REASON DEMANDS BELIEF IN INFINITELY MANY CONTRADICTIONS

Roy Sorensen

I believe that I believe an inconsistent proposition (i.e., one that entails its own negation—as demonstrable by reductio ad absurdum). This meta-belief about the set of propositions I actually now hold is infallible.

Proof by reductio ad absurdum: Assume each proposition I now believe is consistent—including the proposition that I believe an inconsistency. If \( p \) is consistent, then \( \neg p \) is inconsistent is itself inconsistent. For if \( p \) is consistent, it is logically impossible for \( p \) to entail its own negation. Yet if \( \neg p \) is inconsistent is consistent, then it is possible for \( p \) to entail its own negation. Just as it is inconsistent to call a consistent proposition inconsistent, it is inconsistent to say of two consistent propositions that either the first one is inconsistent or the second one is inconsistent. In general, it is inconsistent to say of a set of consistent propositions that at least one member of that set is inconsistent. Therefore, if each proposition that I now believe is consistent, then my meta-belief that at least one of my beliefs is inconsistent is itself inconsistent. That is, if all of my beliefs are consistent, then at least one of my beliefs is inconsistent. Contradiction. Therefore, necessarily, if I believe that I believe at least one inconsistent proposition, then I indeed believe at least one inconsistent proposition.

"The set of propositions I actually now hold" is a rigid definite description (like "the actual inventor of bifocals" which picks out Benjamin Franklin even in those worlds in which Franklin is not the inventor of bifocals). I need a rigid designator because my thesis is de re: Of the set of propositions that I actually believe, at least one of them is inconsistent.

"The set of propositions I actually now hold" is also an indexical expression. Its reference shifts with the circumstances of utterance. However, the phrase picks out the same kind of thing and there is much overlap between what I believed at past stages and what I currently believe (especially with respect to logical truths). Therefore, fruitful generalizations can be made about these belief sets. These generalizations are not merely autobiographical. Nothing special turns on who "I" refers to or which period is designated by "now" or which possible world gets selected by "actual." I am using "I" representatively like Rene Descartes uses "I" in his Meditations. The reader is intended to follow along with his own parallel thoughts about his own current set of beliefs and see his thoughts as representative of human beings in general.
I. BEYOND THE PREFACE PARADOX

If I merely believe that at least one of my beliefs is false, then it is impossible for that belief to be true along with all of my other beliefs. Ironically, this rationally mandatory meta-belief stops my beliefs from collectively constituting a coherent picture of reality (Makinson 1965). I used to take solace in the distinction between believing an inconsistency and merely having jointly inconsistent beliefs. A necessary condition of retaining this comfort is rejection of the agglomeration principle: \((Bp \& Bq)\) entails \(B(p \& q)\).

The surprise is that the denial of agglomeration is not a sufficient condition. My opening reductio secures an inconsistent belief independently of the agglomeration principle.

I must still deny agglomeration because it would saddle me with identifiable inconsistencies. In particular, I would believe the conjunction of my beliefs and the negation of that conjunction. To believe that a proposition is a contradiction is to believe it is false—and hence to not believe it. Lesson: the identities of my logical lapses are personal blindspots; though consistent “\(p\) is inconsistent but I believe it” is just one deduction away from G. E. Moore’s “\(p\) is false but I believe \(p\).” Only others can single out the inconsistencies that I presently believe.

Graham Priest (1987, 120) professes to believe the contradiction that Russell’s set is and is not a member of itself. Perhaps he does. But not simply in virtue of having thoughtfully and sincerely asserted it. Priest uses “contradiction” in a deviant way—he thinks a few contradictions are both true and false! This deprives contradiction of its constitutive role in reductio ad absurdum. Instead of being the point of closure in a premise-less refutation, the “contradiction” becomes a deniable assumption in a modus tollens argument. Consequently, Priest seems as much a skeptic about the existence of contradictions as he seems a daringly open-minded believer in contradictions.

Am I being paternalistic? No more than Professor Priest is when he interprets self-attributions of massive inconsistency by Heraclitus, Lao Tzu, and Engels (Priest and Routley, 1989). To avoid trivializing logic (by counting all propositions as true), Priest accepts the principle that the number of contradictions be minimized. Priest must wield the same principle against self-attributions of triviality. Anyone who accepts all propositions is in total (degenerate) agreement with any other acceptor of triviality. Hence they can only be interpreted as having distinct positions by overriding their thoughtful, sincere self-attributions of maximal inconsistency.

Some may feel that I have nonetheless accompanied Priest too far down the path to contradiction. They will insist “I believe an inconsistency” is less reasonable than “I believe a falsehood.” However, the evidence for my logical fallibility has the same pattern as my generic fallibility. The preface paradox uses my past errors and the errors of people similar to me. A subset of these errors, logical errors, provides ample inductive justification for my logical fallibility. Logicians are just as apt as physicists to apologize for the errors that are sure to exist in their books. Psychological studies of deduction provide ample experimental corroboration of the commonplace observation that human beings are prone to logical errors.

