The Knowledge Attribution Game: Against Interest-Relative Invariantism

Felipe Romero
Washington University in St. Louis

Abstract. In this paper I construct a game-theoretic model of Jason Stanley's third-person bank stakes cases to argue against Interest-Relative Invariantism (IRI). Using the model, I show that in Stanley's cases knowledge attributions cannot be a function of practical circumstances of subjects, as IRI would predict. I conclude that IRI fails to satisfy a desirable requirement for a theory of knowledge attribution, namely, being able to explain how attributors form their knowledge claims. In this respect, Contextualism does a better job.

1. The Problem of Knowledge Attributions

Interest-Relative Invariantism (IRI) is the view that the epistemic standards for knowledge attributing sentences are determined by the practical circumstances of the subject. Consider the following two sentences:

(a) "I know that p"

(b) "John knows that p"

According to IRI, when I utter the sentences it is my practical circumstances and John's practical circumstances that determine whether (a) and (b) are true respectively. Jason Stanley discusses several "stakes cases" to test IRI (Stanley, 2005) and defend it against Contextualism: the thesis that the epistemic standards for knowledge attributions are determined by the context of utterance (DeRose, 1992; Lewis, 1996). In the examples, the context of the person who utters both (a) and (b) determines whether they are true or false. The stakes cases trigger different intuitions, and IRI and contextualism offer competing accounts. Here are the cases and a brief description of the data:

> *Low.* Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. It is not important that they do so, as they have no impending bills. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Realizing that it isn't very important that their paychecks are deposited right away, Hannah says, 'I know the bank will be open tomorrow, since I was there just two weeks ago on Saturday morning. So we can deposit our paychecks tomorrow morning.'
>
> *High.* Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since they have an impending bill coming due, and very little in their account, it is very important that they deposit their paychecks by Saturday. Hannah notes that she was at the bank two weeks before on a Saturday

> morning, and it was open. But, as Sarah points out, banks do change their hours. Hannah says, 'I guess you're right. I don't know that the bank will be open tomorrow.' (Stanley, 2005, pp.3-5)

A common intuition is that Hannah knows in *Low* and does not know in *High*. The IRI-theorist accommodates this intuition pointing out that what is at stake for Hannah changes from *Low* to *High*. That is, Hannah's practical circumstances determine the epistemic standards. Structurally similar cases had been used before by DeRose (1992, p. 913) in his defense contextualism[1]. The contextualist explanation is that Hannah knows in *Low* because she is in a context in which the standards for knowledge are low, whereas *High* is a context in which such standards are higher for her, so she does not pass the bar.

These cases are called *first-person cases* because attributor and subject are the same, and most critics agree that both IRI and contextualism are able to explain the intuitions. Hence, in an attempt to break the tie, philosophers have considered *third-person cases*. In these, attributor and subject come apart. That is, a different person from the subject utters a knowledge attribution, and the stakes for attributor and subject may be different. Stanley discusses the following two, on which I will focus from now on:

> *Low Attributor-High Subject Stakes* [LAHS]. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since they have an impending bill coming due, and very little in their account, it is very important that they deposit their paychecks by Saturday. Two weeks earlier, on a Saturday, Hannah went to the bank, where Jill saw her. Sarah points out to Hannah that banks do change their hours. Hannah utters, 'That's a good point. I guess I don't really know that the bank will be open on Saturday' Coincidentally Jill is thinking of going to the bank on Saturday, just for fun, to see if she meets Hannah there. Nothing is at stake for Jill, and she knows nothing of Hannah's situation. Wondering whether Hannah will be there, Jill utters to a friend, 'Well, Hannah was at the bank two weeks ago on a Saturday. So she knows the bank will be open on Saturday'. (Stanley, 2005, p.4)

> *High Attributor-Low Subject Stakes* [HALS]. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since they have an impending bill coming due, and very little in their account, it is very important that they deposit their paychecks by Saturday. Hannah calls up Bill on her cell phone, and asks Bill whether the bank will be open on Saturday. Bill replies by telling Hannah, 'Well, I was there two weeks ago on a Saturday, and it was open.' After reporting the discussion to Sarah, Hannah concludes that,

[1]The first stakes cases were presented by DeRose (1992, p. 913). See (DeRose, 2009, p.1) for a more recent formulation. Stanley's cases are structurally the same to DeRose's and the data they provide is, for my purposes, the same.

since banks do occasionally change their hours, 'Bill doesn't really know that the bank will be open on Saturday'. (Stanley, 2005, p.5)

A common intuition is that in LAHS, Jill, the attributor, is wrong when she says that Hannah knows the bank will be open on Saturday. In HALS, Hannah is right when she says that Bill doesn't know that the bank will be open on Saturday.

The debate has focused mostly on trying to show which theory of knowledge attribution does a better job meeting the following requirement:

(1) A theory of knowledge attribution should explain our common intuitions in the stakes cases.

Here is were the problem begins, as both sides have acknowledged. Both contextualism and IRI account for the data partially. Both can account for the intuitions in the first-person cases, but not in the third-person cases. Here is how contextualism fails. If the context of the attributor determines the standards for knowledge, then Jill should be right when she attributes knowledge to Hannah in LAHS, but we intuit she is not. Contextualism, however, accounts for HALS. Hannah is right in concluding that Bill does not know, because he has the same evidence she has, but in her context the standards for knowledge are higher. IRI, on the other hand, has troubles with HALS. If the subject's stakes determine the standards for knowledge, then Hannah should say that Bill knows but she doesn't.

