimental data. However, we would not expect these different cultures to assign the same values to the various parameters of the model reflecting cross-cultural inductive biases.

6 Analysis Strategy

Our analysis strategy was to conduct an overall within subjects ANOVA with culture as a between-subjects factor and carry out planned contrasts to test our various hypotheses. However, the experiment used several different topics and reliability manipulations resulting in twelve different set of materials, with each participant seeing only one set. Because of this variation in materials, we also double checked our results using a Bayesian linear mixed effects regression model implemented in the **rstanarm** package in R (Gabry & Goodrich, 2016). We treated participants and materials as random effects and fitted the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). This model was unidentifiable using all five factors. Consequently, to assess the novel experimental hypotheses for reliability and evidentiality, we left out the prior and polarity factors (see, for example, Singmann, Klauer, & Beller (2016) for a similar approach to this issue). The maximal random effects model fitted was: Argument Strength ~ Evidentiality × Reliability × Culture + (Evidentiality × Reliability|Participant) + (Evidentiality × Reliability × Culture|Materials). These analyses used the relatively uninformative default priors for the **stan_lmer** function in the **rstanarm** package. We report the results of these additional analyses fully in the supplementary materials (Supplementary Materials 1: Main Experiment Bayesian Analysis). We only introduce the Bayesian analysis if the results disagree on the predictions based on the experimental hypotheses.

7 Experiment

7.1 *Method*

Participants. The participants were 114 Turkish students (Female = 74, mean age = 19.03 years) from Başkent University (59 students) and Middle East Technical University (55 students), in Ankara, Turkey, and 100 native English speaking participants recruited via Prolific Academic (https://www.prolific.ac) and paid £3 for completing the experiment online (Female = 54, mean age = 33.70 years). Online recruitment was restricted to those whose first language

was English, and who held U.K. or U.S. nationality. This experiment compared cultures rather than language users. Consequently, a mixed UK/US sample may be considered not to be a pure one culture sample. However, the demographics of the Prolific Academic database indicates that 71% of participants are from the UK and only 29% from the US. Moreover, it is unlikely that the UK and the US citizens, while culturally different along many dimensions, have evolved very different ways of regarding evidence, as expressed in their shared language, given that they share a common Anglo-Saxon legal framework.

The English speaking internet sample was on average 13.68 years older than the Turkish sample. However, anticipating the results, they showed less sensitivity to evidential distinctions than their younger Turkish counterparts when additional experience would predict greater sensitivity. There were more females in the Turkish than in the English sample. However, at 65% and 54% respectively this difference was unlikely to be consequential. The English sample was recruited over the Internet, whereas the Turkish sample conducted the study in class. However, even for demanding cognitive and perceptual experiments, there are no performance differences between self-selected internet samples and traditionally recruited and/or lab-tested samples (Germine, Nakayama, Duchaine, et al., 2012). Moreover, as it was the only between subject variable, only language could be affected by differential drop out (Birnbaum, 2000). But there was no drop out in either sample. The different modes of recruitment might lead to the two groups being differentially attentive. However, any such effect could go either way. On the one hand, the Turkish sample may be less attentive as they were distracted by the class setting. On the other hand, the English internet sample may be less attentive as they were motivated by financial incentives and the desire to finish the task. Consequently, any such effects would be expected to counterbalance each other.

Design. The design was mixed, with 2 (prior belief: weak versus strong) × 2 (polarity: positive versus negative) × 3 (source reliability: high, medium, low) × 3 (evidentiality: direct (–DI), indirect (–mIş), neutral) as within-subjects factors, and Culture (Turkish versus English) as a between-subjects factor. The dependent variable was argument convincingness measured on an eleven-point convincingness scale 0 (not convinced at all) to 10 (totally convinced).

Pilot Experiment. A pilot experiment had three goals (for the complete experiment see Supplementary Material 2: Pilot Experiment). First, to confirm that Hypothesis 4 had some credibility. Second, to trial materials for the main study. Third, to confirm that participants' own priors did not interfere with the prior probability manipulation.

Two preliminary studies were carried out before the pilot experiment. First, 33 other participants rated a range of materials, for example, they were asked

how beneficial they found clone technology. Topics were selected which had mean values in the mid-range, that is, topics that did not polarise opinions one way or the other so that both positive and negative arguments made sense for the same topic. Four dialogues were selected on the following topics: the dangers of cloning, the dangers of globalisation, the efficacy of capital punishment, and the efficacy of using robots instead of hiring people in the workplace. The Turkish versions of these dialogues used "–DIr," e.g., "faydalı değildir," indicating the evidence is inferred from general knowledge. Second, four days before the administration of the pilot experiment, the same participants were presented with the conclusions of these dialogues, for example, "Cloning is beneficial for humanity." They were then asked whether they agree on an 11-point Likert scale from *not agree at all* (*0*) to *completely agreed* (*10*) reflecting their own priors.

The third goal was confirmed. There was only one significant correlation, out of 16 possible, between participants' own prior beliefs and those they attributed to the relevant interlocutor in the dialogues. The second goal was also, largely, confirmed. These materials replicated previous studies (Hahn et al., 2005; Oaksford & Hahn, 2004), that is, hypotheses H1 to H3, apart from an effect of polarity (H2). In particular, given the novel experiment hypotheses tested in the current experiment, there was a strong effect of reliability (H3). To assess the first goal, the results of this pilot experiment were compared to the results from Hahn et al. (2005). Turkish participants treated arguments from high-reliability sources as stronger and arguments from low reliable sources as weaker than the English participants in Hahn et al.'s (2005) study (see, Figure 1). This comparison provided tentative support for Hypothesis 4. However, apart from the clone technology topic, different materials were used in Hahn et al. (2005) and the pilot study. In the current experiment, the same materials were used in both language groups.