My belief in inconsistencies can also be established as an inference to the best explanation. The hypothesis that some of the propositions I believe are inconsistent simply accounts for how reductio ad absurdum
can teach me a lesson. This method of proof derives a contradiction from the proposition under attack. So it can refute me only if I believe some inconsistent propositions. The efficacy of *reductio ad absurdum* cannot be better explained by the hypothesis that I *merely* believe that I believe some inconsistencies. For that meta-belief entails that I would really believe an inconsistent proposition.

Many eminent philosophers have maintained that we can only believe what is possible. But their position has an unusual dialectical vulnerability. The mere belief that it is possible to believe the impossible ensures that it is possible to believe the impossible (Sorensen 1996). For if the belief is mistaken, then it is necessarily mistaken. Hence, the belief would itself be a belief in the impossible. Consequently, we need only disagree with the eminent philosophers to prove them wrong. Philosophy never gets easier than this.

The belief that I believe an inconsistency is a consistent report of the bad news that I believe an inconsistency. The meta-belief does not *cause* me to become inconsistent. It is overwhelmingly probable that the *reductio* scenario in which the meta-belief forces me into an inconsistency is logically impossible. Rather than being a spoiler, the meta-belief is almost certainly a non-intrusive observer, a truth that is logically but not causally self-supportive. And it is helpful. Adding “I have inconsistent beliefs” to my stock of beliefs makes it more complete and offers the best explanation of my logical curiosity.

I have considerable non-deductive evidence for believing it anyway. The surprise is that once I make the “inductive leap” to the meta-belief, there is no longer any chance of it being false. There is sugar in this humble pie! This multiplier effect has precedents. A non-introspective empiricist might infer that he has beliefs only on the basis of his behavioral or neural similarity with believers. His inductive argument yields a conclusion that is then infallibly believed.

**II. The Resemblance to the Liar is a False Alarm**

If I say “The next thing you say is true,” you might enmesh us in a liar paradox by saying “What was just said is false.” In addition to risky definite descriptions, there are risky existential generalizations:

List A:

1. There is an inconsistent proposition on list A.
2. Some people have blue eyes.

Since 2 is consistent, 1 is true only if it is inconsistent. If 1 is inconsistent, 1 is false. But then there would be an inconsistent proposition on the list and thus 1 would be true after all. How do I know that my meta-belief is not mired in a liar paradox like the existential generalization in List A?

The quick answer sounds circular: I know because I know my list of beliefs contains some inconsistencies. That is, I know my meta-belief has the same character as

List B:

1. There is an inconsistent proposition on list B.
2. Some people have blue eyes.
3. It is raining and it is not raining.

The existential generalization in B is risky in the sense that it would have been semantically defective (i.e., liar paradoxical) if the other items on the list had all been consistent. This risk is only an epistemic possibility given that the existential generalization is read
de re. That is, if we take “List B” to name the set of propositions composed of the items actually on that list, then in every possible world (B1) is free of semantic defect. For names are rigid designators.

Since my set of actual beliefs contains some inconsistencies, my meta-belief “I believe some inconsistencies” is also necessarily unparadoxical. Notice that this metaphysical assertion differs from the epistemological claim that “I believe some inconsistencies” is certainly unparadoxical. Many necessities are not certainties. After all, many mathematical truths will never even seem probable to us. The point is that my strong inductive evidence puts me in a position to assert that the meta-belief is not a risky sentence. Just as a look at my hand-held calculator puts me in a position to assert that there is no possible world in which 823 + 483 = 1307, reflection on my fallibility puts me a position to assert that there is no possible world in which my de re meta-belief gives rise to a paradox.

Still, one may insist that epistemic possibility is relevant in argumentative contexts. My reductio ad absurdum may be necessarily free of any liar-paradoxical element. But to know that it is a good argument, it seems I must have independent knowledge that I believe some inconsistent propositions. Therefore, my critic could concede that induction gives me this independent knowledge but still complain that the induction renders my original argument superfluous.

