Deception in evidential reasoning: Willful deceit or honest mistake?

Toby D. Pilditch[1,2], Alexander Fries[3], and David Lagnado[1]

[1]Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, UK
[2]University of Oxford, School of Geography and the Environment, South Parks Road, Oxford, OX1 3QY, UK
[3]UCL Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, UK

Abstract

How does one deal with the possibility of deception? Extant literature has mostly focused on identifying deception via cue detection. However, how we reason about the possibility of deception remains under-explored. We use a novel formalism to expose the complexity of this reasoning problem (e.g. separating the uncertainty of an honest mistake, from willful deception), in the process highlighting several reasoning errors regarding deception. Notably, we show reasoners to make substantial errors when reasoning about a (possibly) deceptive source *in isolation* (including base rate neglect errors), but find that reasoning improves when further (independently sourced) corroborative or contradicting reports are introduced.

Keywords: deception; evidential reasoning; probabilistic reasoning; Bayesian Networks; belief updating

Introduction

The question of how to deal with the possibility of deception has long been of interest to police, military and intelligence investigation, among other domains. A potentially deceptive source, more so than a generally unreliable (e.g. incompetent) source, can be particularly deleterious to an investigation, via the wilful sowing of misinformation. Critically, however, investigators seldom have definitive proof of deception, and are therefore placed into the realm of reasoning under uncertainty. In the present paper, we demonstrate a novel Bayesian formalism for capturing the complex uncertainties surrounding (potentially) deceptive sources, such that optimal inferences regarding the likelihood of deception, as well as the hypothesis being informed upon, may be updated with minimised inaccuracy (Pettigrew, 2016). Moreover, we demonstrate that lay reasoners wildly diverge against such a normative expectation.

Deception in Psychology

Deception has typically been researched in terms of lie detection (see Vrij, 2008). Crucially, previous research has noted that individuals struggle with the uncertainties surrounding the possibility of deception (e.g. chance error vs deception) when explaining errors (Schul, Mayo, Burnstein, & Yahalom, 2007).

Research on perceived trustworthiness has shown that it influences attitudes (Cuddy, Glick, & Beninger, 2011; Fiske, Cuddy, & Glick, 2007), persuasive efficacy (Briñol & Petty, 2009), risk perception (Siegrist, Cvetkovich, & Roth, 2000; Earle, Siegrist, & Gutscher, 2010), and advice uptake (Schul & Peri, 2015). But relatively little research has been conducted in regards to not only how people *do reason* about the *possibility of deception*, but also how they *should*. Within evidential reasoning, one can consider deception to be a special case of (dis)trustworthiness. Dual process models in argumentation, like the Heuristic Systematic Model (HSM; Chaiken & Maheswaran, 1994) and Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1984) have argued that cues to the trustworthiness of a source are only attended to in the absence of effortful engagement with the arguments made by that source.

More recently, coherence-based models, such as the Bayesian source credibility model (Hahn, Harris, & Corner, 2009; Harris, Hahn, Madsen, & Hsu, 2015) have provided a framework that moves beyond the directional predictions of earlier models. This has allowed for the integration of a source's trustworthiness (the willingness to impart accurate information) and orthogonally, expertise (the capacity to impart accurate information) into the support provided by a report from a source. Using these models as a normative backdrop, lay reasoners have been shown to take into account the impact of credibility on argument strength (Hahn et al, 2009), and even follow appropriate adjustments in estimations of argument strength and source reliability in light of (shared) compromising reliability information (Madsen, Hahn, & Pilditch, 2018).

Formalising Deception

Taking forward the notion of deception as a special form of (un)reliability, work using Bayesian networks representations to model legal cases has used an idiomatic approach for witness testimony (Fenton, Neil, & Lagnado, 2013). More precisely, when modelling the strength of a witness's testimony, one may consider two possible (non-exclusive) causes of it – the hypothesis being reported on (e.g. guilt of suspect), and the reliability of the witness.

In the same manner, we may model the representation of deception as a possible cause, along with the hypothesis being reported upon (Lagnado, Fenton & Neil, 2013). Fig. 1 below uses an example case of a target hypothesis – "Is the suspect under questioning in fact the mob's hitman?", and a number of informing sources in a police investigation. Two of these, a forensic scientist and an eyewitness (each retaining their own respective reliabilities), and two Inspectors, McGarret and Graham, who are typically accurate in their investigative reports. Critically, each source reports independently of the other, but there is reason to believe McGarret and Graham *may* in fact be in league with the mob, and thus the possibility of deception is introduced (left-most node in Fig. 1).

