An Elementary Resolution of the Liar Paradox[1]

Abstract

The liar paradox is an ancient conundrum of logic. When an attempt is made to assign a truth value to the statement: "This statement is false;" a vicious cycle of contradiction ensues. A careful consideration of context, using a technique involving *context diagrams*, leads to a clear resolution of this paradox. In addition to context diagrams, the mathematical notion of a *meta-context* is also needed for resolving the liar paradox. Meta-contextual resolutions of related paradoxes, such as the strengthened liar paradox, are also described.

Introduction

The liar paradox is an ancient conundrum of logic. It was originally cast in the form of a fable [11]:

> In ancient times, all the inhabitants of Crete were incapable of making a true statement. Epimenides, who lived in Crete, made the following statement: "All Cretans are liars." Is Epimenides lying?

The essence of the liar paradox is captured by the following simple statement:

$$\textit{This statement is false.} \tag{1}$$

Any attempt to assign a truth value to the statement in (1) leads to a vicious cycle of contradiction (sometimes called a vicious circle [3]). If the statement is true, then it is false. But if it is false, then it must be true. Thus we have a vicious cycle of contradictory statements. Vicious cycles in logic, such as the liar paradox, have been studied intensively in recent years in computer science (see [8]).

Now that we have introduced the liar paradox, we outline the contents of this article. We begin by briefly describing some historical solution attempts, and then we introduce the idea of contexts in logic and describe the contradiction inherent in the liar paradox *within every context.* Meta-contexts are then introduced to resolve the liar paradox. Further applications of meta-contexts to other vicious cycle paradoxes, including the strengthened liar paradox, are then described.

Historical solution attempts

Such an ancient paradox has led, over many centuries, to a wide variety of approaches to its resolution. Some of those expounded in the last 30 years are summarized in [4]. One solution, proposed by various people ([6], for example), is to employ a hierarchy of levels of truth values in logic. The level for the word *false* in (1) is level 1, while the level of truth values for the entire statement is level 2. Hence (1) can be rewritten as

$$\textit{This statement is } (\textit{false})_1.$$

We should then attempt to determine if it is $(\text{true})_2$ or $(\text{false})_2$. This method is unsatisfactory precisely because it introduces a hierarchy of truth values. Such a hierarchy seems entirely ad-hoc and is not referred to in statement (1).

Several resolutions of the liar paradox have enlarged the realm of truth values from two to three ([4] and [5], for instance). In a three-valued logic, the truth values are true, middle, and false [12]. For such a logic, statement (1) is not paradoxical. However, (1) can be revised to

$$\textit{This statement is not true.} \tag{2}$$

With a three-valued logic, (2) is paradoxical. The statement in (2) is called the *strengthened liar paradox.* The method described in this paper will apply to a three-valued logic and provides an elementary resolution of (2).

We now briefly describe the idea behind our method of resolving these liar paradoxes. First, consider the fable of Epimenides. Although this fable is somewhat paradoxical, it can be resolved in various ways. For example, it could be asserted that Epimenides is incapable of making such a statement, since he must always lie. Therefore, the report that Epimenides made such a statement must be false. Another resolution would be to deny the truth of the hypothesis: It is impossible for the inhabitants of Crete to *always* make false statements. The key point to note here is that both of these resolutions *widen the context of discourse* beyond a simple historical fable. Our method of resolving the liar paradox (1) and strengthened liar paradox (2) is an abstract version of these two wider contextual resolutions of the fable of Epimenides. As an aid to our abstract, mathematical approach, we make use of illustrations which we call *context diagrams.*

Our method is a simplified version of Barwise's resolution of the liar paradox, based on situational logic. Barwise's work is described in great detail in [2] and [3], and introduced nicely in [9]. To the best of our knowledge, context diagrams are a new technique in situational logic.

Contexts in logic

Instead of considering some absolute notion of truth values, it is more realistic to examine the truth value of a statement within some context. For example, consider the two statements:

$$p: \quad 8+6=14$$

and

$$q: \quad 8+6=2.$$

The statement p is true when the context is integer arithmetic. If the context is mod 12 arithmetic, then statement p is false. Statement q, however, is true in the context of mod 12 arithmetic but false in the context of integer arithmetic.