However, not all of the preconditions of a successful argument are premises. Whenever I argue, I assume my rule of inference is valid. But as illustrated by Lewis Carroll’s tale of Achilles and the Tortoise, categorizing this assumption as a premise generates an infinite regress. A similar infinite regress follows if we categorize as a premise my assumption that I am not equivocating, or my assumption that my premises are jointly consistent, or my assumption that my argumentative discourse is composed of meaningful sentences. People can learn by argument even when they lack proof for the various propositions that constitute the infrastructure of argument. My inability to prove that I did not equivocate does not always stop me from acquiring knowledge by means of the argument in question. A proof that I did not equivocate might merely play the role of confirming that my knowledge of the conclusion was not defeated by an equivocation. In my opening reductio, I did assume that my reasoning is not liar-paradoxical. This assumption is a precondition of expressing a premise. Therefore, when I provide inductive evidence that this precondition is satisfied, I am only showing that the argument is not a pseudo-reductio. The project of showing that someone learned a conclusion from an argument is more onerous than simply learning the conclusion from the ostensible argument. Non-logicians learn much from their deductions without understanding how they do it. So do logicians!

III. Self-reference is Better than Hierarchy

Wise men have advised me to avoid even the appearance of paradox by casting my thesis hierarchically. Instead of (self-referentially) saying that my entire corpus of beliefs contains some inconsistencies, their counsel is to narrow my thesis to the claim that some of my first-order beliefs are inconsistent. If all of these first-order beliefs are consistent, then I can report (at the third order) that this second-order belief is inconsistent.

All of this is true and compatible with my original argument. I prefer my original self-referential argument because it is less artificial and complicated, and hence a better candidate for the aesthetically sensitive
status of “inference to the best explanation.” My grounds for believing that I believe inconsistencies are generic ones about the nature of belief. I have no special reason to think that my first-order beliefs are inconsistent. I form inconsistent beliefs because consistency competes with other desiderata such as simplicity and completeness. Induction confirms the hypothesis that the tradeoffs occur at all levels of beliefs: my beliefs about my beliefs are just as frequently inconsistent as my first-order beliefs. For instance, many of my past beliefs about principles of “doxastic logic” were refuted by reductio ad absurdum. So my only motive to shepherd my beliefs into separate orders is the “risk” of the liar paradox. This danger has already proven to be objectively baseless.

Therefore, my fear should cease. An argument cannot be refuted by its mere resemblance to a paradoxical form of reasoning. Refutation by logical analogy is a danger only to the extent that the resemblance is a sign that there something wrong with the argument itself. One should relax after one finds that the argument is free of the flaw suggested by the resemblance. Those who treat resemblance as a cause fall into the pseudo-scientific thinking that lies behind the voodoo practice of sticking pins in dolls. It is pseudo-scientific to doubt Cantor’s diagonal argument or Godel’s incompleteness theorem or Turing’s result for the halting problem merely because of their historic connection and similarity to the liar paradox. To reconfigure my argument in terms of hierarchy would just pander to a logical superstition.

IV. COUNTER-LOGICALS AND REDUCTIO ILLUSIONS

When I assert “If the vase falls, it will smash,” I conversationally implicate that the vase could fall. This implicature leads us to misinterpret counter-logical conditionals in general and reductio ad absurdum reasoning in particular.

Conditionals with impossible antecedents are standard in reductio ad absurdum proofs. Indeed, such conditionals are indispensable because the reductio fails unless the supposition is inconsistent. Thus in my opening proof, I did not suggest that the antecedent of the following conditional is possible.

(i) If I mistakenly believe that I believe at least one inconsistency, then this meta-belief is itself inconsistent.

Granted, the antecedent is “epistemically possible.” But an epistemic possibility need be no more a possibility than a suspected murderer need be a murderer. The conditional must be understood in the same stern way as other contrary-to-necessity conditionals such as Euclid’s “If there is a largest prime number, then there is a larger prime number less than or equal to the successor of the product of all the prime numbers.” If Euclid were to allow that the antecedent is possible, he would be committed to agreeing that it is necessary. The whole proof method of reductio ad absurdum would backfire.

There is irony in the assertion “My belief that I believe an inconsistency is infallible in virtue of its possible inconsistency.” The irony would be outright absurdity if “possible” here meant more than epistemic “possibility.” Nothing is both possible and possibly impossible. Every possibly inconsistent belief is necessarily inconsistent, so every possibly inconsistent belief is necessarily false—the reverse of infallible.

V. CONTINGENT MEMBERSHIP LETS YOU KEEP ESSENTIAL PROPERTIES

The Barcan formula, (x)□Fx ⊃□(x)Fx, licenses the inference from a de re necessity
to its \textit{de dicto} counterpart. An apparent counterexample is a world that happens to be filled with immaterial things (numbers, God, sets). Each member of this domain possesses its immateriality necessarily. Yet there could have been some material beings in this world.

Now consider a blackboard that happens to hold only consistent propositions. Each proposition is necessarily consistent: \((p)\Box [p \text{ is consistent}]\). However, it is not a necessary truth that the blackboard comprises only consistent sentences. There could have been some inconsistent sentences. (Although the blackboard scenario involves an unrestricted modality, logical possibility, the domain of discourse is restricted to the sentences on the blackboard. Even friends of the Barcan formula say that the formula only holds when \textit{both} the modality and the quantification are unrestricted [Parsons 1995].) Distinguish the true \((p)\Box [p \text{ is consistent}]\) from the false \(\Box (p) [p \text{ is consistent}]\).