Materials. The materials were the same as those trialled in the pilot experiment except that, in the current experiment, they included — DI suffixation, e.g., *-faydalıydı*, and — mIş suffixation, e.g., *-faydalıymış*, in the final paragraph. A neutral case was also included, e.g., *-faydalı*. Suffixation was introduced in the verbs in the closing statement of each dialogue where one interlocutor makes the primary argument for why the other should or should not believe the conclusion. Long sentences were split into two or three to include more instantiations of evidentiality markers in each dialogue to make them more salient. No lexical items or adverbs (e.g. güya [allegedly]) as signs of evidentiality were used to avoid confounding lexical effects with the effects of the morpho-syntactic markers under investigation. (All the Turkish materials and English translations can be found in Supplementary 3: Experimental Materials).

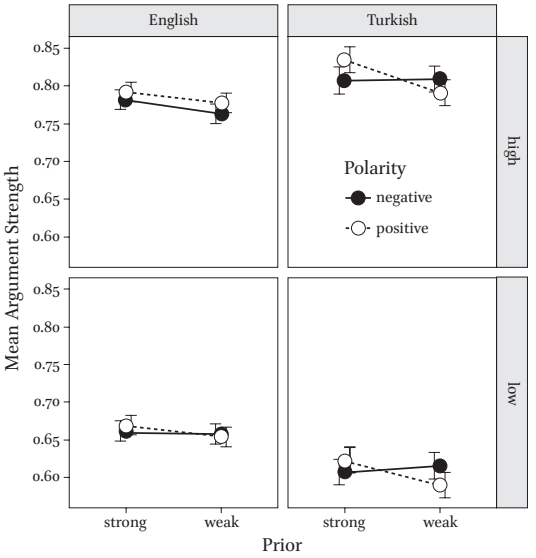

FIGURE 1
Mean argument strength for positive and negative arguments in high and low-reliability conditions with strong and weak prior beliefs in the pilot experiment (Turkish) and in Hahn et al.'s (2005) data (English). Error bars are 95% Highest Density Intervals.

For the English translations of the neutral case, the evidential relationships were not stated. In the indirect case, all these evidential relations were marked as indirect. As in (3), the terms "apparently", "I gather", and "it seems" were used in place of the three occurrences of the indirect marker used in Turkish (see, Göksel and Kerslake, 2005). Three different lexical items were included expressing indirect evidence because using one term repetitively in English sounded infelicitous. In the direct case, it was made clear that the sources carried out the study and are reporting on their own direct observations or results. The description is in the past tense and with factual modality. All translations were checked with Turkish-English bilinguals on the research team.

Source Reliability. In this experiment, in addition to high and low-reliability sources, a medium-reliability source was also introduced. This source was felt necessary to avoid possible ceiling effects of reliability such that there was no room to see effects of the evidentiality markers. A survey study was therefore conducted before the experiment, to identify information sources with varying levels of reliability. 35 participants, all native speakers of Turkish currently living in Turkey, were contacted by e-mail and requested to list as many sources of information as possible under three categories: the most reliable, moderately reliable, and least reliable. To see if the three categories identified in the survey differed 45 students (Mean Age = 19; SD = 0.8) from Çankaya University, were asked to rate 15 items (see Appendix A) from this pool for reliability.

Students' ratings confirmed the e-mail results and the categorisation of information sources with just two differences. The item "local newspapers" was

placed in the medium-reliability category and the item "poll/survey companies" was placed in the low-reliability category. The mean reliability ratings are shown in Appendix A.

When appropriately reclassified, there were significant difference between all levels of source reliability: the high reliability category (mean = 7.16, *SE* = 0.30) was found to be significantly more reliable than the medium-reliability category (mean = 4.87, *SE* = 0.45) and both categories combined were found to be significantly more reliable than the new low reliability category (Mean = 2.78, SE = 0.26), $F(2, 14) = 38.11$, $p < .0001$. All three levels were used in this experiment. Sources were selected from the list in Appendix A in each category and combined with topic to provide three different versions of each topic using different high, medium, and low-reliability sources. So there were effectively 12 different sets of materials.

The inclusion of a medium-reliability level would have meant that participants had to evaluate 144 dialogues, which was considered excessive and likely to lead to fatigue effects. We, therefore, treated materials as a between-subjects variable. Each participant saw all thirty-six conditions of interest but with only one of the combinations of topic and information sources.

Procedure. For the Turkish speaking participants, the dialogues were presented in a booklet. Participants rated the extent to which one of the interlocutors should now believe the conclusion. The booklet took about 20 minutes to complete, and participants were tested during their classes (without talking to each other) in the presence of their instructors and the experimenter. Before the experiment, participants signed an informed consent form. For the English participants, the dialogues were implemented in Qualtrics (https://www.qualtrics.com/), and participants conducted the study online with participants recruited via the Prolific Academic platform (https://www.prolific.ac/).

7.2 *Results*

We first converted the argument convincingness rating scale into the 0–1 probability scale.[3] This scale was used because we subsequently modelled the data and converting means that we can report the results of the data analysis and modelling on the same scale. We first conducted a mixed ANOVA with 2 (prior belief: weak versus strong) × 2 (polarity: positive versus negative) × 3

3 To make this conversion the ratings were transformed by adding ten and dividing by 20. One of the interlocuters was always initially described as believing the conclusion, whether it was a negative ($\neg C$) or a positive (C) argument, i.e., the lowest value the prior could take was .5. Consequently zero on this rating scale is the midpoint, i.e., corresponding to 0.5, on the probability scale. The same transformation was applied in Hahn and Oaksford (2007).

(reliability: high, medium, low) × 3 (evidentiality: direct (–DI), indirect (–mIş), neutral) as within-subjects factors and culture (Turkish versus English) as a between-subjects factor.