If we were to decide which theory is better at satisfying (1) based on the number of cases they get right it would seem, at least in HALS and LAHS and structurally similar cases, that IRI and contextualism are even. To mitigate the failures, both sides have attempted to explain away the intuitions that create conflict proposing error theories. I won't discuss such error theories here[2], but I want to point out that appealing to such theories might make us suspicious of the whole debate[3], because it implies that, while trying to describe how knowledge attributions work, we have to regard speakers as being systematically wrong in some of their knowledge attributions[4].

I will use a game-theoretic model that shows more clearly why contextualism and IRI are even with regard to (1). However, my main purpose is to

[2]To give an idea: Hawthorne, for instance, suggests a psychological explanation: anxious attributors *project* their own worries to subjects. See (Hawthorne, 2004, p.160–166), (Stanley, 2005, pp.98-104), (Schaffer, 2006, pp.92–94) and (DeRose, 2009, pp.230–238) for discussions about error theories.

[3]There are other reasons that could make one suspicious about setting the debate as a problem of stakes. Schaffer (2006), for example, argues that Stanley's cases are presented in a biased form, and he offers alternative versions of the cases to argue that it is *salience* of possibilities of error rather than stakes that triggers our intuitions. Schiffer (2007, pp.190–191) shows how salience of possibilities of error can be manipulated merely by using *pessimistic* agents in the cases. And Russell & Doris (2008, pp.431–433) present cases in which *indifference* about practical matters turns out to be (counterintuitively) knowledge-making according to IRI.

[4]In this line, Williamson (2005, Section II) argues that both IRI-theorists and contextualists endorse a methodological principle of charity which, in the end, they cannot fully respect, i.e. "we should prefer to interpret speakers as speaking and thinking truly rather than falsely (ceteris paribus)".

evaluate the discussion on a different basis. Requirement (1) is neither completely satisfied by either theory, nor enough to favor one over the other. If such a diagnosis is right, we need to evaluate both theories and reveal their strengths and weaknesses using a criteria different from (1). There is a second requirement that allows us to do that, which has not been explicitly discussed in the literature:

(2) A theory of knowledge attribution should explain how attributors form their knowledge claims in the stakes cases.

The difference between (1) and (2) might not be obvious. To meet (1) a theory has to provide an explanation of why we, the readers of the cases, have certain intuitions about who knows, who does not, who is right, wrong, and so on. On the other hand, to meet (2) a theory has to provide an explanation of how attributors (Jill, Hannah, etc.) come to believe the knowledge claims they believe. Succeeding in (2) implies adopting the perspective of the attributors, and explaining their attributions in terms of the information they have available (e.g., whether they know they have impending bills or that the bank was open last Saturday). I argue that IRI fails in doing this for HALS and LAHS. Here is the argument:

(P1) Nothing but the information that attributors have available can explain how they form their knowledge claims in HALS and LAHS.

(P2) IRI cannot explain how attributors form their knowledge claims in HALS and LAHS on the basis of the information they have available.

(C) IRI cannot explain how attributors form their knowledge claims in HALS and LAHS.

To make my argument precise and to play with Stanley's rules I limit my analysis to HALS and LAHS (this and other assumptions will be more explicit at the beginning of next section) but, I will suggest later, the conclusion holds for structurally similar cases. Both premises are supported by game-theoretic considerations that I discuss in the next sections. In section 2. I construct the model and explain better what's behind (P1). Section 3. defends (P2), and shows why this is not a problem for contextualism.

2. Modeling Third-Person Knowledge Attributions

In this section I construct a model that captures the relevant features of LAHS and HALS. That is, the information provided by the descriptions of the cases (i.e., what is at stake for attributors and subjects), and the results of the attributions (i.e., whether we intuit that attributors' claims are true or false). Three main assumptions guide the construction of the model. First, I assume that *speakers's intuitions in LAHS and HALS are as Stanley describes* (Stanley, 2005, p.1-14). This means, such intuitions, as I presented them in the first section, are the right data that a theory of knowledge attribution should explain. There are

reasons for doubt[5]. However, I will show that even if the data is right, there are still some explanations that IRI cannot give.

The second assumption is that *only stakes determine knowledge attributions, and other determiners remain constant.* That is, knowledge attributions change as a function of stakes (either given by the context of the attributor or the circumstances of the subject), and other determiners, such as truth, beliefs, more evidence, etc., do not change from one case to the other. The cases involve many variables that could make us suspicious of whether only the stakes change from one case to the other[6]. However, once we use this assumption, we can focus on the issue of whether "someone knows that p may be determined in part by practical facts about the subject's environment"(Stanley, 2005, p.85) without worrying about other variables, which is what I presume the cases are intended to test. In other words, if IRI is right, stakes are responsible for the changes in our intuitions, so changing them while keeping everything else constant should create the expected effects.

Finally, I assume that *if both the stakes for attributor and subject are low, then "the subject knows" is true. If both the stakes for attributor and subject are high, then "the subject does not know" is true*. We need this assumption to construct a complete model. We don't have cases for such situations, however, here are two reasons to support this assumption. First, consider that in the first-person cases the same person is both subject and attributor, and she doesn't know in *High* and does know in *Low*. Second, in a High Attributor-High Subjects stakes case, independently of our theory of knowledge attribution, given that all standards are raised, the attributor will assert that the subject doesn't know. A similar reasoning applies to a Low Attributor-Low Subject case. If all standards are sufficiently low, the subject will be said to know.