The point can be put unequivocally by fixing the domain of discourse with a rigid designator. Name the set of propositions expressed on the blackboard “B.” Notice that I do not need to know which propositions are on the board. I can use the reference-fixing description “the propositions expressed on that blackboard” to fix the membership of B without knowing which propositions. The blackboard could be covered with a curtain. My christening of the set ensures that B picks out the same propositions in every possible world. Then from \( (p)\Box [p \text{ is consistent}]\) we can conclude \(\Box (p) [p \text{ is a member of } B] \Rightarrow \Box [p \text{ is consistent}]\). Or to put the same conclusion negatively, it is necessarily false that some proposition in B is inconsistent: \(\Box - (\exists p) (p \text{ is a member of } B \& [p \text{ is inconsistent}]\). More generally, it is inconsistent to say of a set of consistent propositions that at least one of them is inconsistent or even that one of them is possibly inconsistent (or even possibly possibly inconsistent, etc.) Thus the opening \textit{reductio} can be easily changed into a proof of the following thesis: if I believe that one of my beliefs is \textit{possibly} inconsistent, then one of my beliefs is \textit{actually} inconsistent.

I am like the blackboard. I do not need any introspective power to specify the set of propositions I believe. And the propositions I hold are only contingently held. But their contingent membership in my belief system is compatible with each of them being necessarily consistent. Likewise, the fact that I only contingently believe a particular inconsistency is compatible with it being necessarily inconsistent. My belief that I believe at least one inconsistent proposition is a belief about the nature of the beliefs I actually hold. Whichever beliefs I happen to hold, there will be an inconsistent belief—given that I also hold, as I ought, that at least one of my beliefs is inconsistent.

\section*{VI. Unreflective Inconsistency}

The full-strength argument springs from a global, self-referential meta-belief and employs \textit{reductio ad absurdum}. However, the basic case for believing contradictions can be made without any of these anxiety-provoking features.

Consider a student who is given a test in propositional logic. He is required to pick as many truths as he can from a list. The list is composed solely of logical truths and logical falsehoods but the student has not been told this. The student believes each of his answers, \(p_1, p_2, \ldots, p_n\). However, he also believes that at least one of these answers is false, i. e., he believes \(\sim(p_1 \& p_2 \& \ldots \& p_n)\).

Here is a direct proof (by constructive dilemma) that the student believes a logical
falsehood. If any of his answers \( p_1, p_2, \ldots, p_n \) are false, then the student believes a logical falsehood (because the only falsehoods on the question list are logical falsehoods). If all of his test answers are true, then the student believes the following logical falsehood: \( \neg(p_1 \land p_2 \land \ldots \land p_n) \). For if \( p_1, p_2, \ldots, p_n \) are true, they are all logical truths. A conjunction of logical truths is itself a logical truth. And the negation of any logical truth is a logical falsehood. Hence, if all the student’s test answers are true, then his belief that \( \neg(p_1 \land p_2 \land \ldots \land p_n) \) is itself a belief in a logical falsehood.

With minimal modal logic (such as the system \( T \)), this proof generalizes to a broader reading of “non-contingent.” Just substitute “necessary” for “logical” in the above two paragraphs. The result is a sound proof about necessary falsehoods (since necessity collects over conjunction). For instance, an essentialist could have the student pick from metaphysical necessities and impossibilities such as “Water is \( \text{H}_2\text{O} \),” “This lectern is made of ice,” “Man is a rational animal,” etc. If this metaphysics student disbelieved the conjunction of his sincere answers, then he would thereby believe at least one necessary falsehood.

If the disbelieved necessary truth is \( \text{a posteriori} \), then the student’s disbelief does not raise issues about his rationality. In contrast, disbelief in an \( \text{a priori} \) truth betokens carelessness. Investigation of this asymmetry threatens to lead us into the labyrinth of modal epistemology. Happily, I can afford a retreat from the metaphysics student to the logic student. For the argument featuring the logic student involves only sentence logic and yet takes us a long way to showing that rational agents believe contradictions.

My assumptions about the logic student are minimal. I do not assume that he is applying the concept of a logical truth. His task is only to identify truths. Nor do I require that the student apply the concept of belief. Some introverted students believe that \( \neg(p_1 \land p_2 \land \ldots \land p_n) \) on the strength of reflections on their personal fallibility. But such meta-beliefs are not essential to the anti-agglomerative belief pattern: \( (Bp_1 \land Bp_2 \land \ldots \land Bp_n) \land B \neg(p_1 \land p_2 \land \ldots \land p_n) \). An unreflective student may base his belief that \( \neg(p_1 \land p_2 \land \ldots \land p_n) \) directly on the improbability of such a long list containing only truths.

