

Truth, Lies and Bullshit

distinguishing classes of dishonesty

Martin Caminada

University of Luxembourg

Abstract. In this paper, we distinguish three important classes of dishonesty that can occur in multi-agent systems, as well as in human society. In particular, the distinction is being made between lies and bullshit, following the work of Harry Frankfurt. The difference is that someone who tells a lie has access to the truth, whereas the concept of bullshit requires no knowledge of the truth at all. That is, the liar knows that what he says is not true, whereas the bullshitter has no proper knowledge to support the statements he or she is making. We point out that different situations, in multi-agent systems as well as in human society, provide strong individual incentives for bullshit. Overall, our analysis is meant to identify some particularly troublesome issues regarding reasoning in social context.

1 Introduction

One of the most salient features of our culture is that there is so much bullshit. Everyone knows this. Each of us contributes his share. But we tend to take the situation for granted. Most people are rather confident of their ability to recognize bullshit and avoid being taken in by it. So the phenomenon has not aroused much deliberate concern, nor attracted much sustained inquiry. In consequence, we have no clear understanding of what bullshit is, why there is so much of it, or what functions it serves. And we lack a consciously developed appreciation of what it means to us. In other words, we have no theory.

Harry G. Frankfurt, "On Bullshit" [8]

In his booklet "On Bullshit" [8] the American philosopher Harry G. Frankfurt provides a characterization of a class of dishonesty that is different and in some sense weaker than plain lies, but is not any less harmful in its capability to distort knowledge in a social setting. Although his booklet, published in 2005, became an almost instant bestseller, it has until now received remarkably little interest from researchers in formal epistemology, perhaps partly due to the somewhat provocative title. Nevertheless, Frankfurt's analysis is relevant not just from the perspective of philosophy, but also for the fields of sociology, formal logic and social epistemology.

Apart from treating Frankfurt’s work (Section 3), we also treat two other classes of dishonesty: lies (Section 2) and deception (Section 4). We then re-examine these three classes in the context of formal argumentation (Section 5 and 6), explain why these classes are relevant from the perspective of mechanism design (Section 7), what are the individual strategies of dealing with dishonesty (Section 8) and treat some real life examples (Section 9) before rounding off (Section 10).

2 On Lies

Defining exactly what a lie is has been the subject of quite some philosophical discussion. In its most simple form, a lie can be defined as the utterance of a statement which the speaker knows not to be true. That is, an agent A is lying on proposition p iff the following holds:

$$utters_A(p) \wedge K_A(\neg p)$$

More complex definitions of lying also explicitly take into account the intent that the hearer will adopt the false belief p [6]. However, for our purposes, the current simpler account of lying will suffice. A logical account of lying has been presented in [13]. Overall, the concept of lying is relatively well-studied and well-understood.

3 On Bullshit

In every day life, it is quite common for people to make statements of things they have no proper knowledge of. This is often done out of the desire to appear knowledgeable, even if one in fact is not. The situation here is different from the liar, who tells things he knows to be incorrect. How can one lie about something one has no knowledge about? Clearly, lying is not the right word to describe the basic concept here.

In the remainder of this paper, statements made without the speaker having sufficient knowledge about their validity will be referred to as “bullshit”, sometimes abbreviated to “BS”. We use this somewhat provocative term not only for its conciseness, but also to be in line with existing literature [8, 9] and to allow the reader to easily relate the phenomena described in this paper to his every day life experiences. As described in [8], the difference between lies and BS is that with lies, there exists a negative relation to the truth, whereas with BS, there is from the perspective of the speaker no relationship at all between his statements and the truth.

Bullshit is inevitable when people are forced to speak about subjects of which they possess no proper knowledge. Frankfurt claims that this is the direct consequence of the fact that in modern democratic society everyone is supposed to have an opinion about the current social and political issues, even if one does not have the time and means to be properly informed on all relevant aspects. In our view, however, there also exists a more mundane reason for the large amounts of ill-informed statements in the world around us. The point is that more and

more people started to make a living in professions that aim at generating, processing and providing information. Examples of this are journalists, business consultants, lawyers, financial analysts and even scientists. In these professions, it is vital to appear knowledgeable, even in situations where this is actually not the case. If it is not an option to honestly admit that one simply does not know, then the only thing to do is to generate BS.

