Paradox of Deniability

Massimiliano Carrara
FISPPA Department, University of Padua, Italy

Peking University, Beijing - 6 November 2018

Introduction. The starting elements

Suppose two speakers disagree if:

Padua is North of Venice.

Disagreement. Two speakers S_1 and S_2 who respectively utter u_1 and u_2 disagree only if they have *incompatible* beliefs, or perform speech acts that cannot be jointly correct.

A standard way to express that one has an *incompatible* belief with another is to *negate* the other's *assertion*.

But, if a sentence A is a dialetheia (or a *glut*), it is both true and false: dialetheists or glut theorists won't in general take assertions of both A and $\neg A$ to express *disagreement*.

Question: *How to express disagreement in a dialetheic framework*?

Answer: Disagreement may still be expressed by *denying* what was said (Priest 2006a, Priest 2006b, Beall 2009).

The Thesis: Assertion and rejection are *incompatible* speech acts. Unlike paraconsistent negation, denial is for dialetheists *exclusive*: A and $\neg A$ may both be true, but *you can't correctly assert and deny A*.

If so, glut theorists must reject, and reject, the right-to-left direction of the classical theory of denial, that to deny A is equivalent to asserting $\neg A$:

> *Classical denial*. A is correctly denied iff $\neg A$ is correctly asserted.

The paraconsistent denial of A is *stronger* than the assertion of $\neg A$.

The result: The exclusivity of negation is lost (or perhaps was never had) in the realm of *logic*, but it is regained at the *pragmatic level* – with denial. The above characterization in terms of incompatibility seems to go in the direction before expressed by *disagreement.*

Question: If giving up the exclusivity of negation is the key to solving the semantic paradoxes, doesn't the exclusivity of denial land us back in paradox?

Aim of the talk

- I argue, first in an informal way, then formally, that intuitive norms for denial (which dialetheists arguably accept) give rise to paradox – *the rejectability paradox*.
- Then, I consider some possible ways out or replies, concerning first the informal way of characterizing the paradox, then the formal language adopted.
- I conclude with a *dilemma*: either denial can serve as means to express disagreement, but the notion of exclusive deniability is not expressible in the glut theorist's language, or deniability is expressible, but denial no longer is a means to express disagreement.

The deniability or rejectability paradox. An informal skecth

As Littman and Simmons (2004) observe since the dialetheist appeals to 'non-standard relations between assertability and deniability', we are owed a full account of these notions. in particular, any such account would need to deal with apparent paradoxes that turn on the notion of assertability or deniability.

The assertability paradox (Littman and Simmon 2004, 320). Take a sentence α as having the form:

(α) α is not assertable.

They argue that (α) is a dialetheia.

Proof.

- Suppose (α) is *true*. Then what it says is the case. So (α) (i.e. (α) α is not assertable) is not assertable. But we have just asserted (α). So (α) is assertable–and we have a contradiction.
- Suppose, on the other hand, that (α) is *false*. Then what (α) says is not the case. So (α) is assertable. So we may assert: (α) is not assertable. Again, we have a contradiction.

A problem with the assertability paradox: The mere supposition that α is *true does not imply its assertability.* Indeed, assertability implies the recognition, not just the mere supposition, of the truth of (α).

The amended assertability paradox.

Proof. Let us prove dialetheically that α is true by distinguishing the following two cases:

- (1) Assume that (α) is false. Then, its negation is true, so (α) is assertable and then it is true.
- (2) Assume that (α) is true. Then it is true.
- According to the Law of the Excluded Middle (LEM), (α) is true. In this way we have a proof–not just a supposition–of the truth of (α).
- And we can assert it. So (α) is assertable, in opposition to what (α) claims (it claims of being not assertable), hence it is false. Therefore it is a dialetheia.

Question: If α is a dialetheia, then it both is and is not assertable. But how is it possible to both assert and not assert a sentence?

Reply. Once the exclusivity of logical negation has been rejected, the non-assertability of a sentence does not exclude its assertability.

The rejectability (or deniability) paradox. Let R be a sentence having the form:

(R) the sentence R is rejectable.

You can both acccept and reject (R).

Proof.

- Assuming that (R) *is true*, then it is rejectable. So, there is a state of knowledge in which one can *reject* it. In such a state, one recognizes that what (R) says is true, so that one is in a position to *assert* (R). So you *both reject and accept* R. Thus, the assumption of (R) leads to a state (of knowledge) in which *one can both assert and reject (R), and that is dialetheically inacceptable.*
- It follows that (R) *cannot be true*. But then we can *reject* it, recognize its truth and *assert* it, which, again, is in opposition to Priest's thesis of *the impossibility of accepting and rejecting the same sentence*. Again: *dialetheically inacceptable*.