The student could shed his belief in a contradiction by gaining new evidence. For instance, if the teacher showed the student the answer key, the student might learn on authority that \( p_1 \) is false and so stop believing the contradiction that \( p_1 \).

However, other contradictions are likely to remain after a genuine contradiction is uprooted. For instance, the student might have a belief in a contradiction that arose from an earlier tautology test. The student answered that \( q_1, q_2, \ldots, q_m \) are each true but the student also believed \( \neg(q_1 \land q_2 \land \ldots \land q_m) \).

A second answer key could eliminate the contradiction dwelling within this anti-agglomerative constellation. However, purging contradictions by appealing to authority becomes less feasible as the number and complexity of the statements increase. Instructors are themselves fallible. True, there are algorithms for deciding whether a given formula is a tautology in sentence logic. But they are too time-consuming to be a complete, practical remedy. For instance, the truth table for a sentence that is composed of \( n \) sub-sentences has \( 2^n \) lines. This is an exponential function that grows unmanageably complex as \( n \) increases. In any case, there is no such algorithm once we turn to predicate logic.
VII. Awareness That One Believes a Contradiction

It might be thought that the contradiction will go away if the student becomes aware of his situation. Very well, let the instructor carefully explain the meaning of “contradiction” and “tautology.” Have the student master the above proof that he believes a contradiction.

This tutorial might have the byproduct of improving the student’s ability to discriminate between the truths and falsehoods on the question list. It might also intensify the student’s desire to eliminate the false belief. Many people feel that belief in a logical falsehood indicts their rationality in a way that belief in a false contingency does not. After all, if one unwittingly believes a contradiction, C, then one is disposed to reject valid inferences such as “C, therefore, A” (where A is an arbitrary proposition). One will also be disposed to accept invalid inferences such as “A, therefore, ‘C is consistent’” (because one will view “C is consistent” as tautology). Hence, concern about one’s rationality may stimulate further inquiry or prompt heightened standards of evidence.

However, even authors of logic textbooks make mistakes. The logician’s belief that his solutions manual has errors is sometimes volunteered explicitly in the preface. Hence, the student’s anti-agglomeratvity can continue even after he has mastered logical theory and emulated his instructor’s scholarly virtues. The student can have a stable, rational belief in a contradiction.

VIII. Meta-beliefs and Anti-agglomerativity

Even more stable would be the belief that one believes a contradiction. This meta-belief is apt to arise from the very process of purging a contradictory belief. A person who discovers that one of his beliefs is a contradiction is like a man who discovers a louse on his head. The detection leads to a response that decreases the amount of harm—immediate termination of the louse. But the detection also constitutes evidence that there are further lice. The man who eliminates a contradiction benefits himself by causing himself to become more rational. However, the process is humbling because it turns up evidence that he has further irrationalities.

This modesty becomes a further basis for believing that one believes an inconsistency. It is logically independent of anti-agglomerativity:

Modesty: B\[\sim(p_1 & p_2 & \ldots & p_n) \& (Bp_1 \& Bp_2 \& \ldots \& Bp_n)\]

Anti-agglomerativity: \(Bp_1 \& Bp_2 \& \ldots \& Bp_n\) & \(B\sim(p_1 & p_2 & \ldots & p_n)\).

Since belief distributes over conjunction, both principles entail disbelief in the conjunction: \(B\sim(p_1 & p_2 & \ldots & p_n)\). However, anti-agglomerativity fails to entail what modesty explicitly asserts: the meta-belief that some of one’s beliefs are false. And only the anti-agglomerative proposition implies that one really has the individual beliefs. \(Bp_1 \& Bp_2 \& \ldots \& Bp_n\).

Anti-agglomerativity would be equivalent to modesty if each individual were omniscient about his own beliefs. A strong logic of belief could approximate this inner omniscience by combining the principle that we are infallible about our beliefs (if BBp then Bp) with the principle that beliefs are self-intimating (if Bp then BBp). Since both of these principles are implausible, no logic of belief challenges the independence of modesty and anti-agglomerativity.

Only anti-agglomerativity implies that the agent believes a proposition that is jointly inconsistent with other propositions.
that the agent believes. Modesty does entail that the agent has a false belief. But this kind of inevitable falsehood need not be in virtue of the propositions believed. Consider a deluded “author” who falsely believes he has written a book. He composes a preface in which he apologizes for the mistakes in the text. Since there is no text, the author’s preface belief “There is a false belief in the text” is false. The author’s meta-belief guarantees that he has at least one false belief. However, the author’s meta-belief does not guarantee that the propositions he believes are jointly inconsistent. For the deluded author only has one relevant belief and the proposition he believes is consistent.