In its simplest form, bullshit can be characterized as follows:

$$utters_A(p) \wedge \neg K_A(p) \wedge \neg K_A(\neg p)$$

As with lies, there is also an intensional aspect related to BS. Although one intends the hearer to believe that p , it is often more important that the hearer will believe that A is knowledgeable about p . While a liar has a very distinct purpose of wanting the hearer to believe p (because such a belief would have consequences that would suit the liar's goal), a bullshitter might be equally well off by telling the hearer that $\neg p$, as long as he appears knowledgeable in doing so.

4 On Deception

A third form of dishonesty to be discussed is that of deception. Although deception can be described in a very broad way [6], for current purposes we are interested in a more focussed concept of deception, as applied in [1]. The basic idea of deception is to provide the hearer with correct information, which the hearer is most likely to use to make an incorrect inference.

As an example, suppose one wants to persuade a friend to come over for the weekend. One could try to persuade him by claiming the newspaper predicts good weather this weekend, even though one knows that the local newspaper weather forecast is notoriously unreliable, and that the much more reliable TV-news predicts rain all weekend. In this case, one did not tell anything untrue, or lacking sufficient backing. The newspaper really *does* predict good weather. But by telling this to one's friend, he will make an inference that one knows to be incorrect, namely that this weekend the weather will probably be good. In essence, deception is a particular form of dishonesty that one can apply even without speaking anything else than the truth.

One of the interesting things about deception is that it depends on non-monotonic reasoning. Deception basically functions by providing some pieces of information and withholding other pieces of information in order to lead the victim to wrong conclusions. If we would tell that Tweety is a bird, without telling that Tweety is a penguin, the hearer would most probably derive that Tweety can fly, which we know to be wrong. With classical (monotonic) logic, this would not be possible. Withholding information in a classical formalism will result in inferences that are *missing*, whereas withholding information in a nonmonotonic formalism results in inferences that are *wrong*. With deception, one makes use of the nonmonotonic inference capabilities of the other person in order to implant wrong beliefs, without having to resort to lying ourselves.

5 Knowledge and Argumentation

In standard epistemic logic (S5), the possession of knowledge is basically a binary phenomenon. One either has knowledge about p or one does not. It is, however, also possible to provide a more subtle account of the extent to which one is knowledgeable about proposition p . Suppose Alex thinks that Hortis Bank is on the brink of bankruptcy because it has massively invested in mortgage backed securities. Bob also thinks that Hortis is on the brink of bankruptcy because of the mortgage backed securities. Bob has also read an interview in which the finance minister promises that the state will support Hortis if needed. However, Bob also knows that the liabilities of Hortis are so big that not even the state will be able to provide significant help to avert bankruptcy. From the perspective of formal argumentation [7], Bob has three arguments at his disposal.

A: Hortis Bank is on the brink of bankruptcy, because of the mortgage backed securities.

B: The state will save Hortis, because the finance minister promised so.

C: Not even the state has the financial means to save Hortis.

Here, argument *B* attacks *A*, and argument *C* attacks *B* (see Figure 1). In most approaches to formal argumentation, arguments *A* and *C* would be accepted and argument *B* would be rejected.



Fig. 1. Argument *C* attacks *B*, and argument *B* attacks *A*.

Assume that Alex has only argument *A* to his disposal. Then it seems to regard Bob as more knowledgeable with respect to proposition p (“Hortis Bank is on the brink of bankruptcy”) since he is better informed of the facts relevant for this proposition and is also in a better position to defend it in the face of criticism. The example also suggests that the common definition of knowledge as justified true belief might be too strong for many practical purposes, since it requires access to the truth in order to determine whether or not someone is knowledgeable. In many cases, such a direct access to the truth is not practically feasible. In our current world, we cannot objectively determine things like how long the credit crisis will persist, what would be the effect of 1 degree of global warming or how long global oil supplies will last. The most feasible way to determine whether someone is knowledgeable on these issues is to evaluate whether he is up to date with the relevant arguments and is able to defend his position in the face of criticism. This gives reason to a weaker definition of knowledge as justified belief. In cases where the objective truth cannot easily be accessed, one can then still say that agent X is more knowledgeable than agent Y iff it has to its disposal a larger set of relevant arguments.