For notation we let a context $\mathcal{C}$ be a set of statements such that when a statement p is true in the context $\mathcal{C}$, we write $p \in T(\mathcal{C})$. On the other hand, if statement q is false in the context $\mathcal{C}$, then we write $q \in F(\mathcal{C})$. These ideas are captured by a *context diagram,* as shown in Figure 1.

We now denote the statement in (1) by the letter p :

$$p = [\text{This statement is false.}]. \tag{3}$$

Given any context $\mathcal{C}$ containing p, the interpretation of p in that context leads us to the following abstract formulation of the liar paradox:

$$p = [\, p \in F(\mathcal{C})\,]. \tag{4}$$

This last formulation leads to the following basic theorem.

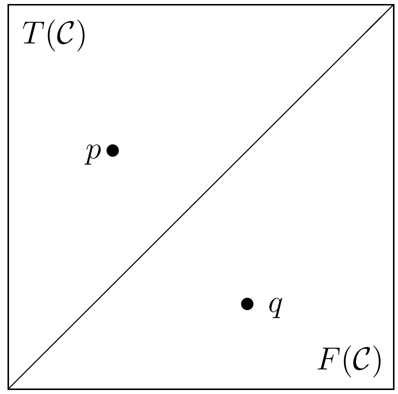


Figure 1: Context diagram illustrating p true and q false in the context $\mathcal{C}$.

Theorem 1. *For every context $\mathcal{C}$ containing p, the statement*

$$p = [\, p \in F(\mathcal{C}) \,]$$

is self-contradictory.

Proof. When reading this proof, it is important to keep in mind that we are assuming at every step that $\mathcal{C}$ is the context where we judge whether p is true or false.

If $p \in T(\mathcal{C})$, then p is true in the context $\mathcal{C}$. Since p asserts that $p \in F(\mathcal{C})$, it follows that $p \in F(\mathcal{C})$. Thus, we have shown that

$$p \in T(\mathcal{C}) \implies p \in F(\mathcal{C}).$$

On the other hand, if $p \in F(\mathcal{C})$, then the assertion of p that $p \in F(\mathcal{C})$ is true, hence $p \in T(\mathcal{C})$. Thus, we have also shown that

$$p \in F(\mathcal{C}) \implies p \in T(\mathcal{C}).$$

These two contradictory implications create the vicious cycle shown in Figure 2. □

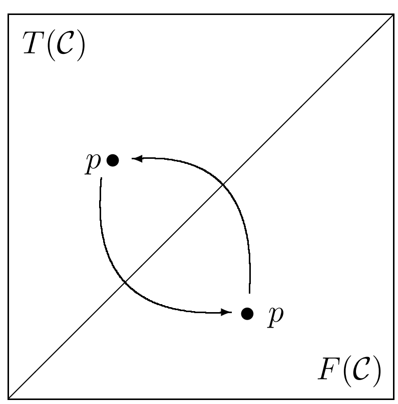


Figure 2: The statement $p = [\, p \in F(\mathcal{C}) \,]$ is self-contradictory in every context $\mathcal{C}$.

On the surface, we seem to have gotten nowhere with our notion of contexts. The liar paradox is still a vicious cycle, no matter what context is employed. However, by introducing the concept of a meta-context, within which an underlying context is analyzed from the outside, we shall be able to resolve the liar paradox.

Meta-contexts and the resolution of the liar paradox

The liar paradox can be resolved by using meta-contexts. Given a context $\mathcal{C}$, its meta-context $\widehat{\mathcal{C}}$ is the context defined as follows:

$$\widehat{\mathcal{C}} = \{\text{statements referring to } T(\mathcal{C}) \text{ or } F(\mathcal{C})\}.$$

For example, the liar paradox statement $p = [\, p \in F(\mathcal{C})\,]$ can be judged within the meta-context $\widehat{\mathcal{C}}$. In this section we show, for an appropriate context $\mathcal{C}$, that this statement can be assigned a truth value within the meta-context $\widehat{\mathcal{C}}$ in a non-contradictory way.

The specific context that we consider is the context $\mathcal{S} = \{\text{self-referential statements}\}$. It is easy to assign a truth value to some self-referential statements. For instance, consider the statement

$$q: \quad \text{This statement is a sentence.}$$

Clearly q is true in the context $\mathcal{S}$, i.e., $q \in T(\mathcal{S})$. On the other hand, consider the statement

$$r: \quad \text{This statement is written in French.}$$

Certainly r is false; hence $r \in F(\mathcal{S})$.