Having made these assumptions, I model LAHS and HALS as a knowledge attribution game[7]. Figure 1. shows the game in its extensive form.

An attributor of knowledge is in a situation in which she has to decide whether the subject knows or does not know. She is playing against Nature[8]. Using the player Nature is useful to model the idea that external factors (e.g., pending bills) determine the practical circumstances of the subject and the attributor. Nature intervenes two times. It determines stakes for the subject (H or L), and stakes for the attributor (h or l). Nature's movements don't require any particular order because subject's and attributor's stakes are independent, so they occur simultaneously at the beginning of the game. Subsequently, the

[5]For example, Schaffer & Knobe (2010) show some surveys that suggest that ordinary speakers don't say what philosophers say they will say about the stakes cases.

[6]For example, minor details in the selection of the words could be considered problematic: one could think that the use of "really" in "Bill doesn't really know" in HALS is doing the change in the intuitions and stakes are secondary.

[7]The problem is essentially decision-theoretic rather than game-theoretic because there is only one player. However, I use the game-theoretic form because it offers a standard syntax for modeling different informational constraints (not available in plain decision theory), which will play an important role in the argument.

[8]The concept of Nature is used as a resource to model probabilistic events that are not the result of choices of players.

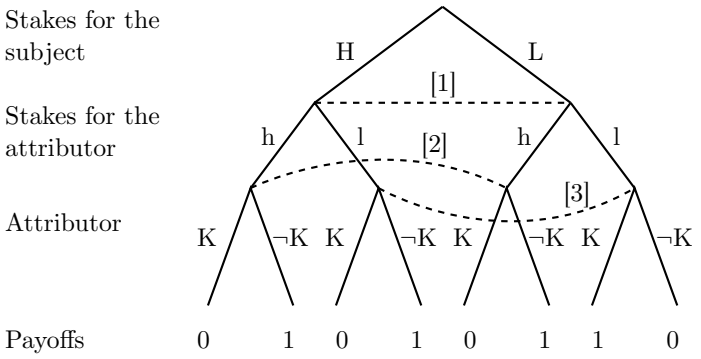

Figure 1: Knowledge Attribution Game

attributor decides whether the subject knows (K) or does not know ($\neg K$)[9]. The payoffs are binary, and represent whether the attribution is true (1) or false (0). These payoffs represent the data Stanley thinks a theory of knowledge attribution should account for, and the data coming from my third assumption. For example, in HALS, Jill attributes knowledge to Hannah, but Stanley considers that she is wrong. We can see this in the model: in the branch H, l, K the payoff is 0.

I also take from LAHS and HALS that whether knowledge attributions are true or false is independent of what the subject does: In LAHS, whether Jill is right when she claims that Hannah knows the bank will be open on Saturday might depend on Hannah's stakes, but doesn't depend on whether Hannah is actually going to the bank on Saturday or not. On the other hand, in HALS, Bill doesn't say that he has any plans of going to the bank or not. The implication for the model is that we don't need to include actions or payoffs for the subject. Technically, the subject is not a player in the game, but we refer to her indirectly because her stakes are determined by Nature.

Now we can model information. LAHS and HALS include two features in this regard: (1) When the attributor utters her attribution, she knows that she and the subject have exactly the same evidence: the bank was open two weeks ago: in LAHS, Jill remembers seeing Hannah in the bank two weeks before; in HALS, Bill tells Hannah that he went to the bank two weeks before. (2) The attributor knows her stakes but ignores the stakes for the subject: in LAHS Jill "knows nothing about Hannah's situation"(Stanley, 2005, p.4); in HALS, Hannah's conversation with Bill only confirms the evidence she already has. In game-theoretic terms, the game is played under incomplete information. To

[9]Arguably, there are some cases in which speakers should remain agnostic about whether someone knows or does not know. That is, there is space for a third option: *don't believe either*. Such an option does not appear in the data we have to explain, thereby I don't include it in the model. I discuss more about this issue in section 4.

model this situation we need to use *information sets*, which are represented by dotted lines. All the nodes touched by the same dotted line are part of the same information set. When the player reaches an information set, she does not know in which node of information set she is. Hence, she has to choose the best action assuming she could be in any node within the set.

Consider the information sets in Figure 1. In information set [2], the attributor knows that her stakes are high, but doesn't know whether the subject's stakes are high or low. On the other hand, in information set [3], the attributor knows that her stakes are low, but doesn't know whether the subject's are high or low. Information set [1] models the idea that Nature defines the stakes for the attributor independently of what the stakes for the subject are and are defined simultaneously before the attributor plays.

I have specified the knowledge attribution game. The model contains a precise representation of the features of HALS and LAHS. The model also captures the assumption that any other truth determiners different from the stakes remain constant from one case to the other. We can consider again (P1) in my argument. Whether an attributor asserts or denies a knowledge claim depends on her reasoning from the information sets she is in. In other words, nothing outside the information sets could be legitimately used in an explanation of how attributors form their knowledge claims.