Hypothesis 1: Priors. There was a main effect of prior belief. When the prior was strong (Mean = .706, SE = .006, 95% CI [.694,.719]), argument strength (the posterior degree of belief) was higher than when it was weak (Mean = .690, SE = .006, 95% CI [.679,.701]), $F(1, 212) = 48.98$, $MSe = .01$, $\eta^2 = .19$, $p < .0001$. This effect was replicated for Turkish participants (Strong: Mean = .690, SE = .008, 95% CI [.673,.707]; Weak: Mean = .671, SE = .007, 95% CI [.657,.685]); $F(1, 113) = 27.13$, $MSe = .01$, $\eta^2 = .19$, $p < .0001$), and for English participants, (Strong: Mean = .723, SE = .010, 95% CI [.704,.742]; Weak: Mean = .709, SE = .009, 95% CI [.691,.727]); $F(1, 99) = 24.21$, $MSe = .01$, $\eta^2 = .20$, $p < .0001$).

Hypothesis 2: Polarity. There was a main effect of polarity. When the argument was positive (Mean = .708, SE = .006, 95% CI [.695,.720]), argument strength was higher than when it was negative (Mean = .689, SE = .006, 95% CI [.677,.701]), $F(1, 212) = 21.94$, $MSe = .03$, $\eta^2 = .09$, $p < .0001$. This effect was replicated for English participants (Positive: Mean = .733, SE = .010, 95% CI [.713,.754]; Negative: Mean = .698, SE = .010, 95% CI [.680,.717]); $F(1, 99) = 20.76$, $MSe = .05$, $\eta^2 = .17$, $p < .0001$), but it was not for Turkish participants (Positive: Mean = .682, SE = .008, 95% CI [.667,.697]; Negative: Mean = .679, SE = .008, 95% CI [.664,.695]); $F(1, 113) < 1$), although the trend was in the right direction.

Hypothesis 3: Reliability. As predicted there was a significant linear trend for reliability. When reliability was high (Mean = .794, SE = .007, 95% CI [.780,.808]), argument strength was higher than when it was medium (Mean = .657, SE = .007, 95% CI [.643,.670]), which was higher than when it was low (Mean = .644, SE = .006, 95% CI [.630,.657]), $F(1, 212) = 373.99$, $MSe = .02$, $\eta^2 = .64$, $p < .0001$. The quadratic component was also significant, $F(1, 212) = 332.81$, $MSe = .02$, $\eta^2 = .61$, $p < .0001$, indicating that, for these materials, as reliability reduces, its effects on the posterior level off quite rapidly. All pairwise comparisons were signifciant at at least $p < .0001$. This finding was replicated for the Turkish participants (High: Mean = .802, SE = .010, 95% CI [.782,.822]; Medium: Mean = .628, SE = .009, 95% CI [.610,.645]); Low: Mean = .612, SE = .009, 95% CI [.595,.630]),]), Linear trend: $F(1, 113) = 261.27$, $MSe = .02$, $\eta^2 = .70$, $p < .0001$; Quadratic trend: $F(1, 113) = 233.49$, $MSe = .01$, $\eta^2 = .67$, $p < .0001$), and for English participants (High: Mean = .786, SE = .009, 95% CI [.768,.805]; Medium: Mean = .686, SE = .011, 95% CI [.664,.707]); Low: Mean = .676, SE = .011, 95% CI [.655,.697]),]), Linear trend: $F(1, 99) = 127.73$, $MSe = .01$, $\eta^2 = .56$, $p < .0001$; Quadratic trend: $F(1, 99) = 112.62$, $MSe = .01$, $\eta^2 = .53$, $p < .0001$). For both Turkish and English participants, all pairwise comparisons were significant at at least $p < .05$.

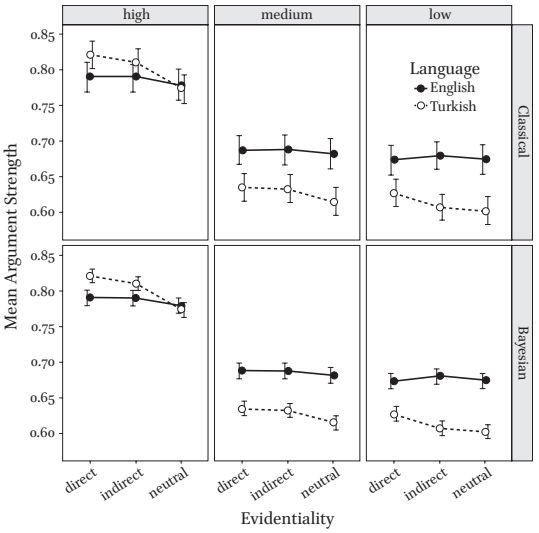

FIGURE 2 Mean argument strength for Turkish and English speakers in high, medium, and low-reliability conditions with direct, indirect, and neutral evidentiality. For the Classical statistical analysis, error bars are the classical 95% confidence intervals. For the Bayesian analysis, error bars are the 95% highest density intervals.

These analyses confirmed Hypotheses 1 to 3 and replicated previous research using these argument forms (Hahn & Oaksford, 2007; Hahn et al., 2005; Oaksford & Hahn, 2004), apart from the lack of a significant polarity effect for Turkish participants. However, neither the polarity nor the prior manipulations were our primary focus. The polarity effects were also small, albeit detectable, in previous research using similar materials (Hahn & Oaksford, 2007; Hahn et al., 2005). Moreover, the trend was in the right direction. Consequently, we now focus on the novel predictions for reliability, evidentiality and culture.

Figure 2 shows the mean argument strength for Turkish and English participants by evidentiality and by reliability showing the classical 95% confidence intervals in the upper panels. The lower panels show the results of the hierarchical Bayesian analysis based on the posterior predictive sample derived from the maximal random effects model and show the 95% highest density intervals.