Unlike de dicto modesty, de re modesty implies the existence of other beliefs. If the author believes of the beliefs expressed in the text, that one of them is false, then the propositions he believes are jointly inconsistent. For if all of his text beliefs are true, then his preface belief is false.

De re modesty implies anti-agglomerativity but not vice versa. The student who disbelieves the conjunction of his test answers need not realize that those are his answers. He might have just picked up a test at random and predicted, on actuarial grounds, that not all the answers are correct. In summary, de dicto modesty is not enough to entail joint inconsistency while de re modesty is more than enough. Anti-agglomerativity has the distinction of being exactly enough.

The deluded author shows how the absence of beliefs can force a distinction between the inevitability of a false belief and a joint inconsistency. However, the distinction is also prompted by the presence of beliefs that a meta-belief excludes:

1. I have no disjunctive beliefs.
2. Either Al Gore will win or Newt Gingrich will win.

Although 1 and 2 are jointly consistent, my belief in both of them would imply that I have a false belief (in particular, the meta-belief is false).

We know that “◊p” does not entail “◊(p & Bp)” because p could be a proposition about belief such as “No one has a belief.” This helps to explain relative blind spots in which each conjunct of a consistent conjunction is consistently believable but the consistent conjunction as a whole is not consistently believable.

IX. META-STATEMENTS

Meta-beliefs are beliefs about beliefs. Meta-statements are statements about other statements such as “‘Some sentences are long’ is a short sentence.”

A meta-statement can be jointly inconsistent with its object statement:

(a) “Snow is white” is false.
(b) Snow is white.

Similarly (b) is jointly inconsistent with
(c) “Snow is white” is inconsistent.

Indeed, since (b) is consistent, it is actually (c) that is inconsistent. For if (b) is consistent, then it is not a logical truth that a contradiction can be derived from (b). But if (c) were true, then it would be a logical truth that a contradiction can be derived from (b). Since a logical falsehood is derivable from (c) (namely, that a logical falsehood is derivable from [b]), (c) must itself be a logical falsehood.

In general, false accusations of inconsistency are themselves inconsistent:

(I) If p is consistent, then [p is inconsistent] is inconsistent.

The principle bears a likeness to the characteristic formula of S5, ◊p ⊃ □◊p, which states that whatever is possible is necessarily possible. The analogy invites the inference to
(II) If \( p \) is inconsistent, then \( \neg p \) is consistent is inconsistent.

The reasoning for (II) echoes that of (I). If \( p \) entails a contradiction, then \( \neg p \) is consistent entails that \( p \) does not entail a contradiction. But \( \neg p \) does not entail a contradiction entails that \( p \) entails a contradiction\( \neg p \) is a contradiction (by principle I).

Combining (I) and (II) yields the moral that statements bear their consistency status in a logically mandatory fashion just as modal statements bear their modal status necessarily. All errors about whether a statement is consistent are inconsistencies.

It follows that a policy of refusing to form opinions about non-contingencies is self-defeating. If I resolve to form beliefs only about contingencies, then I must form beliefs about whether candidates for beliefs are contingent. The problem is that this higher order belief is never itself a belief in a contingent proposition. If I am right about whether \( p \) is contingent, then I am necessarily right, and if wrong, then necessarily wrong.

Principle (II) suggests that there is also a danger in avoiding the incoherence of agnosticism by positively believing that one’s beliefs are jointly consistent. For if they are not jointly consistent, then one’s meta-belief is a logical falsehood. Those who argue that we should believe our beliefs to be jointly consistent are likely to be promoting a belief in an inconsistency. Ironically, their effort to prevent jointly inconsistent beliefs leads to the more severe sort of inconsistency—believing an inconsistency. The inconsistency would be the very assertion of joint consistency!

X. THE VIRAL THEORY OF INCONSISTENCY

J. R. Lucas’s (1964) Godelian argument that he is not a machine requires Lucas to believe that he is consistent. Computers physically instantiate formal systems. A consistent formal system that is powerful enough to prove the theorems of arithmetic is also powerful enough to construct a sentence with the same effect as “This sentence is not a theorem of the system.” The sentence could only be a theorem if its negation was also a theorem. So given that the system is consistent, the sentence is a truth that is missed by the system. Therefore, the system is incomplete. Lucas contends that he is not limited by this incompleteness result.

We have adequate, more than adequate, reason for affirming our own consistency and the truth, and hence also the consistency, of informal arithmetic, and so can properly say that we know, and that any machine representation of the mind must manifest an output expressed by a formal (since it is a machine) system which is consistent and includes Elementary Number Theory (since it is supposed to represent the mind). (Lucas 1996, 121)

I regard total consistency as an unattainable ideal. To maximize consistency, we must adopt policies that precipitate some inconsistency as a side effect. Just as our immune system is constantly combating incipient infections, our belief system constantly struggles to detect and eliminate contradictions. Both systems operate under the assumption that enemies already lurk within the gates. This assumption is probably hard-wired, making the meta-belief in my own inconsistency innate. Since this de re meta-belief in my own inconsistency is infallible, it would follow that I am an innately inconsistent being.