3. Information Constraints and the Problem for IRI

I will defend premise (P2) showing that IRI is not able to explain how attributors form their knowledge claims in the model. To do so, I first discuss how the required explanation should look. That is, how such explanations should be formalized in game-theoretic terms. Then, I will show that such a formalization, in the case of IRI, cannot be done.

Given a game, a player can play different *strategies* which might lead to very different outcomes. A strategy is a sequence of actions that the player decides to do, each of which depends on the stage of the game and the information she has available at that stage. The knowledge attribution game has only one stage. That is, the attributor has to make only one decision (i.e., utter "the subject knows" or "the subject does not know") and the information she has available depends on which information set she is in. Now remember that IRI and contextualism attempt to predict the result of the decision, and the difference in their predictions depends ceteris paribus on the theory's purported epistemic standards. To satisfy requirement (2), contextualism (and IRI) should explain how the contextualist (or the IRI-ist) agent makes the decision. In game-theoretic terms, they should provide the strategy that the contextualist (or the IRI-ist) agent plays in the game. To model such strategies we have to formalize the decision of attributing (K) or denying knowledge ($\neg K$) as a function of the theory's epistemic standards. In short, having a function that formalizes the strategy of an attributor in the knowledge attribution game is precisely having an explanation of the way she forms her knowledge claims, which satisfies requirement (2).

Here is how it works for contextualism. According to contextualism, the epistemic standards are provided by the context of utterance. In the knowledge attribution game this context is represented by the attributor's stakes. Hence, the contextualist strategy is to be specified as a function of the attributor's stakes (h or l). Informally, the strategy consists in denying knowledge when the attributor's stakes are high, and attributing it when the stakes are low. Formally, let s_{CTX} be the strategy, which is:

$$\begin{aligned} s_{\text{CTX}}(h|H) &= \neg K \\ s_{\text{CTX}}(h|L) &= \neg K \\ s_{\text{CTX}}(l|H) &= K \\ s_{\text{CTX}}(l|L) &= K \end{aligned}$$

Given that the stakes for the subject don't do any work for the contextualist strategy for this game, we can simply write:

$$\begin{aligned} s_{\text{CTX}}(h) &= \neg K \\ s_{\text{CTX}}(l) &= K \end{aligned}$$

Figure 2. represents s_{CTX}. Bold lines represent what the attributor does in any given node.

Remember that contextualism does not account completely for the data (see the discussion about the first requirement in the first section). In line with this, we can see that s_{CTX} does not always obtain the highest payoffs. Specifically, s_{CTX} fails in l after a history of H, because it attributes knowledge but it shouldn't. Up to this point, I'm stating Stanley's and DeRose's analysis formally, and there is nothing new philosophically speaking.

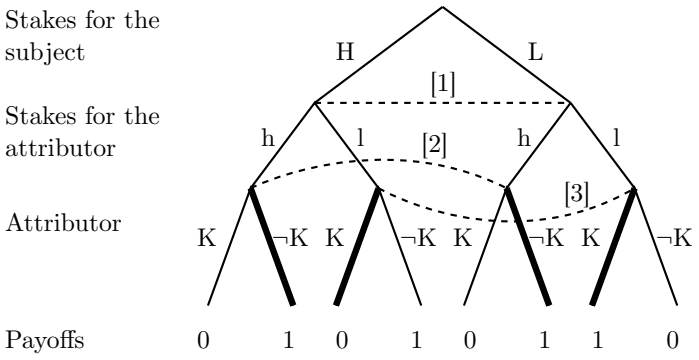

Figure 2: Contextualist Strategy

We get a more interesting result when we try to specify IRI's strategy. According to IRI, the epistemic standards are provided by the practical circum-

stances of the subject. Thus, the strategy s_{IRI} for the game would have to be:

$$\begin{aligned} s_{\text{IRI}}(h|H) &= \neg K \\ s_{\text{IRI}}(h|L) &= K \\ s_{\text{IRI}}(l|H) &= \neg K \\ s_{\text{IRI}}(l|L) &= K \end{aligned}$$

The strategy s_{IRI} would have to look like Figure 3. The problem is that if we consider the constraints that information sets impose, s_{IRI} is not a possible strategy. That is, s_{IRI} is not a strategy that attributors can play. In other words, it is not possible to specify s_{IRI} because it is not consistent with the idea that, once the player reaches an information set, she does not know which node of the information set she is in and, therefore, is unable to distinguish the outcomes of her available actions. Formally, the constraint is that given the information sets [2] and [3], for any possible strategy s, it should be true that $s(h|H) = s(h|L)$ and $s(l|H) = s(l|L)$. This is true for s_{CTX} but false for s_{IRI}.

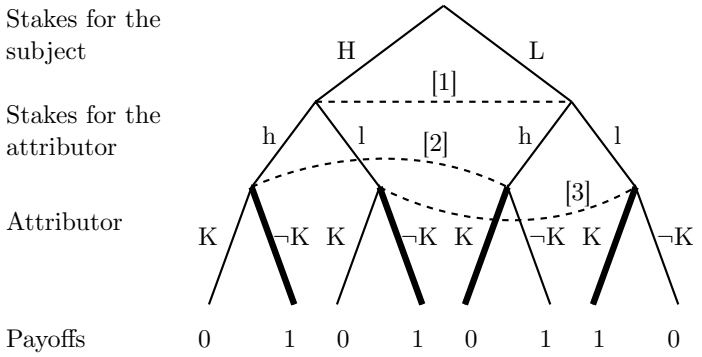

Figure 3: IRI Strategy (impossible)