Hypothesis 4: Culture and Reliability. To test Hypothesis 4, we first carried out planned contrasts comparing the two cultures at each level of reliability (the values of the relevant means are shown above). For both low and

medium-reliability conditions, Turkish participants found the arguments significantly less convincing than English participants (Low: $F(1, 212) = 21.74$, $MSe = .01$, $\eta^2 = .09$, $p < .0001$; Medium: $F(1, 212) = 17.46$, $MSe = .01$, $\eta^2 = .09$, $p < .0001$). For the high reliability conditions, Turkish participants found the argument *more* convincing than English participants, but this result was not significant, $F(1, 212) = 1.21$, $MSe = .01$, $\eta^2 = .01$, $p = .27$). However, in the Bayesian analysis, the posterior probability that this difference in argument strength, δ, was zero or less was zero, within the accuracy reported in R (mean difference, $\delta = .015$, HDI = [.007,.024], $Pr(\delta \leq 0) = 0$). The reason for the discrepancy is obvious from Figure 2. In the hierarchical Bayesian model, although the means are the same, because of shrinkage (Kruschke, 2010), the HDIs are a lot narrower than the classical confidence intervals (CIs). The HDI is the interval which includes the mean with a probability of.95 given the data, an interpretation that cannot be given to classical CIs. So, these data provide support for Hypothesis 4. Moreover, on either analysis, English participants were not more convinced than Turkish participants in the high-reliability condition, thereby ruling out a general politeness explanation of these results (Bhatia & Oaksford, 2015; van Eemeren, Garssen, & Meuffels, 2009).

Hypothesis 5: Evidentiality. As predicted there was a significant linear trend for evidentiality. When evidentiality was marked as direct (Mean = .706, SE = .006, 95% CI [.694,.718]), argument strength was higher than when it was indirect (Mean = .701, SE = .006, 95% CI [.690,.713]), which was higher than when it was neutral (Mean = .688, SE = .006, 95% CI [.675,.700]), $F(1, 212) = 48.21$, $MSe = .002$, $\eta^2 = .19$, $p < .0001$. The quadratic component was also significant, $F(1, 212) = 6.49$, $MSe = .001$, $\eta^2 = .03$, $p < .025$, indicating that, for these materials, the effect on the posterior of the unmarked, neutral, case falls away rapidly compared to the two marked cases. All pairwise comparisons were signifciant at at least $p < .05$. This finding was replicated for the Turkish participants, apart from the quadratic trend (Direct: Mean = .694, SE = .008, 95% CI [.679,.710]; Indirect: Mean = .683, SE = .007, 95% CI [.669,.698]; Neutral: Mean = .664, SE = .008, 95% CI [.647,.680; Linear trend: $F(1, 113) = 54.16$, $MSe = .003$, $\eta^2 = .32$, $p < .0001$; Quadratic trend: $F(1, 113) = 3.11$, $MSe = .001$, $\eta^2 = .03$, $p = .08$). For Turkish participants, all pairwise comparisons were significant at at least $p < .001$. However, for English participants, neither trend was significant (Direct: Mean = .717, SE = .009, 95% CI [.699,.735]; Indirect: Mean = .719, SE = .009, 95% CI [.701,.737]); Neutral: Mean = .712, SE = .010, 95% CI [.692,.731]; Linear trend: $F(1, 99) = 3.53$, $MSe = .001$, $\eta^2 = .03$, $p = .063$; Quadratic trend: $F(1, 99) = 3.53$, $MSe = .001$, $\eta^2 = .03$, $p = .070$). Moreover, for English participants, none of the pairwise comparisons were significant, $p > .076$ for all comparisons.

The effects for the high reliability condition alone showed a significant interaction between evidentiality and culture, $F(2, 424) = 9.28$, $MSe = .01$, $\eta^2 = .04$,

p <.0001 (see, Figure 2). In particular, when evidentiality was neutral there was no difference in argument strength between Turkish and English participants. This pattern of interaction suggests that when reliability is high, Turkish participants do not judge the source to be more reliable than English participants. Rather, in this condition, they find an argument more convincing than English participants because of the presence of the evidentiality marker.

These analyses confirmed Hypotheses 4 and 5. Turkish participants found arguments from high-reliability sources more convincing and arguments from medium and low-reliability sources less convincing than English participants. Moreover, only Turkish participants showed an effect of evidentiality. They found arguments with evidence marked as direct more convincing than arguments with evidence marked as indirect and arguments with evidence marked as indirect more convincing than arguments when evidence was unmarked or neutral. We did not observe similar effects for English participants. The interaction for the high-reliability condition modified these hypotheses suggesting that the effect of culture in this condition was due to the presence of the morpho-syntactic evidentiality markers. These results show that in Turkish, but not in English culture, marking evidentiality is treated evaluatively influencing how strongly one should believe the conclusion of an argument.

Hypothesis 6: Model Fitting. To test this hypothesis, we fitted the Bayesian model in Eq. 2 to the data. Previous model fitting for similar data (e.g., Hahn & Oaksford, 2007) only fitted the pooled data, that is, the means, which can lead to overfitting and a lack of generalizability to new data. Moreover, it does not allow comparisons between the parameters of the model. We, therefore, adopted a cross-validation approach to address the issue of overfitting and also fitted the unpooled data for each participant. We used eleven free parameters to model the 36 data points for each participant. Separate sensitivity ($\Pr(a|C, R)$) and specificity ($\Pr(\neg a|\neg C, R)$) parameters were fitted for the three separate levels of evidentiality, making six parameters in all. We did this because we modelled the effects of evidentiality as affecting the likelihood ratio ($\Pr(a|C, R)$)/(1 - ($\Pr(\neg a|\neg C, R)$). We also included three reliability ($\Pr(R)$, high, medium, low) and two prior degrees of belief parameters ($\Pr(C)$, weak vs strong). The priors were constrained such that $\Pr(C)$ or $\Pr(\neg C)$ >.5. As we observed in describing the transformation to the probability interval (Footnote 3), the first interlocutor (e.g., Margaret in (3)) was always described as initially believing the conclusion to some extent ($\Pr(C)$ or $\Pr(\neg C)$ >.5), rather than disbelieving it ($\Pr(C)$ or $\Pr(\neg C)$ <.5). We used the *optim* function in R (Nash, 2014) to minimise the residual sum of squares between the model's predictions and the data and computed the coefficient of determination,[4] R^2, as an index of fit.

4 $R^2 = 1 - \left(\frac{RSS}{TSS}\right)$, where RSS = residual sum of squares and TSS = total sum of squares.