Notice that, differently from the Assertability paradox, the Deniability paradox goes against dialetheist thesis that assertion and rejection are *incompatible* speech acts.

The irrationalist paradox (Priest (2006a, 111–112 and 2010 121, my revision). Let I be a sentence having the form:

(I): it is not rational to accept I

Again you can both accept and and non-accept (I).

Proof.

- Suppose that one accepts it; then one accepts I and *it is not rational to accept I*. This, presumably, is irrational. Hence it is not rational to accept I. That is, we have just proved I. So it is rational to accept it. Let Rat be an operator expressing *rational acceptance* and $R = \neg\mathsf{Rat}(R)$, Priest derives R from the schema

  $$(P)\ \neg\mathsf{Rat}(A \wedge \neg\mathsf{Rat}(A)),$$

  as follows:

  $$\cfrac{\cfrac{\cfrac{\neg\mathsf{Rat}(R \wedge \neg\mathsf{Rat}(R))}{\neg\mathsf{Rat}(R \wedge R)}}{\neg\mathsf{Rat}(R)}}{R}\ .$$

  That is, R, and hence $\neg\mathsf{Rat}(R)$, is deducible from P: $P \vdash R$.

- Assuming that rational acceptance (Rat) is closed under single-premise deducibility, and that P is rationally acceptable (*and it seems to be: if someone believes A, and, at the same time, believes that it is not rationally permissible to believe A, that would seem to be pretty irrational – not something that is itself rationally permissible*), we have that $\mathsf{Rat}(P) \vdash \mathsf{Rat}(R)$, and we have a contradiction.

Reply. Priest calls such kind of paradox a *rational dilemma*. He observes that a dialetheist cannot rule out a priori the occurrence of rational dilemmas:

> Arguably, the existence of dilemmas is simply a fact of life (Priest 2006b, 111).

Moreover, he maintains that the irrationalist's paradox is much more problematic for a classicist than for a dialetheist. For the latter it is not irrational to believe both a sentence α and that it is irrational to believe α, if such a belief is also rational, an option clearly closed to a fan of classical logic. This argument is in line with the comment to the *assertability paradox*: if negation is non-exclusive, a dialetheist can rightly assert a non-assertable sentence, if she has recognized that it is also true.

Replies to the reply. The irrationalist paradox with rejection

First. You can produce a paradox with assertion and rejection.

- (I): it is not rational to believe I

I is assertable and rejectale

Proof.

- Suppose that a rational human being believes I. Then she should accept to have an irrational belief, against her rationality. So she can rightly *reject* it.
- But, then, she recognize that it is irrational to believe it, namely that I is true. Therefore, we are right to *assert* it.

Both the *irrationalist's paradox* and the *rejectability paradox*, lead to the conclusion that it is rational both to *assert* and *reject* it, against Priest's thesis that *assertion* and *rejection* are incompatible speech acts.

Second. About *rational dilemma*: Consider a classicist on the strengthened liar. He could observe that the fact that there are rational grounds both for the truth and for the untruth of the strengthened liar sentence is nothing but a rational dilemma. It is simply a fact of life. There is nothing to add.

Problem: let us seriously consider the dialetheist analysis of assertion and denial. Is possible to provide a logic of exclusive denial for glut-theorist?

A logic for the Dialetheist assertion and denial

The logic is Priest's Logic of Paradox LP, which may or may not be augmented with a suitable conditional $\rightarrow$ satisfying $A \rightarrow A$ and *modus ponens*.

LP. Semantically, LP is a *three-valued* logic whose language is that of CPL and admissible valuations are all the total maps from the set of well-formed formulae WFF to the set $\{1, 0.5, 0\}$ satisfying:

($\neg$) $v(\neg A) = 1 - v(A)$

($\vee$) $v(A \vee B) = \max\{v(A), v(B)\}$

($\wedge$) $v(A \wedge B) = \min\{v(A), v(B)\}$

The designated values are $\{1, 0.5\}$: an argument is LP-valid if it never gets you from either 1 or 0.5 to 0. Notable invalidities include the principle of *Ex Contraditione Quodlibet*

$$(\mathsf{ECQ})\ A, \neg A \vdash B.$$

Assertion and *denial* are external manifestations of, respectively, the mental states of acceptance and rejection can be represented by means of *signed formulae*: $+A$ for assertion and $-A$ for denial, where $+$ and $-$ are nonembeddable *force signs* (Rumfitt 2000).