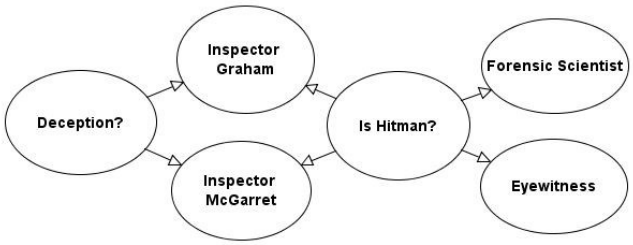

Figure 1. Graphical representation of deception scenario.

To tease apart the levels of uncertainty introduced by the possible deception cause, it is necessary to look at an example conditional probability table (CPT) that represents a (possibly) deceptive agent:

Table 1. Conditional probability table (CPT) representation of a potentially deceptive source, reporting on a hypothesis (Hyp) as either true (T) or false (F).

	Deception = False		Deception = True	
	Hyp = F	Hyp = T	Hyp = F	Hyp = T
Rep = Yes	α	β	γ	δ
Rep = No	(1 -α)	(1 - β)	(1 - γ)	(1 - δ)

In Table 1, α is the probability that the source reports that the suspect is a mob hitman (Rep = Yes) given that he *is honest (Deception = false)* and the suspect is not a mob hitman (Hyp = false). β is the probability that the source reports that the suspect is a mob hitman (Rep = Yes) given that he *is honest (Deception = false)* and the suspect is a mob hitman (Hyp = True). γ is the probability that the source reports that the suspect is a mob hitman (Rep = Yes) given that he *is dishonest (Deception = true)* and the suspect is not a mob hitman (Hyp = false). Finally, δ is the probability that the source reports that the suspect is a mob hitman (Rep = Yes) given that he *is dishonest (Deception = true)* and the suspect is a mob hitman (Hyp=true). δ and (1 - γ) may be due to an imperfect deception, or due to long-run motivations to keep the deception in place. For this initial proof of principle, we simplify the notion of deception by removing this possibility (δ = 0; γ = 1). Put another way, in the scenario we model (and present to participants), deceivers will a) have insider knowledge (i.e. know the true state of "Hyp"), and b) always lie.

Placing Table 1 within the context of the model (and scenario) outlined in Fig. 1, there are a number of important inferences of which to take note.

Firstly, the difference between the probability of a report due to honest error and due to wilful deception (α versus γ), plays a pivotal role in the potential diagnosticity of the report for both P(Deception) and P(Hyp), such that as γ – α increases, the report becomes more diagnostic of deception.

This in turn has a multiplicative effect when considering the second elements: the prior probability of deception (P(Deception)) and – critically – the prior probability of hypothesis being true (P(Hyp)). More precisely, if a report confirms a hypothesis that is *likely* (e.g. P(Hyp) > .5), then P(Deception) should *decrease*, whilst if the report confirms an *unlikely* hypothesis (e.g. P(Hyp) < .5), then P(Deception) should increase. These inferences can best be explained with consideration of how *surprising* a report would be from an honest agent. If unsurprising (e.g. they are saying something *expected*), then an alternative explanation of the report (e.g. deception) is less warranted, and vice versa.

Thirdly, subsequent testimony from independent witnesses will lead to intercausal inferences of P(Deception). For instance, if potentially deceptive agents have their reports *corroborated* by independent testimony, then the increasing probability of P(Hyp) *explains away* the possible deception explanation, lowering P(Deception). Conversely, if independent testimony contradicts the reports of the potentially deceptive sources, then P(Hyp) becomes a less likely explanation of their reports, and again via explaining away, P(Deception) becomes a more likely explanation.

Finally, we seek to provide reasoners with one further clue to deception inferences. The common-cause structure of the deception explanation (left-most node of Fig. 1) – where if deception is true, it explains *both* Inspector Graham and Inspector McGarret's reports, in conjunction with "always liars" (δ = 0; γ = 1) element of their CPTs, allows for an observation-based way of dismissing the possibility of deception. More precisely, given the above, it is not possible for P(Deception) to be true if the two Inspectors contradict each other.