We know from Theorem 1 that the statement $p = [\, p \in F(\mathcal{S})\,]$ is self-contradictory in the context $\mathcal{S}$. However, the self-referential statement p does refer to $F(\mathcal{S})$, so it makes sense to consider its truth value in the meta-context $\widehat{\mathcal{S}}$. This leads us to the following theorem.

Theorem 2. *The statement $p = [\, p \in F(\mathcal{S})\,]$ is false in the meta-context $\widehat{\mathcal{S}}$. That is, $p \in F(\widehat{\mathcal{S}})$.*

Proof. Suppose $p \in T(\widehat{\mathcal{S}})$. Since p asserts that $p \in F(\mathcal{S})$, we conclude that $p \in F(\mathcal{S})$. When $p \in F(\mathcal{S})$, then $\mathcal{S}$ is a context for p. Hence, as in the proof of Theorem 1, we obtain a contradiction. Thus, by *reductio ad absurdum* as shown in Figure 3, we conclude that $p \in F(\widehat{\mathcal{S}})$.

Notice that $p \in F(\widehat{\mathcal{S}})$ does *not* imply that $p \in T(\mathcal{S})$. In the meta-context $\widehat{\mathcal{S}}$, the consequence of $p \in F(\widehat{\mathcal{S}})$ is that $p \notin F(\mathcal{S})$, but it is only in the context $\mathcal{S}$ that we would then be led to conclude that $p \in T(\mathcal{S})$. In the meta-context $\widehat{\mathcal{S}}$, the condition $p \in F(\widehat{\mathcal{S}})$ is consistent with $p \notin F(\mathcal{S})$ *and* $p \notin T(\mathcal{S})$. The context diagram in Figure 4 illustrates this point.

This completes the proof that $p \in F(\widehat{\mathcal{S}})$. □

Theorem 2 has resolved the liar paradox. We now can say, unequivocally, that the statement in (1) is false when judged in the meta-context $\widehat{\mathcal{S}}$. Furthermore, it is easy to see that for any other appropriate context $\mathcal{C}$ for which (1) can be expressed as $p = [\, p \in F(\mathcal{C})\,]$, it also follows that p is false in the meta-context $\widehat{\mathcal{C}}$. The proof is just like the proof of Theorem 2, with $\mathcal{S}$ and $\widehat{\mathcal{S}}$ replaced by $\mathcal{C}$ and $\widehat{\mathcal{C}}$, respectively.

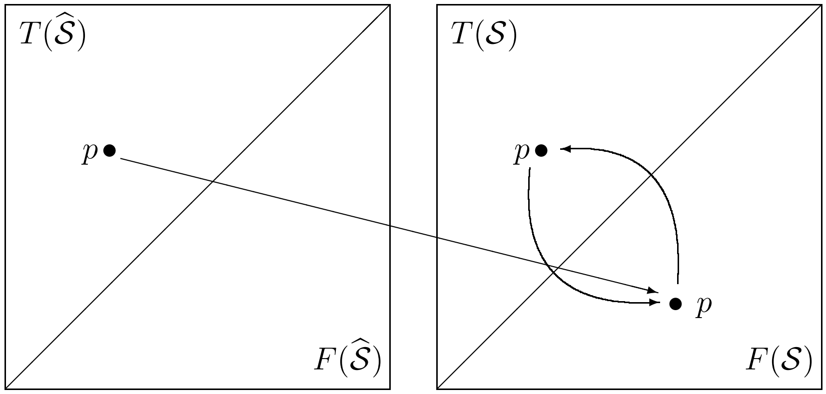


Figure 3: Assuming $p = [p \in F(\mathcal{S})]$ is true in the meta-context $\widehat{\mathcal{S}}$ leads to a self-contradiction in context $\mathcal{S}$. This is a *reductio ad absurdum* in the meta-context $\widehat{\mathcal{S}}$.