Why is it a problem that s_{IRI} is not a possible strategy to play the knowledge attribution game? The problem is not that IRI has limitations to account completely for the data. Whether a strategy is possible is different from whether it can account for the data properly (i.e., getting the highest payoffs in the game). As we saw, contextualism can't account completely for the data either but even so we can model a possible strategy s_{CTX} for the game. And if s_{IRI} were possible[10] for this game, it would account for the data as well as s_{CTX}. The problem is that, in HALS and LAHS, what IRI says that attributors are doing cannot be a function of the practical circumstances of the subject, because the attributor does not have access to the information about such circumstances. This means, IRI cannot provide the explanation of how attributors form their knowledge

[10]See section 5. for a discussion about a game in which an IRI strategy is possible.

claims in the very same cases that Stanley uses to show IRI's intuitive appeal. That is, (P2) is true, and so is (C), which also implies that IRI does not satisfy requirement (2).

IRI's intuitive appeal comes from the fact that we (i.e., the readers of the cases) are not informationally constrained, and have intuitions about what speakers will say in the cases, and such intuitions match with IRI's predictions (at least to the point in which it is not possible to say whether IRI or contextualism is better). However, from speakers' perspective, it turns out that such predictions cannot be effectively operationalized. Even in the cases in which the attributor plays in the way IRI predicts (e.g., choosing $\neg K$ in l after a history of H), IRI cannot be the explanation of why she is doing so. More generally, attributors in HALS and LAHS (are said to) believe knowledge claims: they believe that subjects know or do not know. But the stakes for the subject cannot be part of the explanation of why they are doing so, so it is a fact that still needs to be explained.

Where can we find the missing explanation? An evident option is to favor contextualist semantics, given that in the game the context of utterance can effectively determine speakers' knowledge claims. Thus, here we have not only an argument against IRI, but also an argument in favor of contextualism. Other option is that there is something wrong with the presuppositions of the whole debate, and knowledge attributions should not be explained in terms of what is at stake for attributors or subjects. Given my assumptions I endorse the first alternative and I will continue exploring its consequences. Now I will review possible objections to (C).

4. How to Play the Game

The IRI-theorist could try to deny that a IRI (and maybe any theory of knowledge attribution) should satisfy requirement (2). She could say that IRI sets standards to establish whether knowledge attributions are true or false, and that job is unrelated to the factors that determine how attributors come to believe the attributions they believe. Specifically, she could say "if attributors do not know the stakes for the subjects, then *they should not* utter knowledge claims. And, if they do it, then they are systematically wrong because they are in no position to do it". This objection presupposes that IRI has a normative character, because it appeals to an alleged constraint that speakers should observe. In this section I will explore such a normative character and show that it is unwarranted.

Whether theories of knowledge attribution are descriptive or normative is one of the least clear issues in the literature. One could think that the motivation for proposing a theory of knowledge attribution is to explain why common-sense intuitions about knowledge claims change from one test case to the other. In this sense, the project is primarily explanatory or descriptive. Indeed, probably Stanley would agree with this. He says "my philosophical tendency is to preserve as much as possible of common-sense intuition" (Stanley, 2005, p.v.). Nonetheless, given that both theories fail in that explanatory

project, a way out is to use an ad hoc move: *the theory is not only descriptive but also normative. It is not a problem that the theory fails to explain common sense intuitions. The theory is right, and speakers are mistaken.* This is the reasoning that supports objections as the aforementioned.

I think this move is questionable. I will show that if the project is to find a normative theory of knowledge attribution, then IRI is not a good result. A sensible approach to ground the normativity of a theory of knowledge attribution (and the only one I can think of) is demanding speakers to be rational. We can analyze this demand formally, and see that IRI does not respect it.

We can find *solutions* to the knowledge attribution game. A solution is a strategy that is at least as good (in terms of payoffs) as every other strategy available to the agent, according to her preferences. In this sense, solutions are rational. In the knowledge attribution game, a solution would maximally fit the common intuitions in the stakes cases. Thus, whether the attributor should utter K, $\neg K$ or nothing in a given circumstance reduces to whether such a strategy is a solution. Now I will compare IRI with the solutions to the knowledge attribution game.

To find a solutions we first need to find the expected utilities of any possible action by the attributor (See Appendix 1 for details). Let p be the probability that the subject is in high stakes. When the attributor is in high stakes, we have:

$$\begin{aligned} U(K|h) &= 0 \\ U(\neg K|h) &= 1 \end{aligned}$$

On the other hand, when the attributor is in low stakes we have:

$$\begin{aligned} U(K|l) &= 1-p \\ U(\neg K|l) &= p \end{aligned}$$

It is important to stress that these expected utilities are not for the actions of going to the bank on Friday or Saturday. These expected utilities measure *the number of knowledge attributions that the attributor gets right*. Players increase their utility by succeeding in their attributions.