In sum, we use the above formalism to test lay reasoners on 3 different elements of the uncertainty surrounding deception: the prior probabilities of deception (and reported hypothesis) as explanations, the conditional probabilities, and observation-based inference.

The Experiment We present lay reasoners with the above scenario of a police investigation looking into whether a suspect in custody is a mob hitman. The key element to this scenario is to assess how well lay reasoners can integrate the influence of the possibility of deception when integrating testimony from what may otherwise be considered reliable sources.

Of interest is whether reasoners are able to make the following key inferences as more evidence comes in from the available sources:

1. Will reasoners sufficiently account for the likelihood of an honest report when estimating the probability of deception? I.e. If the source is reporting the (a priori) more likely state of the world, then P(Deception) should in fact *decrease*?
2. Will reasoners sufficiently account for the common-cause element of this form of deception? Namely that if the two potentially deceptive sources contradict one another, then they cannot (both) be (all-knowing, perfect liar) deceivers.
3. Will reasoners sufficiently account for the impact of independent sources, when their reports either a) corroborate the deceptive agents (and thus P(Deception) should decrease) or b) contradict the

deceptive agents (and thus P(Deception) should increase)? However, reasoners are likely to get the qualitative direction of these latter inferences.

Method

Participants 180 UK participants were recruited and participated online through the Prolific Academic platform. Participants were native English speakers, with a median age of 28.5 (*SD* = 11.3), and 113 participants identified as female. All participants gave informed consent, and were paid 1.30GBP for their time (*Median* = 8.69 minutes, *SD* = 4.38).

Procedure & Design Participants are provided with a brief background to the scenario, in which they are investigating whether a suspect is in fact a hitman hired by the local mob. They are instructed that they have a number of sources to inform their investigation: two highly reliable inspectors, Graham and McGarret, a Forensic Expert, and an eyewitness – all of whom provide assessments independently of one another. Critically, along with being provided with a prior probability of the suspect being the hitman (P(Hitman) = .1), participants are told there are some logs that suggest the two investigators may be in league with the mob. It is explained to participants that although this is unlikely (P(Deception = .1), if true, the two inspectors will both know the truth (they know the identity of the hitman) and will be motivated to always lie (make sure the innocent suspect takes the fall, or prevent the guilty suspect from going to jail). All the necessary probabilities to populate the underlying model (e.g. error rates of each source) and structures (e.g. common-cause structure of P(Deception)) were provided to participants.[1]

Having had the background explained to them, participants then repeated back the prior probabilities for P(Hitman) and P(Deception):

P(Hitman): *"Until you receive the assessments of other professionals investigating whether the suspect is in fact the hitman, you can safely assume a fairly low (10%) chance of the suspect being the hitman. Please indicate you understood the initial (baseline) probability of the suspect being the hitman:"*

P(Deception): "*... there is only a 10% probability that the two criminal investigators are in fact compromised ...*

Please indicate you understood the initial (baseline) probability of the two criminal investigators being in league with the mob boss:"

Using the gRain package in R (Højsgaard, 2012), these elicited prior probabilities were used to outfit a Bayesian Network (BN) model (Fig. 1) for each participant, creating individually fitted BNs (hereafter termed Behaviorally Informed Bayesian Networks; BIBNs). The remaining structure and parameters were taken from the background information presented to all participants. Thus, a fitted normative comparison could be made for inferences on the participant level.

Following the elicitation of priors, participants then saw three stages of observations, with questions asked at each stage.

(T1) Firstly, participants heard from both the potentially deceptive agents ("DecAgents"). This was manipulated between-subjects, as: Both Report Hitman=True, Both Report Hitman=False, One Contradicts the other.

(T2) Participants then heard from the Forensic Expert, followed by the eyewitness (T3), in separate elicitation stages. These ("OtherAgent") reports were also manipulated between-subjects, as: Both Report Hitman=True, Both Report Hitman=False.