We will now provide a more formal account of how the concept of knowledge could be described using formal argumentation. An *argumentation framework*

[7] is a pair (Ar, att) where Ar is a set of arguments and att is a binary relation on Ar . An argumentation framework can be represented as a directed graph. For instance, the argumentation framework $(\{A, B, C\}, \{(C, B), (B, A)\})$ is represented in Figure 1.

Arguments can be seen as defeasible derivations of a particular statement. These defeasible derivations can then be attacked by statements of other defeasible derivations, hence the attack relationship. Given an argumentation framework, an interesting question is what is the set (or sets) of arguments that can collectively be accepted. Although this question has traditionally been studied in terms of the various fixpoints of the characteristic function [7], it is equally well possible to use the approach of argument labellings [4, 5, 2]. The idea is that each argument gets exactly one label (accepted, rejected, or abstained), such that the result satisfies the following constraints.

1. If an argument is labelled accepted then all arguments that attack it must be labelled rejected.
2. If an argument is labelled rejected then there must be at least one argument that attacks it and is labelled accepted.
3. If an argument is labelled abstained then it must not be the case that all arguments that attack it are labelled rejected, and it must not be the case that there is an argument that attacks it and is labelled accepted.

A labelling is called complete iff it satisfies each of the above three constraints. As an example, the argumentation framework of Figure 1 has exactly one complete labelling, in which A and C are labelled accepted and B is labelled rejected. In general, an argumentation framework has one or more complete labellings. Furthermore, the arguments labelled accepted in a complete labelling form a complete extension in the sense of [7]. Other standard argumentation concepts, like preferred, grounded and stable extensions can also be expressed in terms of labellings [4].

In essence, one can see a complete labelling as a reasonable position one can take in the presence of the imperfect and conflicting information expressed in the argumentation framework. An interesting question is whether an argument *can* be accepted (that is, whether the argument is labelled accepted in at least one complete labelling) and whether an argument *has to be* accepted (that is, whether the argument is labelled accepted in each complete labelling). These two questions can be answered using formal discussion games [10, 14, 3, 2]. For instance, in the argumentation framework of Figure 1, a possible discussion would go as follows.

Proponent: Argument A has to be accepted.

Opponent: But perhaps A 's attacker B does not have to be rejected.

Proponent: B has to be rejected because B 's attacker C has to be accepted.

The precise rules which such discussions have to follow are described in [10, 14, 3, 2]. We say that argument A can be *defended* iff the proponent has a winning strategy for A . We say that argument A can be *denied* iff the opponent has a winning strategy against A .

If knowledge is defined not as justified true belief, but simply as justified belief, and justified is being interpreted as defensible in a rational discussion, then formal discussion games can serve as a way to examine whether an agent has knowledge with respect to proposition p , even in cases where one cannot directly determine the truth or falsity of p in the objective world. An agent knows p iff it has an argument for p that it is able to defend in the face of criticism.

The dialectical approach to knowledge also allows for the distinction of various grades of knowledge. That is, an agent X can be perceived to be at least as knowledgeable as agent Y w.r.t. argument A iff either X and Y originally disagreed on the status of A but combining their information the position of X is confirmed, or X and Y originally agreed on the status of A and in every case where Y is able to maintain its position in the presence of criticism from agent Z , X is also able to maintain its position in the presence of the same criticism.