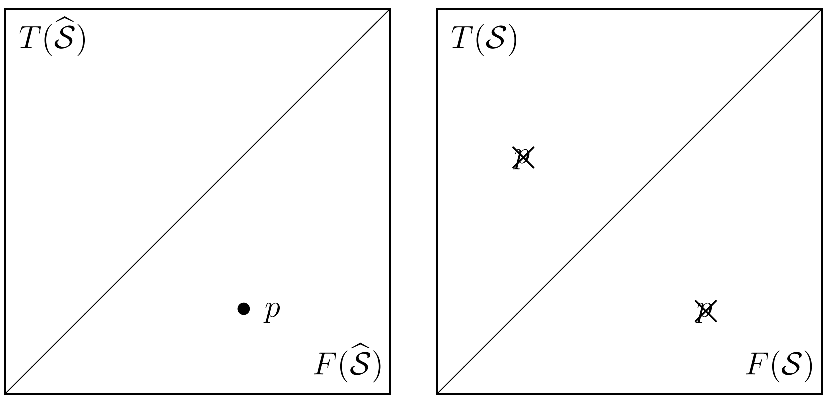


Figure 4: The conclusion $p \in F(\widehat{\mathcal{S}})$ of Theorem 2 is consistent with p not belonging to either $T(\mathcal{S})$ or $F(\mathcal{S})$.

An *implication diagram* can be drawn to illustrate our resolution of (1):

By labeling the statement as p, and then pointing outside the context (symbolized by the box drawn around the sentence comprising the statement), we predict the conclusion that p is false. It is interesting that this type of diagram can also be used to predict the more rigorous meta-contextual resolution of other paradoxes (see the concluding section).

Analogies to other approaches

We now provide a brief description of the relations between the meta-contextual resolution just described and other work on the liar paradox. There are loose analogies between the meta-contextual resolution and aspects of each of these other approaches.

For the resolution of the liar paradox which uses a hierarchy of truth values, the truth values in context $\mathcal{C}$ would be at level 1, while the truth values in the meta-context $\widehat{\mathcal{C}}$ would be at level 2. This is only a loose analogy, however, because consideration of context in logic does not require distinct notions of truth value for different contexts. The contexts are distinguished from each other, not the truth values.

Gödel's Incompleteness Theorem has sometimes been discussed in connection with the liar paradox [10]. There is a loose analogy here if the word undecidable is interpreted in a broad enough sense. The conclusion $p \in F(\widehat{\mathcal{S}})$ of Theorem 2 is consistent with p not belonging to either $T(\mathcal{S})$ or $F(\mathcal{S})$. In other words, p could be called undecidable in the context $\mathcal{S}$ while false in the meta-context $\widehat{\mathcal{S}}$. This shows some relation to Gödel's theorem. The relation is not a strong one, however, because the undecidability of p in context $\mathcal{S}$ (or any other context) consists in p being *self-contradictory.* So, unlike the situation in Gödel's theorem—where it is impossible to derive a proof of either p or of $\neg p$ within a given logical system—we have here a case of p and $\neg p$ being logically equivalent.

There is also a very loose connection between the context diagrams in Figures 3 and 4 and recent work [1] on a quantum mechanical perspective on the liar paradox. In [1], the statement p for the liar paradox is interpreted as a self-referent entity in a quantum mechanical setting. When this entity p is measured it enters into a true-false cycle, and when left unmeasured it remains fixed in its initial state. Figures 3 and 4 can be given (in a very loose sense) that kind of interpretation. In Figure 3, the arrow pointing from p in $T(\widehat{\mathcal{S}})$ to $p \in F(S)$ is analogous to the measurement of p (from the perspective of $\widehat{\mathcal{S}}$) which initiates the true-false cycling within $\mathcal{S}$. In Figure 4, the lone element $p \in F(\widehat{\mathcal{S}})$ corresponds to the lack of measurement of p.

Finally, there are resolutions ([4], [5], for example) of the liar paradox that make use of three-valued logic. In a three-valued logic, the statement in (1) can be assigned the value middle without causing any contradiction. There is a loose relation between this resolution and the meta-contextual resolution, in that the truth value of false in the meta-context $\widehat{\mathcal{C}}$ is outside of the context $\mathcal{C}$ with its two truth values of $T(\mathcal{C})$ and $F(\mathcal{C})$, and could thus be regarded as analogous to middle.

Three-valued logic resolves the liar paradox, but it cannot resolve the strengthened liar paradox. The statement in (2) is self-contradictory in a three-valued system. However, the method of meta-contexts is strong enough to resolve this strengthened liar paradox. We discuss this resolution in the next section.