Across these 3 stages, participants were asked two sets of questions:

Probability Estimates (sliders from 0-100%, no default):

- **Hitman Hypothesis:** *"Based on the evidence so far, what do you believe is the current probability of the suspect being the hitman?"*
- **Deception Hypothesis:** "*Based on the evidence so far, what do you believe the current probability is that the criminal investigators are in league with the mob boss?*"

Qualitative Judgments (forced choice; response options: "Increased" / "Decreased" / "Same"; randomized presentation order.):

- **Hitman Hypothesis:** *"Based on the evidence so far, do you believe the probability of the suspect being the hitman has increased, decreased, or remained the same?"*
- **Deception Hypothesis:** "*Based on the evidence so far, do you believe the probability that the criminal investigators are in league with the mob boss has increased, decreased, or remained the same?*"

Thus, this 3 (DecAgents reports) x 2 (OtherAgents reports) x 3 (Elicitation stage) x 2 (Hypothesis) design allows for the testing of the influence of explanation priors, internal (within DecAgents) contradiction, and independent corroboration/contradiction, on estimates (both quantitative and qualitative) of the probability of the hypothesis, and the probability of deception.

Results

Bayesian statistics were employed throughout[2] using the JASP statistical software (JASP Team, 2018). For the sake of brevity, analyses are not reported exhaustively here.

[1] Using the notation of Table 1, participants were given values $\alpha = .05$ (honest false positive); $\beta = .95$ (honest true positive); and $\gamma = 1$, $\delta = 0$ (deception = always lie) for deceptive agents. For full details of the materials used, as well as the collected data, please see https://osf.io/4hvu6/.

[2] Bayes Factors (BF_{10}: likelihood ratio of data given hypothesis, over data given null), may be interpreted as: 1 – 3 = anecdotal support; 3-10 = substantial; 10-30 = strong; 30-100 = very strong; >100 = decisive (Jeffreys, 1961). Conversely, Bayes Factors < .33 can be considered substantial support for the *null* (Dienes, 2014). All analyses used an objective (uninformed) prior. Sample sizes for a given analysis (*N*), and Bayesian Credibility Intervals (95% CI) are indicated wherever appropriate.

Hypothesis 1: Priors and Deception (Base rate neglect)
To understand the impact of priors, we look at estimates and judgments relating to the introduction of the potentially deceptive agents reports (i.e. Baseline to T1), on both P(Hitman) and P(Deception) estimates and judgments, with greater errors predicted for the latter.

P(Hitman) estimates (black lines, Fig. 2). A repeated measures ANOVA was run using elicitation stage (Baseline, T1) and Observed vs Predicted (Data vs BIBN Model) as within-subject factors, and DecAgents condition (restricted to Hitman=True vs Hitman=False reports) as a between-subject factor. This found main effects of elicitation stage (positive trend), $BF_{Inclusion} > 10000$, Observed vs Predicted (data > model), $BF_{Inclusion} = 4.905$, DecAgents condition (Hitman=True > Hitman=False), $BF_{Inclusion} > 10000$, decisive deviations from expectation over time, $BF_{Inclusion} = 5.334$, and opposing trends based on DecAgents condition (increases with Hitman=True, decreases with Hitman=False), $BF_{Inclusion} > 10000$. Crucially, there was no evidence for an interaction of Observed vs Predicted with DecAgents condition, $BF_{Inclusion} = 1.112$, or in conjunction with elicitation stage, $BF_{Inclusion} = 2.178$, indicating no influence of reported base rates on the correctness of P(Hitman) estimates.[3]

P(Deception) estimates (grey lines, Fig. 2). Not only are the same background terms all decisive ($BF_{Inclusion}$'s all > 10000), but there are decisive interactions of Observed vs Predicted and DecAgents condition, $BF_{Inclusion} > 10000$, and the three-way including elicitation stage, $BF_{Inclusion} > 10000$. As can be seen in Fig. 2 by looking at grey solid (participant) vs grey dashed (BIBN model) lines in the middle row (DecAgents reports Hitman=False) vs bottom row (DecAgents reports Hitman=True), estimates *increase* when they should *decrease* in the former, and insufficiently increase in the latter.[4]

Qualitative judgments. Correct responding proportion for the change in P(Hitman) to P(Hitman|DecAgents) did not differ between the DecAgents reports Hitman=True (.39) and DecAgents reports Hitman=False (.25) conditions (N = 121), $BF_{10} = 0.852$. However, in line with probability estimate data, there was substantial evidence for correct responding proportions for the change in P(Deception) to P(Deception|DecAgents) being worse in the DecAgents reports Hitman=False (.1) than DecAgents reports Hitman=True (.28) conditions (N = 121), $BF_{10} = 3.99$.