When $AF_1 = (Ar_1, att_1)$ and $AF_2 = (Ar_2, att_2)$ are argumentation frameworks, we write $AF_1 \sqcup AF_2$ as a shorthand for $(Ar_1 \cup Ar_2, att_1 \cup att_2)$, and $AF_1 \sqsubseteq AF_2$ as a shorthand for $Ar_1 \subseteq Ar_2 \wedge att_1 \subseteq att_2$. Formally, agent X is at least as knowledgeable about argument A as agent Y iff:

1. A can be defended using AF_X (that is, if X assumes the role of the proponent of A then it has a winning strategy using the argumentation framework of X), A can be denied using AF_Y (that is, if Y assumes the role of the opponent then it has a winning strategy using the argumentation framework of Y), but A can be defended using $AF_X \sqcup AF_Y$, or
2. A can be denied using AF_X , A can be defended using AF_Y , but A can be denied $AF_X \sqcup AF_Y$, or
3. A can be defended using AF_X and can be defended using AF_Y , and for each AF_Z such that A can be defended using $AF_Y \sqcup AF_Z$ it holds that A can also be defended using $AF_X \sqcup AF_Z$,
4. A can be denied using AF_X and can be denied using AF_Y , and for each AF_Z such that A can be denied using $AF_Y \sqcup AF_Z$ it holds that A can be denied using $AF_X \sqcup AF_Z$.

Naturally, it follows that if $AF_Y \sqsubseteq AF_X$ then X is at least as knowledgeable w.r.t. each argument in AF_Y as Y .

In the example mentioned earlier (Figure 1) Alex has access only to argument A , and Bob has access to arguments A , B and C . Suppose a third person (Charles) has access only to arguments A and B . Then we say that Bob is more knowledgeable than Alex w.r.t. argument A because Bob can maintain his position on A (accepted) while facing criticism from Charles, where Alex cannot. A more controversial consequence is that Charles is also more knowledgeable than Alex w.r.t. argument A , even though from the global perspective, Charles has the “wrong” position on argument A (rejected instead of accepted). This is compensated by the fact that Bob, in his turn, is more knowledgeable than Charles w.r.t. argument A . As an analogy, it would be fair to consider Newton as more knowledgeable than his predecessors, even though his work has later been attacked by more advanced theories.

6 Argumentative Knowledge and Dishonesty

It is interesting to re-examine the earlier mentioned classes of dishonesty when the notion of knowledge is not justified true belief (S5) but simply justified belief (as was discussed in the previous section). The concept of lies is the most straightforward. If Bob would say that argument A should be rejected, he would be lying, since A has to be accepted in the argumentation framework Bob has at his disposal.

Bullshit can still be defined as making claims without having proper knowledge about them. The fact that knowledge has become a relative concept implies that BS has become a relative concept as well. For instance, Alex's claim that A should be accepted is more BS than Bob's claim that A should be accepted. In general, in order to make a claim in a knowledgeable way, one should try to be aware of possible counterclaims and the associated ways to dismiss them. If one simply makes a claim as soon as one sees a plausible reason for it, the result is very likely to be BS.

Also deception can quite easily be described in terms of argumentation. Suppose Bob has a reason for wanting Alex to reject argument A. Then he gives to Alex a subset of the arguments Bob has to his disposal (in this case B) such that these arguments, when merged with Alex's own arguments, change the status of A from accepted to rejected. So again, we can see deception as giving correct information ("The finance minister was on TV yesterday, promising that the state will save Hortis Bank.") that will lead the hearer to make an inference that one knows to be wrong.

7 Dishonesty and Mechanism Design

The difference between lies, bullshit and deception is also relevant from the perspective of mechanism design [11], because different mechanisms can provide (undesirable) incentives for different forms of dishonesty. As an example, consider the case of a financial adviser who advises his clients on investment products. If the adviser is paid on commission basis depending on the products his clients buy, then the incentive will be to advise those products that yield the highest commission. Thus, there will be an incentive for lies or, more likely, for deception. However, if the adviser is paid not on commission basis but directly by the client (say, based on an hourly fee) then he has no intrinsic bias anymore to advise product X above product Y. However, this is then replaced by a new kind of problem. For the adviser to earn his money, it is not required to actually give the best advice to his clients. After all, how would his clients be able to measure the quality of his advice? The fact that they are paying money for advice implies that the clients are to a large extent ignorant about the domain of expertise of the consultant. Thus, what matters is merely that the consultant *appears* to provide the right advice, that he *appears* to be knowledgeable. However, the task of gaining and maintaining real expertise is one that requires significant resources. Would it not be more attractive to give an advice that is perhaps not as informed as it appears to be?