For the cross-validation analysis, we used the random sampling or Monte-Carlo method (Picard & Cook, 1984). We randomly split the data into two halves for both the Turkish and English speakers. One-half of each group was designated a training-set and the other half a test-set. The best fitting parameter values for the aggregate of the training-set were then computed. We used these parameter values to generate predictions which we evaluated against the aggregate of the test-set. The evaluation was carried out by computing R^2 between the test-set and the predictions. We did this 2000 times for each language group. For Turkish participants, the mean R^2 = .93 (SE = 0.001, median = .94). For English participants, the mean R^2 = .80 (SE = 0.004, median = .86). The Bayesian epistemic model was able to account for a greater proportion of the variance in unseen data for Turkish than for English participants. Nonetheless, the model's generalisation performance was good across both cultures.

Fitting the model to the unpooled data, for Turkish participants, the mean R^2 = .73 (SE = 0.02, median = .78), and the fit was comparable for English participants, mean R^2 = .70 (SE = 0.02, median = .77). The fit was seemingly better for the Turkish participants. However, this result may arise from Turkish participants' greater sensitivity to source reliability, which has led to greater variation in their responses and so a higher total sum of squares and consequently a higher R^2 (see, Footnote 5). We, therefore, calculated the Akaike Information Criterion (AIC) for each participant, assuming the residuals were independent identical normal distributions with zero mean (Akaike, 1974; Burnham & Anderson, 2004).[5] For AIC, if anything, the fit was better for English participants (mean = -97.42, SE = 11.09) than for Turkish participants (mean = -89.03, SE = 12.44), that is, the mean AIC value was lower. In sum, the epistemic norm provided by Bayes theorem seems to adequately characterise how members of both cultures think people should revise their beliefs in informal argumentation.

The final question we addressed was whether the model's parameters behaved as expected given the experimental manipulations and the finding of cross-cultural differences. Models can provide good fits to the data but not necessarily for the right reasons. For both cultures, the prior ($\Pr(C)$ or $\Pr(\neg C)$) was higher in the strong (Turkish: mean $\Pr(C)$ = .572, SE = 0.008; English: mean $\Pr(C)$ = .591, SE = .010) than in the weak prior belief condition (Turkish: mean $\Pr(C)$ = .549, SE = 0.006; English: mean $\Pr(C)$ = .575, SE = 0.009), Turkish: $t(113) = 4.98, p < .0001$; English: $t(97) = 4.16, p < .0001$. However, for the weak prior, $\Pr(C)$ was significantly higher for English than for the Turkish participants, $t(210) = 2.41, p < .025$.

5 On this assumption, AIC = $2k + n.\ln(\text{RSS})$, where k is the number of parameters and n is the number of data points, that is, 12 and 32 respectively for each participant.

We next looked at reliability, Pr(R). For both cultures, Pr(R) was significantly higher in the high (Turkish: mean = .925, SE = .014; English: mean = .895, SE = .014) than in the moderate reliability condition (Turkish: mean = .299, SE = .026; English: mean = .412, SE = .032), Turkish: $t(113) = 18.51$; $p < .0001$, English; $t(97) = 13.64$, $p < .0001$. However, although Pr(R) was higher in the moderate than in the low reliability condition (mean = .242, SE = .033) for Turkish participants, $t(113) = 2.53$, $p < .025$, it was not for English participants (mean = .398, SE = .034), $t(97) = .63$, $p = .53$. There was no significant difference in Pr(R) between Turkish and English participants in the high reliability condition, $t(210) = 1.59$, $p = .11$, but it was significantly lower for Turkish than for English participants in the medium, $t(210) = 2.73$, $p < .01$, and in the low reliability conditions, $t(210) = 3.82$, $p < .0005$. These differences are consistent with the presence of evidentiality markers being the cause of the higher argument strength for Turkish participants than English participants in the high reliability condition.

We assumed that evidentiality affects the likelihood ratio (Pr($a|C, R$))/(1 - (Pr($\neg a|\neg C, R$)). We, therefore, computed the log-likelihood ratio, $\log_{10}$LR, for each evidentiality marker and for each participant.[6] For Turkish participants, $\log_{10}$LR did not differ significantly between the direct (mean = 1.018, SE = .103) and indirect markers (mean = .993, SE = .096, $t(113) = .25$, $p = .81$), but both were significantly higher than in the neutral condition (mean = .666, SE = .077; direct vs. neutral: $t(113) = 3.63$, $p < .0005$; indirect vs. neutral: $t(113) = 3.30$, $p < .005$). For English participants, none to these differences were significant (direct: mean = .912, SE = .080; indirect: mean = .828, SE = .072; neutral: mean = .791, SE = .082; direct vs. indirect: $t(97) = 1.17$, $p = .24$; direct vs. neutral: $t(97) = 1.26$, $p = .21$; indirect vs. neutral: $t(97) = .45$, $p = .65$). For the overall data, there were no differences between cultures. We therefore looked at the participants in the upper 50th percentile of R^2 values (cut-off = .775). For this group, for direct and indirect markers, $\log_{10}$LR was significantly higher for Turkish participants (direct: mean = 1.320, SE = .150; indirect: mean = 1.184, SE = .150) than for the English participants (direct: mean = .791, SE = .100; indirect: mean = .668, SE = .100), direct: $t(101) = 2.72$, $p < .005$, indirect: $t(101) = 3.04$, $p < .005$. There was no significant difference for the neutral case (Turkish: mean = .697, SE = .098; English: mean = .639, SE = .083; $t(101) = .43$, $p = .33$).

The model that emerges is one in which, when there is no marking of evidentiality, the likelihood ratio or measure of argument force (Hahn & Oaksford,

6 We tested whether the log transformed variables were normally distributed using the Shapiro-Wilks test (Royston, 1995). For all three evidentiality markers, $\log_{10}$LR was normally distributed, W ranged between .70 and .81 and $p < .0001$ in each test.