This latter effect, in conjunction with the P(Deception) estimates, confirms the neglect of the report base rates when considering the possibility of deception, leading to substantial overestimation.

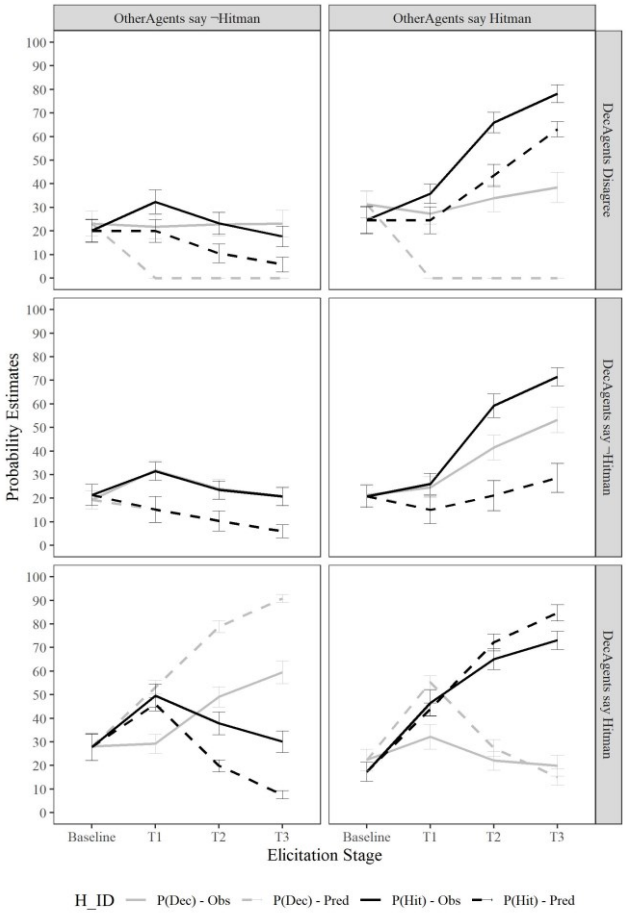

Figure 2. P(Deception) estimates (solid grey lines) and P(Hitman) estimates (solid black lines) across elicitation stages, split by condition. BIBN model predictions are also shown (dashed lines). Error bars reflect standard error.

Hypothesis 2: Common-cause, logic and Deception
To address hypothesis 2, we turn to the DecAgents disagree condition (top row, Fig. 2). Here we focus again on the change in P(Deception) estimates from baseline to T1, as well as the correctness of qualitative judgments. The logic of the structure and conditional probabilities dictate that disagreement between deceptive sources *disproves deception*. However, the repeated measures ANOVA found a decisive deviation from expectation in P(Deception) estimates when moving from baseline to T1, $BF_{Inclusion} > 10000$[5] – an effect corroborated by a t-test showing participants P(Deception) estimates at T1 to be decisively above 0 (N = 59, M = 24.61, SD = 25.44), $BF_{10} > 10000$, δ = 0.937 (95% CI: [0.657, 1.240]).

This error was further confirmed qualitatively, with correct responses (i.e. "Probability decreases") no different from chance level (0.33) responding (N = 59), $BF_{10} = 0.162$, δ = 0.309 (95% CI: [0.203, 0.432]) in a binomial test,

[3] The model with only the above significant terms yielded the best fit, $BF_M = 14.099$, and was significant overall, $BF_{10} = 1.160 * 10^{19}$.

[4] The model with all terms included yielded the best fit, $BF_M = 1.929 * 10^9$, and was significant overall, $BF_{10} = 3.098 * 10^{27}$.

[5] The model including this interaction term yielded the most significant fit, $BF_M = 733042.66$, and was significant overall, $BF_{10} = 3.958 * 10^{15}$.

further confirming an ignorance of the structure and logic based capacity to refute the possibility of deception.
Taken together, these results show that when reasoning about deception, inferences based on structural relations (and logic) alone are highly error prone, once more leading to substantial deception overestimation.

Hypothesis 3: Corroboration, contradiction, and Deception (Explaining Away)

To step through participant estimations of the impact of corroboration / contradiction of possibly deceptive agents, we look first at quantitative estimates (P(Hitman) and P(Deception)) across elicitation stages 1 to 3 – assessing deviation from normative expectation. Second, the correctness of qualitative judgments are assessed over these same stages. This is split by each 2x2 cell (Corroborating Hitman=True, corroborating Hitman=False, contradicting Hitman=True, contradicting Hitman=False).