We can use these utilities to know when it is rational for the attributor to attribute knowledge, given her stakes. First, since $U(\neg K|h) > U(K|h)$ for any values of p, we know that $\neg K$ (saying "the subject does not know") is the rational choice when the stakes for the attributor are high. Second, when the stakes for the attributor are low, we have that $U(K|l) > U(\neg K|l)$ iff $1-p > p$. So she should attribute knowledge when $p < \frac{1}{2}$. Given the attributor's uncertainty about the value of p, to specify a solution we need to specify beliefs[11]. The first

[11]Strategies often require specifying *beliefs* (denoted by β) when there is incomplete information. Beliefs are probabilities that the player assumes. The strategy s^* maximizes expected utility under the assumption that β^* is true.

solution is the following[12]:

$$\begin{aligned} s_1^*(h) &= \neg K \\ s_1^*(l) &= K \\ \beta_1^*(l|H) &< \frac{1}{2} \end{aligned}$$

Informally, this strategy states that whenever the attributor is in h, she should say that the subject does not know, and when she is in l, she should say that the subject knows if her belief that the subject is in H is less than $\frac{1}{2}$. One might wonder what happens if the attributor believes that H and L are equally likely. In such a case, the attributor's strategy has to satisfy $U(K|l) = U(\neg K|l)$. From a game-theoretical stance, the strategy should be flipping a coin to decide between K and $\neg K$ if her stakes are low[13], and it would be supported by a belief $\beta_2^*(l|H) = \frac{1}{2}$. We might not want knowledge attributions to be probabilistic. Even if we accept that the epistemic standards change (with context, stakes, salience of possibilities of error, etc) we might want to hold on to the idea that, once we know what the relevant standards are in a given situation, we are able to tell whether the subject knows, does not know or that we are not in a position to ascribe knowledge.

I think it is philosophically interesting that we have to specify beliefs to support any optimal strategy. A rational attributor has to make assumptions in order to play. For example, if a rational attributor in a low stakes situation says "X knows" in ignorance of X's stakes, it is because she believes (either implicitly or explicitly) that the probability that X is in a high stakes situation is lower than a half. Hence, if we want to say that attributors in HALS and LAHS are rational, we have to explain the source of their beliefs, which is not in any of the features captured by the model. The lesson is that stakes (keeping all other possible knowledge determiners constant) are not enough to fully determine the intuitions in HALS and LAHS.

Now we can compare the normative solution s^* with IRI. We have two cases. First, when the attributor is in h, she can always succeed in her attributions regardless of the stakes for the subject, and regardless of whether she knows the stakes for the subject. Thus, demanding not to utter knowledge claims in this case, as the objection suggest, is not justified. When the attributor is in l, the fact that the strategy depends on the attributor's beliefs about the stakes for the subject might be seen as an argument for IRI. But the asymmetry with the first case suggests that it is not the stakes for the subject per se that matter, for if they did, then they would make a difference in the first case. What matters is to have beliefs about stakes in general. If the attributor didn't know her own stakes, she would have to support any strategy with beliefs about her

[12] A note on notation: s denotes a strategy whereas s^* denotes a *dominant strategy*, a strategy that gives the player a payoff larger than any other strategy.

[13] This is called a *mixed strategy*. The actions are decided probabilistically. The attributor plays K and $\neg K$ with probability $\frac{1}{2}$.

own stakes as well. In short, there is no clear relation between a normative solutions to the game and IRI.

The IRI theorist could note that s_{CTX} is also different from s^*. Nonetheless, it is worth noting that s_{CTX} yields the same outcome as s^* when β_1^* is true. In terms of the optimal strategy, this means that a contextualist agent that ignores the stakes for the subject always believes that such stakes are low. This belief fails sometimes, but is both consistent with contextualism, and explains why s_{CTX} deviates from s^*.

Summing up, it is not entirely clear whether IRI is a normative theory or not. But some objections against the argument presented in section 3. presuppose it is. My answer is that even if the project is to build a theory of what attributors should and should not say, IRI fails at this task. In the next section I will explore whether my conclusions hold in different scenarios.

5. Cases Without Informational Constraints

Someone might want to object that the model oversimplifies the cases, because I'm not considering other variables which, in addition to the stakes, drive knowledge attributions[14]. However, as it stands, the model allows to understand decision making under the informational constraints embedded in the cases (which is a feature of the cases that has been overlooked in the literature). In particular, no matter how complicated the model turns out to be, as long as the information about the subject's stakes remains unaccessible to the attributor, it won't be possible to specify a strategy s_{IRI}, which is what (P2) requires to work. The question is what happens when we change such constraints. The IRI theorist could say "You are not being charitable enough with your reading of the stakes cases. Informational constraints should not be taken too seriously. With a slight modification in the design of the cases there would be no informational constraints and (P2) in the argument would no longer be true".

Let's assume that Stanley's remark that in LAHS Jill "knows nothing about Hannah's situation"(Stanley, 2005, p.4) and the implied fact that in HALS Hannah ignores Bill's stakes, shouldn't be taken too strictly. So what happens in cases in which attributors have complete information about the subject's stakes? As I will show, requirement (2) could be satisfied. However, this is not good enough news for IRI, since contextualism does a far better job.