The example of the financial adviser is interesting because it again illustrates an important difference between lies and BS. With lies (or deception), one has a clear interest in making the hearer believe very specific claims about the object world. With BS, at the other hand, one has no intrinsic interest in letting the hearer believe X or Y. All that matters is that the hearer believes that the speaker is knowledgeable about the claims it makes. If the financial adviser is paid by the client, he has no incentive whatsoever to lie. If he happens to have real knowledge available, he might as well tell his client the truth. However, if his expertise is limited, then it is in his best interest to conceal this from his client, and provide an advise that appears to be based on a level of expertise the adviser in fact does not have. It can hence be seen that different mechanisms provide incentives for different forms of dishonesty. If the adviser is paid on commission base, the incentive is to deceive. If he is paid by his clients directly, there is an incentive for BS.

8 Strategies for Dealing with Dishonesty

In the light of the above discussion, it is interesting to examine what strategies are available for dealing with dishonesty. For the consultant, the interests are relatively straightforward. The aim is to appear knowledgeable, without having to go through a great effort. There is a clear incentive to base the advise on a relatively small set of arguments, because obtaining more information would cost resources like time, effort and money. For the client, the interest is quite the opposite. The client is willing to pay money for the advise, as long as it is well-informed and takes into account everything that one might reasonably argue should be taken into account. How can the client evaluate the quality of the consultant's output even though she is not an expert?

A possible solution would be if the client had a small field of expertise that is a subfield of the expertise of the consultant. This then allows the client to take a "sample" of the advise of the consultant, and evaluate its well-informedness. This works not only for consultants, but in principle also for any information source. Consider the example of a magazine that specializes in international politics. Although one may not be an expert in international politics oneself, it is interesting to see what they write about one's own country (especially coming from a relatively small country). If on this particular topic it turns out that the information source is ill-informed, then it is a fair assumption that same information source will also be ill-informed on other topics.

Suppose the client would adopt such a strategy. What would be the best way for the consultant to react? Clearly, it still desires to appear knowledgeable in order to sell its advise (or to sell newspapers or magazines) but, still, doing extensive research costs time and money and should therefore preferably be avoided. It appears that one may still want to do a minimal effort, while at the same time prevent "being caught" on this. That is, the chance that the client is more informed than the consultant should be minimal. The client should not have a "larger" argumentation framework that allows the consultant's advise to

change status (from accepted to rejected or vice versa). Therefore, the consultant should have a good impression of the set of arguments that are most likely to be known by the clients. As an example, consider again the magazine on international politics and economics, whose name we shall not mention. One of its recent articles started with: “There is a Chinese saying ‘may you live in interesting times’.” I, as the reader, happen to know that no such Chinese saying exists, but that it is commonly believed in the West that it does. Therefore, as long as the readership consists of mainly Westerners, the chance that they will lose subscribers because of ill-informed claims is pretty minimal. All that matters is that the magazine is aware of what their readers are most likely to know and not to know. It would not dare to make the same mistake regarding English sayings.

There is yet another strong reason not to deviate from the group consensus when being a consultant. Whatever position one takes, there is always the risk of being wrong. When being wrong while participating in the group consensus, one can always claim that “we could not have known that...” or “at the knowledge that was available at that time, it seemed reasonable to assume that...”. One can simply claim to have been hit by a *Black Swan* [12]. The chance for a consultant to be singled out when the effects of bad advice finally become clear is significantly less if one tightly sticks to the group consensus.

This kind of behavior has consequences for the emergent behavior of the system as a whole. It simply implies that once there is a set of arguments that becomes fairly well-known, it will be in the consultant’s interest to amplify these arguments, whereas the relatively little-known arguments will not receive any attention and will therefore fail to become well-known. This then easily leads to a consensus of which most participants are not aware of how ill-informed it is. Great is their surprise when, often after considerable time, the group consensus turns out to be fatally wrong.