Corroborating Hitman=True (bottom-right facet, Fig. 2). Repeated measures ANOVAs (elicitation stages T1-T3, and Observed vs Expected) reveal participants do not differ from normative expectation for P(Hitman) estimates, $BF_{Inclusion}$ = 1.777, and track this expectation across elicitation stages, $BF_{Inclusion}$ = 1.436. However, P(Deception) estimates are shown to decisively differ from normative expectation (underestimation), $BF_{Inclusion}$ > 10000, but this deviation decreases across stages, $BF_{Inclusion}$ > 10000.

Table 2 below reveals that whilst qualitative judgments at T1 (when only DecAgents have reported) are correct no better than chance, correct responding at T2 and T3 are greater than chance.

Table 2. Proportion of correct responding in corroborating Hitman=True group. N = 30.

Stage	Hypothesis	Proportion	$\neq .33$ (BF_{10})
T1	Hitman	0.433	0.443
	Deception	0.300	0.218
T2	Hitman	0.900	> 10000
	Deception	0.600	21.16
T3	Hitman	0.733	5313.25
	Deception	0.567	7.527

Corroborating Hitman=False (middle-left facet, Fig. 2). Repeated measures ANOVAs reveal participants decisively differ from normative expectation for P(Hitman) estimates (overestimation), $BF_{Inclusion}$ > 10000, but this deviation does not change across elicitation stages, $BF_{Inclusion}$ = 0.390. Similarly, P(Deception) estimates are shown to decisively differ from normative expectation (overestimation), $BF_{Inclusion}$ > 10000, and this does not change across stages, $BF_{Inclusion}$ = 0.45.

Table 3 below reveals that once again whilst qualitative judgments at T1 (when only DecAgents have reported) are correct no better than chance, correct responding at T2 and T3 are again greater than chance.

Table 3. Proportion of correct responding in corroborating Hitman=False group. N = 32.

Stage	Hypothesis	Proportion	$\neq .33$ (BF_{10})
T1	Hitman	0.25	0.306
	Deception	0.031	707.137†
T2	Hitman	0.781	> 10000
	Deception	0.563	8.124
T3	Hitman	0.656	248.427
	Deception	0.594	22.385

† = Decisively *worse* than chance level.

Contradicting Hitman=True (bottom-left facet, Fig. 2). Repeated measures ANOVAs reveal participants decisively overestimate P(Hitman), $BF_{Inclusion}$ > 10000, and there is strong evidence that this overestimation increases across elicitation stages, $BF_{Inclusion}$ = 17.66. However, participants decisively *underestimate* P(Deception), $BF_{Inclusion}$ > 10000, a trend that does not change across elicitation changes, $BF_{Inclusion}$ = 0.656. Table 4 below reveals that qualitative judgments at T1 (when only DecAgents have reported) are again correct no better than chance, whilst correct responding at T2 and T3 are decisively greater than chance.

Table 4. Proportion of correct responding in contradicting Hitman=True group. N = 31.

Stage	Hypothesis	Proportion	$\neq .33$ (BF_{10})
T1	Hitman	0.355	0.220
	Deception	0.258	0.282
T2	Hitman	0.677	499.34
	Deception	0.774	> 10000
T3	Hitman	0.774	> 10000
	Deception	0.677	499.34

Contradicting Hitman=False (bottom-left facet, Fig. 2). The final repeated measures ANOVAs reveal participants again decisively overestimate P(Hitman), $BF_{Inclusion}$ > 10000, and that this overestimation increases across elicitation stages, $BF_{Inclusion}$ = 334.9. Similarly, participants decisively overestimate P(Deception), $BF_{Inclusion}$ > 10000, but this does not change across elicitation changes, $BF_{Inclusion}$ = 1.965.

Finally, Table 5 below reveals that qualitative judgments at T1 (when only DecAgents have reported) are once again correct no better than chance, whilst correct responding at T2 and T3 are substantially greater than chance.

Table 5. Proportion of correct responding in contradicting Hitman=False group. N = 31.