Lucas pictures contradictions as occasional intruders—like squirrels in your attic. The squirrels can be detected with reasonable vigilance and then promptly expelled:
The fact that we are all sometimes inconsistent cannot be gainsaid, but from this it does not follow that we are tantamount to inconsistent systems. Our inconsistencies are mistakes rather than set policies. They correspond to the occasional malfunctioning of a machine, not its normal scheme of operations. Witness to this is that we eschew inconsistencies when we recognize them for what they are. If we really were inconsistent machines, we should remain content with our inconsistencies, and would happily affirm both halves of a contradiction. Moreover, we would be prepared to say absolutely anything—which we are not. (Lucas 1964, 53)

I agree that detected contradictions are instantly abandoned. (We cannot believe what we regard as false.) But this very process of elimination creates a selective advantage for hidden contradictions. Contradictions rarely parade as $P \& \neg P$. Most are small, quiet, and colorless. Like biological parasites, contradictory beliefs jeopardize their own survival if they invite their hosts to make reckless deductions. A contradictory belief that did little harm or which even helped its host would have better long-term prospects.

Each belief is formed under pressure for completeness. Consequently, beliefs tend to crowd each other in logical space. Any combination of beliefs might be jointly inconsistent. Given a set of $n$ propositions, there are $2^n - 1$ combinations of propositions. Therefore, the search space for a check of consistency grows exponentially. Since this version of the satisfiability problem is NP complete, it is intractable even for future super-computers. (The most accessible discussion of NP completeness is the ninth chapter of William Poundstone’s *Labyrinths of Reason*. The classic presentation is Michael Garey’s and David Jonson’s *Computers and Intractability*.)

Inconsistencies come in quantity. Our belief system is under relentless, massive assault. Contradictions are not like squirrels; they are like viruses. By strength of numbers, some contradictions inevitably slip through.

The only feasible way to cope with contradictions is with fallible heuristics rather than algorithms. Recognition of this necessity forces the adoption of a policy that guarantees, as a foreseeable side effect, that one will have some contradictions. For instance, instead of devoting all resources to preventing entry, we devote substantial resources to the pursuit and elimination of the contradictions that do slip through. We deliberately sacrifice opportunities to be contradiction-free for the certainty of having a tolerable frequency of contradiction. An analogy: after disease fighters learned that cowpox immunized dairy workers from smallpox, they deliberately infected all of their patients with the cowpox disease. The physicians realized that most unvaccinated people end up in the ideal position of contracting neither smallpox nor cowpox. But they calculated that it was better to accept the certainty of a small loss rather than risk a much larger loss.

The most lucrative investment I could make with the dollar in my pocket is to buy the winning ticket in the state lottery. But prudence dictates that I should act in way that would preclude this optimal outcome—by investing in something with better odds. The good should not be sacrificed in pursuit of the perfect. Well, let’s not underestimate: the perfect *should* be sacrificed in pursuit of the good. Instead of maximizing my chance at achieving a maximal outcome, I should maximize my expected utility—even at the cost of rendering a maximal outcome impossible. I should burn my bridges to perfection—for the right price.

This point also holds for epistemic value. If I believe that I have at least one false
belief, then I preclude the optimal outcome of having entirely true beliefs. But it is a good bargain. The probability of having entirely true beliefs is so low that I lose little by further reducing the probability to zero. In exchange, I acquire an interesting, true belief. The same holds for more humbling self-attributions—the belief that some of my beliefs are inconsistent, ill-founded, and so forth. These contribute to a useful self-profile that helps me to tailor my cognitive practices to my strengths and weaknesses. The meta-beliefs put me out of contention for the title of “ideal believer” but it would be irrational for me not to take myself out of the running. It is rationally mandatory to acknowledge cognitive imperfections even when this acknowledgement is logically self-fulfilling.

Epistemologists have naturally simplified the agents with an assumption of perfect rationality. Of course, they have been aware that ordinary agents are imperfect: human beings are not logically omniscient, consistent, and so forth. Yet the commentators have still forbidden the agent from forming beliefs that would foreclose the attainment of perfectionist ideals. For instance, Keith Lehrer says that I ought not to believe “At least one of my beliefs is false” because that would doom me to having jointly inconsistent beliefs (1974, 203). Such counsel against self-injury may seem even more compelling when it comes to contradictions rather than mere falsehoods. But this is a perfectionist illusion.