2007), is the same for Turkish and English participants. There is a monotonic increase in argument force, such that direct > indirect > neutral, for Turkish but not for English participants. While both cultures show a monotonic increase in Pr(R), such high > medium > low, Turkish participants assign much lower values of Pr(R) for medium and low-reliability sources than English participants. However, when reliability is high, they assign the same value of Pr(R). In this condition, the increase in argument strength for Turkish participants is caused by increased levels of argument force for the direct and indirect markers.

8 Discussion

Replicating previous research (Hornikx & Hoeken, 2011; Hornikx & ter Haar, 2013), this study revealed significant differences in how two different cultures, Turkish and English, evaluate informal arguments. However, fitting the Bayesian epistemic model to these data, showed that we could explain these differences by culture-dependent inductive biases rather than by these cultures adopting different epistemic norms. These inductive biases consisted of Turkish participants being less trusting of low and medium-reliability sources than English participants, although both cultures were equally trusting of high-reliability sources. An ordering over evidentiality modulated this basic pattern, such that, direct > indirect > neutral, for Turkish but not for English participants. We first discuss possible explanations of the main findings. We then explore the implications for the two issues we outlined in the introduction concerning the universality of the norms of reasoning and argumentation and the role of normative theories in the psychology of reasoning.

That Turkish participants showed a monotonic increase in argument strength with evidentiality seems to be directly related to the Turkish language incorporating these distinctions morpho-syntactically rather than lexically as in English. Linguists like Everett (2013) view language as a cultural tool for communicating efficiently and effectively about the distinctions that a culture finds important (see also, Christiansen & Chater, 2016). It is a reasonable conjecture that the more important a culture finds a distinction, the more likely it is to become sedimented into the obligatory morphosyntactic structure of a language over the course of its evolution. As we argued in the introduction, however, we needed to empirically test whether evidential marking in a language is evaluative or only informative. This experiment showed that Turkish culture treats the distinction between direct and indirect evidence evaluatively. Evidence marked as direct led to greater increases in peoples degree of belief in the conclusion of an informal argument than indirect evidence, that is,

in the Bayesian model, the fitted value of the likelihood ratio was higher. We observed no such effects for English participants.

Previous explanations of Turkish participants greater trust of reliable sources have focused on the notion that habitual and obligatory use of evidential markers sensitises Turkish speakers to source reliability (Lucas et al., 2013). However, our findings did not show that Turkish participants treated high reliable sources as any more reliable than English participants in an argument evaluation paradigm. The principal finding was that Turkish participants were far less trusting (lower Pr(R)) of medium and low-reliability sources than English participants. This result seems to show that English participants were more credulous or gullible than their Turkish counterparts. Social learning appears to depend on a degree of gullibility (Boyd & Richerson, 2007). It is vital to the intergenerational transmission of knowledge and skills that learners are relatively gullible and accepting of the instruction given by their elders. However, it is also important to maintain a degree of scepticism to avoid acquiring silly beliefs (Kurzban, 2007). The balance between gullibility and scepticism may vary between cultures predicting that English participants may be more open to learning from less reliable sources. That is, our results may generalise beyond argument evaluation to a range of learning tasks, an issue that we can address in future research.

These results suggest that the underlying epistemic norm provided by Bayes theorem may be universal for human argumentation. Of course, testing one norm between two linguistic cultures could not constitute an exhaustive test of the universality of this formal rational norm.[7] However, in good falsificationist fashion (Popper, 1959), this experiment did put the hypothesis at risk, i.e., this is a critical test it has survived. Our results, therefore, bear on the debate over the universality of the principles of reasoning and whether particular cultures have fundamentally different ways of thinking and reasoning (Mercier, 2011, 2013; Nisbett, 2003; Nisbett, Peng, Choi, & Norenzayan, 2001; Peng, & Nisbett, 1999). Nisbett and colleagues have argued that Western-individualistic cultures are more analytical and likely to apply rules of reasoning like the principle of non-contradiction and Bayes rule. Eastern oriental cultures are more likely to reasoning holistically and are much more tolerant of contradictions. Mercier (2011) has argued that, while some cross-cultural differences in reasoning exist, claims that formal norms of reasoning — derived from logic or probability theory — are not universal may not stand up to scrutiny. For example, apparent differences between Eastern and Western cultures in their adherence to

7 That is, it is derived from mathematical probability theory. Other informal norms, perhaps like those governing dialogical exchanges may not be universal.

the principle of non-contradiction (Peng & Nesbitt, 1999) may be illusory. Both cultures seem able to apply this principle (Mercier, 2011; Mercier et al., 2015).

By fitting the Bayesian model of argumentation to these data, this study has shown that both Turkish and English cultures follow the same norm in informal argumentation but have different inductive biases. A similar model fitting approach might inform the debate over the principle of non-contradiction. The principle of non-contradiction is the foundation of logic, which is taken to underlie our analytic abilities. Mercier (2011) has cast the difference between Western and Eastern cultures in terms of modern dual process theories (Evans & Stanovich, 2013), which distinguish between reflective, analytic processes and intuitive, heuristic processes. Recently, related dual source models have been fitted to reasoning data to tease out the relative contributions of reflective/logical and intuitive/knowledge-based reasoning in classical reasoning tasks (Singmann, Klauer, & Beller, 2016). These models include a parameter that reflects the relative weighting of these two reasoning processes. If non-contradiction were not a factor in Easterners' reasoning, we would expect the value of this parameter to be zero, indicating no analytic contribution to the reasoning process. Values intermediate between 0 and 1 indicate a mix of both styles of reasoning. We suspect that fitting such models to individual participants is likely to yield a lot of variation in the weighting parameter but that people from Western and Eastern cultures would both show intermediate values. However, Easterners may show a systematic tendency toward the low end indicating a lower involvement of analytic processes. This modelling approach is one that we can pursue in future research.