Stage	Hypothesis	Proportion	$\neq .33$ (BF_{10})
T1	Hitman	0.25	0.307
	Deception	0.179	0.897
T2	Hitman	0.857	> 10000
	Deception	0.714	1164.74
T3	Hitman	0.821	> 10000
	Deception	0.607	20.143

Hypothesis 3 Summary. Taking these 4 sets of analyses together, it is clear that participants can qualitatively appreciate the influence of both corroboration and contradiction from independent sources on potentially deceptive sources, for both P(Hitman), via diagnostic inference, and P(Deception), via an explaining away inference. This is in stark comparison to the substantial qualitative error rates at T1, when only potentially deceptive agents have been observed (see Hypothesis 1). However, estimation data reveals participants consistently overestimate P(Hitman), irrespective of condition (with the exception of corroborating hitman=True). In line with Hypothesis 1, P(Deception) is overestimated when the potentially deceptive agents are reporting the a priori more likely hypothesis (Hitman=False), and underestimated when reporting the less likely hypothesis (Hitman=True). This again suggests a base rate neglect component to assessments of deception.

Conclusions

The issue of how to deal with the possibility of deception when reasoning under uncertainty is as complex as it is potentially deleterious. We present novel findings that lay reasoners are prone to several systematic errors when integrating the possibility of deception, often leading to substantial overestimation.

Using a Bayesian Network formalism, we disentangle the underlying components of deception, including the base rates of deception and the hypothesis the (potentially deceptive) source is reporting on (here, P(Hitman)), structural and logical components, as well as internal (potentially deceptive source reports) and external (corroborative / contradicting reports) observation.

Crucially, we show lay reasoners to be ignorant of the influence of base rates (leading to overestimation of deception, both qualitatively and quantitatively), and structural relations / logic-based negations (again, resulting in deception overestimation). Lay intuitions regarding the impact of corroborative / contradicting testimony on P(Deception) – via explaining away - are (although conservative) shown to qualitatively correspond to normative expectations.

Taken together, this shows erroneous inferences are highest when dealing with potentially deceptive reports alone (where base rates, conditional probabilities, and logical structure are the only active elements to integrate), but accuracy improves when a reference point (other reports / observations) comes into play. This suggests a note of caution for investigative domains in which deception is a possibility (e.g. intelligence analysis), where estimation errors are likely to be substantial until independent evidence (e.g. corroborating testimony) is gathered.

Further work is proposed to incorporate inaccurate / long-run deception motives (i.e. δ), something that we argue may be captured in the present formalism.

Open Practices

All data and materials have been made publicly available via the Open Science Framework at https://osf.io/4hvu6/.

References

Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, *20*(1), 49–96.

Chaiken, S., & Maheswaran, D. (1994). Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgement. *Journal of Personality and Social Psychology*, *66*(3), 460–473.

Cuddy, A. J. C., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, *31*, 73–98.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, 5, 1-17.

Earle, T. C., Siegrist, M., & Gutscher, H. (2010). Trust, risk perception and the TCC model of cooperation. In *Trust in risk management: Uncertainty and scepticism in the public mind* (pp. 1–50).

Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive science*, *37*(1), 61-102.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Hahn, U., Harris, A. J. L., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*(4), 337–367.

Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2015). The Appeal to Expert Opinion: Quantitative Support for a Bayesian Network Approach. *Cognitive Science*, *39*(7), 1–38.

Højsgaard, S. (2012). Graphical independence networks with the gRain package for R. *Journal of Statistical Software*, *46*(10), 1-26.

JASP Team (2018). JASP (Version 0.9)[Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.

Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning. *Argument & Computation*, *4*(1), 46-63.

Madsen, J. K., Hahn, U., & Pilditch, T. D. (2018). Partial source dependence and reliability revision: the impact of shared backgrounds. In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 722-727). Austin, TX: Cognitive Science Society.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Petty, R. E., & Cacioppo, J. T. (1984). Source Factors and the Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research*, *11*, 668–672.

Schul, Y., Mayo, R., Burnstein, E., & Yahalom, N. (2007). How people cope with uncertainty due to chance or deception. *Journal of Experimental Social Psychology*, *43*(1), 91-103.

Schul, Y., & Peri, N. (2015). Influences of Distrust (and Trust) on Decision Making. *Social Cognition*, *33*(5), 414–435.

Siegrist, M., Gutscher, H., & Earle, T. (2005). Perception of risk: the influence of general trust, and general confidence. *Journal of Risk Research*, *8*(2), 145–156.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.