Here are two possible cases to illustrate the situation. The case HALS* is

[14]The game-theoretic representation allows us to describe clearly the logical space of possibilities in the stake cases, so it could be used to represent the cases more accurately, and make the game tree as precise as we want. We could model specific circumstances that determine the stakes for subjects and attributors. For example, the bank could be a player, who decides to open or not on Saturday; the fact that agents point out possibilities of error (e.g., when Sarah points out that banks change their hours) could be modeled as actions; and Nature could be used to determine whether subjects have to pay a bill.

proposed by Schaffer[15] and LAHS* is a modification of LAHS[16]:

> [HALS*] On Friday afternoon, Sam is driving past the bank with his paycheck in his pocket. The lines are long. Sam would prefer to deposit his check before Monday, but he has no pressing need to deposit the check. He has little at stake. Sam remembers that the bank was open last Saturday, so he figures that the bank will be open this Saturday. He is right – the bank will be open. You, by the way, have your entire financial future at stake here. If Sam doesn't deposit his check before Monday, Sam's check to you will not clear in time to save you from impending bankruptcy. Sam has not bothered to look into whether the bank might have changed its hours. (Schaffer, 2006, p. 91-92)

> [LAHS*]. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since they have an impending bill coming due, and very little in their account, it is very important that they deposit their paychecks by Saturday. Two weeks earlier, on a Saturday, Hannah went to the bank, where Jill saw her. Sarah points out to Hannah that banks do change their hours. Hannah utters, 'That's a good point. I guess I don't really know that the bank will be open on Saturday'. Coincidentally Jill is thinking of going to the bank on Saturday, just for fun, to see if she meets Hannah there. Nothing is at stake for Jill, *but she decides to call Hannah to be sure that she will find her there. Hannah tells Jill about her impending bill and her worries about the bank not being open tomorrow*. At the end of the conversation Jill tells to a friend 'Well, Hannah might go today to the bank, because she doesn't know whether it will be open on Saturday '.

HALS* has complete information for the attributor because we, as readers, have complete information, and we, at the same time, are the attributors. The first thing to note about HALS* is that, in spite of the change in informational constraints, our intuitions are not different from our intuitions about the original HALS: we are right in saying "Sam does not know". Remember also that IRI predicts the opposite. Hence, unsurprisingly, the situation does not improve for IRI. Demanding attributors to know about the subject's stakes before uttering a knowledge claim is not providing any help. On the other hand, in LAHS*, after knowing about Hannah's situation, Jill does not attribute knowledge to Hannah. I take it that most readers would intuit Jill's claim is right. Hence, in terms of the intuitions there is no difference for us, as readers, in comparison to LAHS. For Jill, however, knowing about Hannah's stakes makes her say that Hannah does not know (remember that she says that Hannah does know in LAHS). Both LAHS and LAHS* can be explained

[15]Schaffer's aim with HALS* is different from mine here. Indeed, he does not mention the informational difference between his HALS* and Stanley's HALS, nor such difference seems to do any work in his arguments.

[16]The graphical representation of the game for these cases is the same as in Figure 1 except for the information sets [2] and [3] (dotted lines) which would not exist for these.

appealing to IRI. Hence, again, the situation doesn't change here for IRI. Summing up, IRI can satisfy requirement (2) in cases in which the attributors know the subject's stakes. In these cases, however, everything remains as in the original cases with regard to requirement (1). The explanations that IRI cannot give about HALS are still to be given in HALS*, and IRI's success in LAHS is the same as in LAHS*.

If IRI can explain LAHS* and HALS* just as well as LAHS and HALS, avoiding the problem that I am revealing regarding requirement (2), the IRI-theorist could be tempted to say "from now on, I will use LAHS* and HALS* to defend IRI, and bite the bullet (or use some sort of patch) in cases in which IRI cannot satisfy requirement (2)". This, I will show, does not work. To understand why, we have to analyze the job that the informational constraints are doing in LAHS and HALS.

LAHS and HALS have a clear demarcation of (and separation between) the circumstances of evaluation of a knowledge attribution and the context of utterance. First, because subjects and attributors come apart (as in any third-person case), but more importantly, because of the informational constraints. When we remove the informational constraints, the circumstances of evaluation of sentences such as "Hannah does not know" remain the same (e.g., what is at stake for Hannah does not change). However, the context of utterance (which determines the epistemic standards according to contextualism) may change. This is because the information about the circumstances of the subject may make some possibilities of error salient for the attributor, and she has to consider such possibilities when uttering her attribution. This actually happens for the attributor in HALS* and LAHS*. Lewis's contextualism helps to clarify this point:

> *S knows* that *P* iff *S*'s evidence eliminates every possibility in which not-*P* - Psst! - except for those possibilities that we are properly ignoring. (Lewis, 1996, p.554)

When the attributor in HALS* and LAHS* comes to believe some possibility of error (by knowing about the circumstances of the subject) she cannot properly ignore such possibility anymore. Hence, knowing about them raises the (contextualist) bar for knowledge[17]. In other words, in HALS* and LAHS* the circumstances of the subject (which determine the epistemic standards according to IRI) also introduce possibilities of error for the attributor (which raise the epistemic standards according to contextualism). This is not directly a problem for IRI. But the fact that HALS* and LAHS* are messier cases than HALS and LAHS make the former less useful than the latter to resolve the dispute between contextualism and IRI. I believe that's the reason why Stanley chose to embed the informational constraints in HALS and LAHS. Even so, contextual-

[17]This is an application of Lewis's *Rule of Belief*: "A possibility that the subject believes to obtain is not properly ignored, whether or not he is right to so believe. Neither is one that he ought to believe to obtain - one that evidence and arguments justify him in believing - whether or not he does so believe."(Lewis, 1996, pp.555)

ism does a far better job than IRI in LAHS* and HALS* regarding requirement (1), as I show now.