As we just observed, non-contradiction provides the intuitive foundation of formal logic: *if you want to avoid contradictions, you should reason according to standard logic*. The Dutch book theorems provide the intuitive foundation of probability: *if you want to avoid making bets you are bound to lose, you should reason according to probability theory*. These conditional formulations establish the *instrumental* rationality of logic and probability (Elqayam & Evans, 2011). If you want to achieve these practical goals, this is how you ought to reason. Elqayam and Evans (2011), deny that the unconditional formulation, that is, *you should reason according to probability theory*, which is an evaluative normative claim, has any role in the psychology of reasoning. If there is more than one competing normative theory of a task, then we must decide between them empirically. But if we then treat the normative theory selected on this basis as an evaluative norm, one has committed the naturalistic fallacy (Moore, 1903) of inferring an ought from an is, first stated in David Hume's *A Treatise of Human Nature* (1739/2000). However, we often drop the instrumental conditional formulation if the antecedent is considered universal (Oaksford, 2014).

For example, we believe that ripe apples should fall, and would evaluate an apple that did not fall as inedible. We would not feel compelled to formulate this claim as *if gravity is in force, then ripe apples should fall*, because gravity is in force universally in our experience. This experiment shows that the hypothesis that avoiding bets one is bound to lose commands universal assent has withstood at least one attempted falsification. Belief updating in both Turkish and English cultures conformed to Bayes rule which is a trivial consequence of the Kolmogorov axioms that the Dutch book theorems are used to justify. We would argue therefore that evidence for the universality of a norm of reasoning is evidence for an evaluative norm, which, we argue, play an important role in the psychology of reasoning.

9 Conclusion

Despite clear cross-cultural differences in the effects of different linguistic marking of evidence and the reliability of an informant, in informal argumentation, both Turkish and English participants combine this information using the epistemic norm provided by Bayes theorem. These results imply that Bayes theorem provides a universal norm of reasoning and argumentation. These results support recent proposals that many norms, in particular, those derived from logic and probability theory, are not relative to particular cultures, for example, Eastern or Western (Mercier, 2011; Mercier et al., 2015). The universality of these norms also suggests that they can be treated evaluatively in judging the quality of human reasoning. These conclusions are of course tentative as they stand in need of replication comparing a broader range of cultural differences.

References

Akaike, H. (1974). *A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716–723.* doi:10.1109/TAC.1974.1100705, MR 0423716.

Aikhenvald, A. Y. (2004). *Evidentiality*. New York: Oxford University Press.

Aydin, C., & Ceci, S. J. (2013). The role of culture and language in avoiding misinformation: Pilot findings. *Behavioral Sciences & the Law, 31*, 559–573. doi:10.1002/bsl.2077.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68* (3), 255–278.

Bhatia, J.-S., & Oaksford, M. (2015). Discounting testimony with the argument ad hominem and a Bayesian conjugate prior model. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*, 1548–1559.

Birnbaum, M. H. (2000). *Psychological experiments on the Internet*. San Diego, CA, US: Academic Press.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.

Boyd R. & Richerson, P. J. (2007). Cultural adaptation and maladaptation: Of kayaks and commissars. In S. W. Gangestad, J. A. Simpson, S. W. Gangestad, J. A. Simpson (Eds.), *The evolution of mind: Fundamental questions and controversies* (pp. 327–331). New York, NY, US: Guilford Press.

Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language*. Cambridge: Cambridge University Press.

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes, 41*, 390–404. doi:10.1016/0749-5978(88)90036-2.

Burnham, K. P., & Anderson, D. R. (2004), Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods & Research, 33, 261–304.*

Christiansen, M. H., & Chater, N. (2016). *Creating language*. Cambridge, MA: MIT Pres

Corner, A., & Hahn, U. (2013). Normative theories of argumentation: are some norms better than others? *Synthese*, 190, 3579–3610.

Corner, A., Hahn, U., & Oaksford, M. (2011). The psychological mechanism of the slippery slope argument. *Journal of Memory and Language, 64*,133–152.

Eemeren, F. H. van, Garssen, B., & Meufells, B. (2009). *Fallacies and judgements of reasonableness: Empirical research concerning pragmadialectical discussion rules*. Dordrecht: Springer.

Eemeren, F. H. van & Grootendorst, R. (2004). *A systematic theory of argumentation. The pragma-dialectical approach*. Cambridge: Cambridge University Press.

Eemeren, F. H. van, Grootendorst, R., & Snoeck Henkemans, F. (1996). *Fundamentals of argumentation theory*. Mahwah, N.J: Lawrence Erblaum.

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting 'ought' from 'is': descriptivism versus normativism in the study of human thinking. *Behavioural & Brain Sciences, 34*, 233–248. doi: 10.1017/S0140525X1100001X.

Elqayam, S., & Over, D. E. (Eds.) (2016). *From is to ought: The place of normative models in the study of human thought*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-896-2.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. http://dx.doi.org/10.1177/1745691612460685.

Everett, D. (2012). *Language: The cultural tool*. New York: Pantheon Books.

Gabry, J., & Goodrich, B. (2016). rstanarm: Bayesian applied regression modeling via stan [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=rstanarm (R package version 2.13.1).

Gerrig, R. J., & Banaji, M. R. (1994). Language and thought. In R. J. Sternberg (Ed.), *Thinking and problem solving: Handbook of perception and cognition* (pp. 233–261). San Diego, CA: Academic.

Germine, L., Nakayama, K., Duchaine, B. C. et al. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*, 847–857. doi:10.3758/s13423-012-0296-9.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14*, 357–364. doi:10.1016/j.tics.2010.05.004.

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*, 303–314.

Hahn, U., Harris, A. J. L., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic, 29*, 337–367.

Hahn, U., & Oaksford, M. (2006). A Bayesian approach to informal argument fallacies. *Synthese, 152*, 207–236.

Hahn, U. & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review, 114*, 704–732.

Hahn, U., & Oaksford, M. (2012). Rational argument. In K. Holyoak, & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 277–299). Oxford: Oxford University Press.

Hahn, U., Oaksford, M., & Bayındır, H. (2005). How convinced should we be by negative evidence? In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, (pp. 887–892), Mahwah, N.J.: Lawrence Erlbaum.