In the long run, we can describe the process of informedness in multi-agent systems as follows. It starts with a relatively new problem that becomes analyzed, and in the process of doing so, a particular set of arguments and points of view becomes dominant, and serves as a basis for the group consensus. Then, after a while, reality starts to break in, and it becomes clear that the group consensus was built on quicksand. Then follows a period of chaos, which finally results in a new group consensus, which lasts until again reality cannot be ignored anymore. The tragedy of this is that again and again decisions are being taken based on ill-informed analysis that results from flawed forms of collective reasoning.

9 On the Flaws of Collective Epistemics

Much of the ideas outlined above have been the result of some personal experiences of the author. For quite some time, I have been involved in a non-profit organization that tries to raise awareness of resource depletion, especially of mineral oil. We repeatedly had talks with people at (government) agencies on this issue, and it surprised us that almost every time we had more expertise on this

subject than our discussion partners. For instance, it is an often cited “truth” that the world still has 40 years worth of production in proven reserves in the ground. This is usually backed by referring to the official OPEC reserve data. However, less known is that these reserve data have been artificially inflated in the 1980s. The point was that at that time, the OPEC production quota came into effect, and these were based on reported reserves. So if a country has $x\%$ of the reserves in the ground then it gets $x\%$ of the production quota. This then provided a clear incentive to over report the reserves, especially since there was no independent auditing of it. For instance, Iraq under Saddam Hussein at some moment increased its reported reserves to a nice round 100 billion barrels, which was later increased to 112,5 (he simply added $1/8$) in order not to lose quota when other countries also increased. These reported reserves then became official data, used for predictions on how long the world’s oil reserves would last. From an abstract point of view, there are two relevant arguments: one that the oil reserves will last 40 years because this follows from the official data, and one that these official data are likely to be very unreliable. The second argument attacks the first one. Yet, in spite of everyone being aware of the first argument, almost nobody (not even at high levels) was aware of the second argument (which came from a small group of independent geologists). Also, we noticed how difficult it was to get people to discuss the real issues using real arguments. The people we talked to were often trained as economists, who were lacking any specific background on oil production and were simply not capable of assessing the quality of our arguments. Their general attitude was that the consensus was that there is still 40 years worth of oil reserves available, and that anyone who disagrees with this consensus is most likely to be wrong.¹ Although frustrating for us, this attitude was quite rational from their point of view. After all, if one does not have the expertise to assess the quality of arguments, or the ability to generate possible counterarguments (if applicable) one has to consider the risk of being deceived. The most rational thing to do would be to reject our argument altogether, and instead rely on collective opinion, hoping that this opinion has been shaped by people better informed than themselves.

People’s tendency to rely on group opinion when lacking knowledge individually is natural and quite understandable. An example from my own personal experience comes from the time I was staying in Japan. When taking the train, it sometimes occurred that something was announced (in Japanese only) and that then everybody would get out. So I also went out, not because I knew what was going on, but because I assumed my fellow passengers knew what was going on, and that copying their behavior was probably the most rational thing to do. However, this kind of behavior (following the crowd) depends on a critical assumption: that the group as a whole is better informed than oneself. Another interesting example would be the selection of a lawyer. It is for lay persons very

¹ Much of the “authoritative” analysis on oil supplies comes from the International Energy Agency, which is expected to provide quantitative analyses to the OECD governments, even though in many cases no reliable data regarding reserves and production is available.

difficult to assess the quality of a lawyer. Also, one cannot be guided by simple criteria such as the success rate, because no lawsuit is equal, and it might be that the lawyer has specialized in either easy or difficult cases. When we cannot assess the quality ourselves, then perhaps we should rely on the group consensus. In a free market, the group consensus with respect to the value of a product is reflected by its price. So, if we are involved in a lawsuit which is important to us and which we really have to win, then we should select the lawyer who charges the highest fee. This partly helps to explain why legal services are relatively high priced. Clients have little ways of evaluating their quality other than price. However, if nobody is actually able to objectively assess the quality of a lawyer, then the entire group consensus, reflected by the price, is not based on any knowledge at all.