XI. THE SCALE OF THE INCONSISTENCY

The logic student illustrates the anti-agglomerative path to contradiction. His beliefs just naturally fail to collect over conjunction. This leads him to a rational belief in contradictions even if he does not engage in reflections about his fallibility. However, if he does engage in those reflections, he acquires another basis for believing that he believes an inconsistency. The path of modesty turns on the self-fulfilling nature of the meta-belief. Believing so, makes it so.

The modest meta-belief bears a risk of liar-paradoxicality that is inversely proportional to the size of the belief system. If there are few beliefs, then the meta-belief has a significant chance of causing me to become inconsistent. But in the global variation, it is far more likely that the belief that I believe an inconsistency is a consistent report of the bad news that I believe an inconsistency. The meta-belief in the high-strength version covers a much larger number of beliefs. There is safety in numbers. The more comprehensive the meta-belief, the more likely that some other belief is the inconsistency. And in that case, the reductio scenario in which the meta-belief forces me into an inconsistency is logically impossible.

Just how many contradictions do I believe? Anti-agglomerativity is a general pattern of belief, so I should expect to believe many contradictions in the same manner as the logic student.

Belief holism ensures that anyone who believes one inconsistency will believe many inconsistencies. As illustrated by the impact of Duhem’s thesis on the verification principle, beliefs operate within collectives. I cannot have exactly one false belief. Any error infects some ancillary beliefs. Since the ancillary beliefs of a logical belief are themselves logical, holism makes any logical imperfection a corporate setback.

Indeed, any logical error precipitates infinitely many logical errors. Those who believe an inconsistency, X, at least tacitly believe the conjunction of X and T where T is a perceived tautology. For the conjunction has the same probability as X. (Agglomeration only fails when the additional conjunct lowers the probability.)
Everyone believes “\(n = n\)” is a tautology, for each \(n\). Hence everyone believes infinitely many tautologies. Obviously, these beliefs cannot all be occurrent. They are dispositions. A finite object can have infinitely many relational dispositions when there are infinitely many relata. I have the ability to lift a 1 kilogram weight, a 1.1 kilogram weight, a 1.11 kilogram weight, and so on. Similarly, I am disposed to affirm \(1 = 1, 2 = 2\), and so on. Therefore, I have infinitely many inconsistent beliefs of the form “\(X + n = n\).”

Logic equips us with many ways of showing how error multiplies. Instead of conjoining \(X\) with other propositions, one could simply consider successive double negations. If \(X\) is inconsistent, so is \(\sim \sim X, \sim \sim \sim X, \text{etc.} \) Modal error can also issue from non-logical error. Correctly informed that \(Y \sim Z\), the believer in \(Z\) will infer \(Y\). When \(Y\) turns out to be an inconsistency, a new inconsistency has come home to roost. (Notice this pattern simply trades on mistaken belief in \(Z\)—it works even if \(Z\) is meaningless.)

Incompleteness has the same infectious character. Godel’s incompleteness theorem does not merely entail that a consistent, arithmetically competent system must have at least one unprovable truth. It entails that such a system must have infinitely many. For instance, the conjunction of an unprovable truth with a theorem is an unprovable truth. Since there are infinitely many theorems of a system strong enough to express arithmetic, there are infinitely many unprovable truths.

A corollary of Gödel’s theorem is that there are as many unprovable truths as provable truths. If there is any incompleteness, there is infinite incompleteness. Therefore, it is impossible to minimize the incompleteness in the sense of reducing the total number of unproved truths.

Similarly, if there is some inconsistency, there is infinite inconsistency. Since reason demands belief in at least one contradiction, I cannot minimize inconsistencies in the sense of reducing their total number. The only numeric goal is to minimize the frequency of inconsistencies in small subsets of my beliefs. If every hundredth belief of mine is a contradiction, then I would be increasing my rate of consistency by changing to a belief system in which only every thousandth belief was a contradiction. One sign of this higher consistency is a reduced vulnerability to deductive refutation. A belief can only be deductively refuted by exploiting inconsistencies. A deductive argument must exploit inconsistencies that can arise from finitely many premises. Human limits on working memory and attention radically reduce the size of psychologically effective premise sets to about 7 chunks of information (Miller 1956). Therefore, if my inconsistencies are spread thinly, others will rarely be able to show that my belief in the premises conflicts with my lack of belief in the conclusion. I can become more and more logical in the sense of having a lower probability of being actually refuted by someone. But all along, I will believe infinitely many contradictions—just as reason demands.

New York University
NOTES

This paper was improved by audience reactions at Stockholm University and at the 1997 meeting of the Society for Exact Philosophy in Montreal. I also thank Jack Copeland, Frederick Kroon, Kirk Ludwig, Graham Oddie, Graham Priest, and Greg Ray.

BIBLIOGRAPHY