Hahn, U., Oaksford, M., & Harris, A. J. L. (2013). Testimony and argument: A Bayesian perspective. In F. Zenker (Ed.). *Bayesian argumentation* (pp. 15–38). Dordrecht: Springer.

Hamblin, C. L., 1970, *Fallacies*, London: Methuen.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA, US: Belknap Press of Harvard University Press.

Hornikx, J. (2011). Epistemic authority of professors and researchers: Differential perceptions by students from two cultural-educational systems. *Social Psychology of Education, 14*, 169–183. doi:10.1007/s11218-010-9139-6.

Hornikx, J., & Hoeken, H. (2007). Cultural differences in the persuasiveness of evidence types and evidence quality. *Communication Monographs, 74*, 443–463. doi:10.1080/03637750701716578.

Hornikx, J., & ter Haar, M. (2013). Evidence quality and persuasiveness: Germans are not sensitive to the quality of statistical evidence. *Journal of Cognition and Culture, 13*, 483–501. doi:10.1163/15685373-12342105.

Hornikx, J., & de Best, J. (2011). Persuasive evidence in India: An investigation of the impact of evidence types and evidence quality. *Argumentation and Advocacy, 47*, 246–257.

Hume, D. (1739/2000). *A treatise on human nature*. Oxford: Oxford University Press. doi: 10.2307/2216614.

Korb, K. (2004). Bayesian informal logic and fallacy. *Informal Logic*, 24, 41–70.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14*, 293–300.

Kurzban, R. (2007). Representational epidemiology: Skepticism and gullibility. In S. W. Gangestad, J. A. Simpson, S. W. Gangestad, J. A. Simpson (Eds.), *The evolution of mind: Fundamental questions and controversies* (pp. 357–362). New York, NY, US: Guilford Press.

Leech, G. (1983). *Principles of pragmatics*. London, New York: Longman Group Ltd.

Lucas, A. J., Lewis, C., Pala, F., Wong, K., & Berridge, D. (2013). Social-cognitive processes in preschoolers' selective trust: Three cultures compared. *Developmental Psychology, 49*, 579–590. doi:10.1037/a0029864.

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

Mercier, H. (2011). On the universality of argumentative reasoning. *Journal of Cognition and Culture, 11*, 85–113. doi:10.1163/156853711X568707.

Mercier, H. (2013). Introduction: Recording and explaining cultural differences in argumentation. *Journal of Cognition and Culture, 13*, 409–417. doi:10.1163/15685373-12342101.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–74. doi:10.1017/S0140525X10000968.

Mercier, H., Zhang, J., Qu, Y., Lu, P., & Van der Henst, J. (2015). Do Easterners and Westerners treat contradiction differently? *Journal of Cognition and Culture, 15*, 45–63. doi:10.1163/15685373-12342140.

Moore, G. E. (1903). *Principia ethica*. Cambridge, UK: Cambridge University Press.

Nash, J.C. (2014). On best practice optimization methods in R. *Journal of Statistical Software, 60*, 1–14.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York, NY, US: Free Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291–310. doi:10.1037/0033-295X.108.2.291.

Oaksford, M. (2014). Normativity, interpretation, and Bayesian models. *Frontiers in Psychology, 5*:332. doi: 10.3389/fpsyg.2014.00332.

Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology*, 58, 75–85.

Oaksford, M., & Hahn, U. (2013). Why are we convinced by the ad hominem argument?: Bayesian source reliability and pragma-dialectical discussion rules. In F. Zenker (Ed.). *Bayesian argumentation* (pp. 39–59). New York, NY, US: Springer Science.

Ögel-Balaban, H., Aksu-Koç, A., & Alp, İ. (2012). Kaynak Belleği ile Dildeki Kaynak Göstergeleri Arasındaki İlişkinin 3–6 Yaş Çocuklarında İncelenmesi (The relationship between source memory and linguistic encoding of source: A study of 3–6 year-olds). *Türk Psikoloji Dergisi, 27*, 26–47.

Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition, 103*, 253–299. doi:10.1016/j.cognition.2006.04.001.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist, 54*, 741–754. doi:10.1037/0003-066X.54.9.741.

Picard, R.R., & Cook, R.D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association, 79*, 575–583.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Robinson, E. J. (2009). Commentary: What we can learn from research on evidentials. In S. A. Fitneva &, T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development*. *New Directions for Child and Adolescent Development, 125*, 95–103.

Royston, P. (1995). Remark AS R94: A remark on Algorithm AS 181: The *W* test for normality. *Applied Statistics, 44*, 547–551.

Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology, 88*, 61–87. doi:10.1016/j.cogpsych.2016.06.005.

Slobin, D. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 157–192). Cambridge: MIT Press.

Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language, 69*, 121–134. doi:10.1016/j.jml.2013.03.004.

Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language, 87*, 157–176. doi:10.1016/j.jml.2015.10.009.

Zenker, F. (Ed.) (2013). *Bayesian argumentation: The practical side of probability*. New York, NY, US: Springer Science. doi:10.1007/978-94-007-5357-0_1.

Appendix

APPENDIX A The reliability of the information sources used in main Experiment.

Information sources			
High reliability		**Mean**	**SD**
1.	Documentaries	7.97	1.51
2.	Academic Scientific Publications	7.7	1.69
3.	University Databases and Resources	7.17	1.99
4.	Wikipedia	6.62	2.43
5.	Regularly-followed authors/writers	6.35	1.73
Moderate reliability			
6.	Sci-Tech Sections of Newspapers	6.3	2.3
7.	News Portals and Agencies	5.32	2.28
8.	Local Newspapers	4.87	2.52
9.	Regularly-followed Internet Forums	4.22	1.92
10.	Facebook-Youtube-Twitter	3.62	2.61
Low reliability			
11.	Poll/Survey Companies	3.45	2.14
12.	GSM Operators	3.32	2.55
13.	Political Parties	2.62	2.07
14.	Politicians and Ministers	2.5	2.34
15.	Internet Advertisements	2	1.92