Further applications of meta-contexts

The method of meta-contexts can be used to resolve several other paradoxes. As examples, we give the resolution of the problem of P. Jourdain's calling card [7], and of the strengthened liar paradox. We describe only the essential features of the arguments here, complete details being left to the reader.

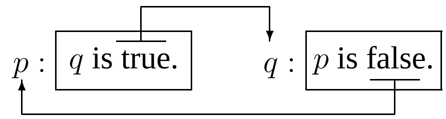
First, let's consider the problem of P. Jourdain's calling card. On the front, the card had this statement:

The statement on the other side of this card is true.

and on the back was this statement:

The statement on the other side of this card is false.

These two statements produce a vicious cycle of contradiction. In terms of a context $\mathcal{C}$, they can be formulated as $p = [\, q \in T(\mathcal{C}) \,]$ and $q = [\, p \in F(\mathcal{C}) \,]$. We leave to the reader the proof that these statements are contradictory within any context $\mathcal{C}$ whatsoever. However, by choosing an appropriate meta-context $\widehat{\mathcal{C}}$, the paradox can be resolved. The following implication diagram



predicts that $p \in F(\widehat{\mathcal{C}})$ and $q \in T(\widehat{\mathcal{C}})$. A rigorous analysis, using context-diagrams, confirms this result.

The method of context-diagrams and meta-contexts can also be extended to three-valued logics and used to resolve the strengthened liar paradox. Since a three-valued logic has truth values of true, middle, and false, the context diagram then consists of three parts, labeled $T(\mathcal{C})$, $M(\mathcal{C})$, and $F(\mathcal{C})$. To visualize this context diagram, the reader need only enlarge the diagonal that separates $T(\mathcal{C})$ and $F(\mathcal{C})$ in Figure 1 into a diagonal strip and label this strip $M(\mathcal{C})$.

In conclusion, we describe the meta-contextual resolution of the strengthened liar paradox. The statement in (2) can be rewritten in formal terms as

$$p = [\, p \in F(\mathcal{C}) \text{ or } p \in M(\mathcal{C}) \,].$$

In an appropriate meta-context $\widehat{\mathcal{C}}$ (for example, $\widehat{\mathcal{S}}$), the assumption that $p \in T(\widehat{\mathcal{C}})$ yields a *reductio ad absurdum.* The conclusion being that p is not true in the meta-context $\widehat{\mathcal{C}}$, i.e., $p \notin T(\widehat{\mathcal{C}})$. Since this last statement is consistent with $p \notin T(\mathcal{C}) \cup M(C) \cup F(C)$, there is no paradox. The following implication diagram illustrates this meta-contextual resolution:

Thus, we see that the strengthened liar paradox yields as easily to meta-contextual resolution as the original liar paradox does.

References

[1] D. Aerts, J. Broekaert, and S. Smets, The liar paradox in a quantum mechanical perspective. *Found. Sci.*, **4** (1999), 115–132.

[2] J. Barwise and J. Etchemendy, *The Liar.* Oxford Univ. Press, 1987.

[3] J. Barwise and L. Moss, *Vicious Circles.* CSLI Lecture Notes 60, 1996.

[4] S.V. Bhave, The liar paradox and many-valued logic. *Philos. Quarterly,* **42** (1992), 466–479.

[5] D.A. Bochvar, On a three valued logical calculus and its application to the analysis of contradictories. *Matematiceskij Sbornik (N.S.),* **4** (1938), 287–308. English transl. by M. Bergmann: On a three valued logical calculus and its application to the analysis of the paradoxes of the classical extended functional calculus. *Hist. Philos. Logic*, **2** (1981), 87–112.

[6] T. Burge, Semantical paradox. *J. of Philosophy*, **76** (1979), 174–196.

[7] J. Casti, *Complexification.* HarperCollins, New York, NY, 1994.

[8] G.J. Chaitin, *The Unknowable.* Springer-Verlag, 1999.

[9] K. Devlin, *Goodbye Descartes.* Wiley, 1997.

[10] J. Humphries, Gödel's proof and the liar paradox. *Notre Dame J. Formal Logic,* **20** (1979), 535–544.

[11] W. Poundstone, *Labyrinths of Reason.* Doubleday, 1988.

[12] H. Putnam, Three-valued logic. *Phil. Studies,* **8** (1957), 73–80.