One can observe a similar phenomenon on the stock markets, where the price-volatility tends to increase significantly in times of crisis and uncertainty. When it becomes more difficult to assess the real value of a particular stock, the best one can do is closely follow the opinion of other participants in the market, under the assumption that they are better informed than oneself, therefore amplifying any existing movements in the market.

As an aside, it might be interesting to examine the current credit crisis in the context of our analysis. To some extent, the credit crisis can be seen not so much as a failure of free markets, but as a failure of collective epistemics. When banking transformed itself from simple savings and lending to include more and more complex products like asset backed securities and credit default swaps, the precise risks and values of these new products were difficult to assess, which made it the task of a group of mathematical whizkids at financial institutions' risk assessment departments, whose positions depended on their perceived abilities to precisely quantify risks and valuations of these products, even though their actual capability to do so was hard to assess by anyone not belonging to this group. Apart from that, there were also rating agencies who were often paid by those involved in creating and selling the complex financial products and therefore had an incentive to deceive. In essence, the system had inherent vulnerabilities and it should not come as a surprise that these vulnerabilities have been exploited. Great was the surprise of European bankers when their balance sheets turned out to be full of American mortgages of people who essentially couldn't afford home ownership. Had the market been properly informed at an earlier stage, there would likely have been a correction in the valuation of these products, and the situation would not have gotten out of hand to the extent that we experience today. The current crisis can to a great extent be regarded as a failure of collective epistemics. How exactly to model this failure is one of the challenges of our research field.

10 Roundup

In this paper, we have provided a semi-formal account of three classes of dishonesty: lies, bullshit and deception. In particular, we have provided a theory

on what bullshit is, what purposes it serves for those who generate it, and why there is so much of it on a collective level. The current work can be considered as a first but necessary step in constructing formal computation models that take these phenomena into account. In particular, we are interested in constructing a multi-agent simulation system of clients and consultants in which both classes of agents try to apply their optimal strategy outlined in this paper. It would be interesting to examine to which extent the resulting consensus would be informed, and to which extent the little known “dissident” arguments would be pushed out of the process altogether. In general, we believe that the problem of collective irrationality is an important one since it touches the way in which our society is functioning.

References

1. Jonathan E. Adler. Lying, deceiving, or falsely implicating. *The Journal of Philosophy*, 94(9):435–452, 1997.
2. Martin Caminada and Yining Wu. An argument game of stable semantics. *Logic Journal of IGPL*, 17(1):77–90, 2009.
3. M.W.A. Caminada. *For the sake of the Argument. Explorations into argument-based reasoning*. Doctoral dissertation Free University Amsterdam, 2004.
4. M.W.A. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence; 10th European Conference, JELIA 2006*, pages 111–123. Springer, 2006. LNAI 4160.
5. M.W.A. Caminada. An algorithm for computing semi-stable semantics. In *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007)*, number 4724 in Springer Lecture Notes in AI, pages 222–234, Berlin, 2007. Springer Verlag.
6. Roderick M. Chisholm and Thomas D. Feehan. The intent to deceive. *The Journal of Philosophy*, 74(3):143–159, 1977.
7. P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n -person games. *Artificial Intelligence*, 77:321–357, 1995.
8. Harry G. Frankfurt. *On Bullshit*. Princeton University Press, 2005.
9. Harry G. Frankfurt. *On Truth*. Alfred A. Knopf, 2006.
10. H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.
11. Jeffrey S. Rosenschein and Gilad Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. The MIT Press, 1994.
12. Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.
13. Hans van Ditmarsch, Jan van Eijck, Floor Sietsma, and Yanjing Wang. On the logic of lying. 2007.
14. G. A. W. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA-00)*, number 1919 in Springer Lecture Notes in AI, pages 239–253, Berlin, 2000. Springer Verlag.