In LAHS and HALS the attributor's stakes provide the possibilities of error that the attributor ought to consider. In LAHS* and HALS*, from the perspective of the attributor, knowing the subject stakes makes salient additional possibilities of error. Specifically, Hannah's high-stakes in LAHS* make salient for Jill the possibility that the bank has changed hours. This explains one fact: Jill in LAHS* says "Hannah does not know" and gets her attribution right. The epistemic standards she is appealing to have changed, because she has information that makes one possibility of error salient, a possibility that she did not consider before talking to Hannah (and that she does not consider at all in LAHS). In terms of the game, the contextualist strategy should capture the idea that knowledge about high stakes introduces possibilities of error that raise the epistemic standards. That is, if someone (anyone) is in a high stakes situation, the attributor denies that the subject knows. If no one is in a high stakes situation, the attributor says that the subject knows. Thus, we can now specify the contextualist strategy for LAHS* and HALS*:

$$\begin{aligned} s_{\text{CTX}^*}(h|H) &= \neg K \\ s_{\text{CTX}^*}(h|L) &= \neg K \\ s_{\text{CTX}^*}(l|H) &= \neg K \\ s_{\text{CTX}^*}(l|L) &= K \end{aligned}$$

Notice that s_{CTX^*} accounts completely for the data. Remember also that the stakes for the subject don't change from HALS and LAHS to HALS* and LAHS*. This implies that an IRI strategy for the latter two (say s_{IRI^*}) is the same as s_{IRI}. This is problematic for the IRI-theorist: s_{CTX^*} satisfies requirement (1) completely and s_{IRI^*} doesn't (it still fails in h after a history of L). Summing up, in HALS* and LAHS* IRI can satisfy requirement (2), but contextualism also satisfies requirement (2) and does a better job than IRI in satisfying requirement (1).

6. Conclusion

Contextualism and IRI offer standards to determine whether knowledge attributions are true or false. In addition to such standards, a theory of knowledge attribution should be able to explain how people get the intuitions they have (at least approximately) in the stakes cases. Such an explanation requires showing how the intended epistemic standards (i.e., the practical circumstances of the subject, or the context of the attributor) interact with the attributor's utterances. IRI fails to show how this interaction is supposed to work in HALS and LAHS, because in such cases attributors don't have epistemic access to the practical circumstances of the subject, but still they are able to utter knowledge attributions. Contextualism does a better job in offering the required explanation.

References

DeRose, K. (1992). Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research*, *52*(4), pp. 913-929.

DeRose, K. (2009). *The Case for Contextualism: Knowledge, skepticism, and context*. Oxford University Press.

Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford University Press.

Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy*, *74*(4), 549 – 567.

Russell, G. K., & Doris, J. M. (2008). Knowledge by Indifference. *Australasian Journal of Philosophy*, *86*(3), 429 – 437.

Schaffer, J. (2006). The Irrelevance of the Subject: Against Subject-Sensitive Invariantism. *Philosophical Studies*, *127*(1), 87-107.

Schaffer, J., & Knobe, J. (2010). Contrastive Knowledge Surveyed. *Nous*, 1–34.

Schiffer, S. (2007). Interest-Relative Invariantism. *Philosophy and Phenomenological Research*, *75*(1), 188-195.

Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford University Press, USA.

Williamson, T. (2005). Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowledge. *Philosophical Quarterly*, *55*(219), 213–235.

Appendix 1: Utilities to the Game

To find a solution to the game we first need to find the expected utilities of any possible action by the attributor. Let p be the probability that the stakes for the subject are high, and q the probability that the stakes for the attributor are high. It is easy to see that $U(K|h) = 0$ because when the attributor is in high stakes, attributing knowledge to the subject gives her a payoff of 0, independently of the subject's stakes. A similar reasoning suggests that $U(\neg K|h) = 1$. Nonetheless, here are the calculations:

$$
\begin{aligned}
U(K|h) &= U(K|h \cap H)P(H|h) + U(K|h \cap L)P(L|h) \\
&= 0 \times P(H|h) + 0 \times P(L|h) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
U(\neg K|h) &= U(\neg K|h \cap H)P(H|h) + U(\neg K|h \cap L)P(L|h) \\
&= \frac{P(H)P(h|H)}{P(H)P(h|H) + P(L)P(h|L)} + \frac{P(L)P(h|L)}{P(H)P(h|H) + P(L)P(h|L)} \\
&= \frac{pq}{pq + (1-p)q} + \frac{(1-p)q}{pq + (1-p)q} \\
&= 1
\end{aligned}
$$

When the attributor is in low stakes we have:

$$
\begin{aligned}
U(K|l) &= U(K|l \cap H)P(H|l) + U(K|l \cap L)P(L|l) \\
&= P(L|l) \\
&= \frac{P(L)P(l|L)}{P(L)P(l|L) + P(H)P(l|H)} \\
&= \frac{(1-p)(1-q)}{(1-p)(1-q) + p(1-q)} \\
&= 1 - p
\end{aligned}
$$

$$
\begin{aligned}
U(\neg K|l) &= U(\neg K|l \cap H)P(H|l) + U(\neg K|l \cap L)P(L|l) \\
&= P(H|l) \\
&= \frac{P(H)P(l|H)}{P(H)P(l|H) + P(L)P(l|L)} \\
&= \frac{p(1-q)}{p(1-q) + (1-p)(1-q)} \\
&= p
\end{aligned}
$$

(8,905 words)