Classically, one defines a set of *correctness-valuations* C for signed formulae such that every member is induced by the set of admissible truth-valuations of Classical Logic by the following correctness clauses:

(C1) $v_c(+A) = \mathbf{1}$ iff $v(A) = 1$;

(C2) $v_c(-A) = \mathbf{1}$ iff $v(A) = 0$.

One may correctly assert (deny) A just in case A is true (false). Validity for signed formulae is defined in the obvious way: an argument is valid iff it preserves value **1**, i.e. correctness. This won't work in a dialetheist framework, however.

The natural dialetheist counterparts of *C1* and *C2* are:

(C1*) $v_c(+A) = \mathbf{1}$ iff $v(A) \subseteq \{1, 0.5\}$;

(C2*) $v_c(-A) = \mathbf{1}$ iff $v(A) \subseteq \{0, 0.5\}$

In *short*: you may assert A iff A is *true* (and possibly false, too), and you may deny A iff A is *false* (even if it turns out to be also true). But, if one could correctly deny what may be correctly asserted – a straightforward consequence of C1*, C2* and the existence of gluts – *the denial* of A would no longer be guaranteed to express disagreement.

If denial/rejection is a means of expressing disagreement, denial must be *exclusive*: it must be "impossible jointly to accept and reject the same thing' (Priest 2006a, p. 103).

But, then, C2 must *not* be revised (with C2*): the exclusive denial of A is correct iff A is (classically) *false only*, i.e. false but not true.

We thus get a *logic of exclusive denial*, call it LP*, whose language is the language of LP supplemented with signs expressing assertion ($+$) and denial ($-$).

The semantics is given by a set of admissible correctness valuations V^*_{LP}, each member of which is induced by the admissible valuations of LP via C1* and C2.

LP* validates many of the classical rules given in (Smiley 1996) and (Rumfitt 2000), but not all. For instance, the following two classically valid rules for negation

$$+\text{-}\neg\text{-E}\ \frac{\Gamma \vdash +(\neg A)}{\Gamma \vdash -A} \qquad -\text{-}\neg\text{-I}\ \frac{\Gamma \vdash +(A)}{\Gamma \vdash -(\neg A)}$$

are LP*-invalid.

The semantics of LP* validates the following, highly intuitive, coordination principles:

$$\text{Coord}_1\ \frac{\Gamma, +A \vdash}{\Gamma \vdash -A} \qquad \text{Coord}_2\ \frac{\Gamma \vdash +A \qquad \Delta \vdash -A}{\Gamma, \Delta \vdash}\ .$$

About:

$$\text{Coord}_1\ \frac{\Gamma, +A \vdash}{\Gamma \vdash -A}$$

> An argument against an opponent who holds A to be true is rationally effective if it can be demonstrated that A entails something that ought rationally to be rejected B. For, it then follows that they ought to reject A. (Priest 2006, p. 86)

About:

$$\text{Coord}_2\ \frac{\Gamma \vdash +A \qquad \Delta \vdash -A}{\Gamma, \Delta \vdash}\ .$$

> it is impossible jointly to accept and reject the same thing . . . acceptance and rejection are mutually incompatible. (Priest 2006a, p. 103)

Deniability and LP*

Suppose the language is rich enough to express *deniability*. That English contains some such predicate seems beyond doubt, as the following examples show:

(*) The judge is confident that everything Marc said is deniable.

(**) If what I say is deniable, why is nobody objecting?

Let T be a dialetheic extension of PA, with underlying logic LP*, let us add to T's language a fresh predicate $\mathcal{D}(x)$ expressing *correct deniability*; such that:

$\mathcal{D}(\ulcorner A \urcorner)$ is true iff A is correctly deniable. Then, $\mathcal{D}(x)$ will at least satisfy the following:

(D) $v(\mathcal{D}(\ulcorner A \urcorner)) = 1$ iff $v_c(-A) = \mathbf{1}$.

$$\mathcal{D}\text{-I}\ \frac{\Gamma, +A \vdash}{\Gamma \vdash +\mathcal{D}(\ulcorner A \urcorner)} \qquad \mathcal{D}\text{-E}\ \frac{\Gamma \vdash +A \qquad \Delta \vdash +\mathcal{D}(\ulcorner A \urcorner)}{\Gamma, \Delta \vdash}\ .$$

Let D be a sentence which say of itself (only) that is deniable or rejectable, our (R)

Then, D satisfies:

$$\frac{+D}{+\mathcal{D}(\ulcorner D \urcorner)} \qquad \frac{+\mathcal{D}(\ulcorner D \urcorner)}{+D}$$

One may then reason thus:

$$\mathcal{D}\text{-I}\ \dfrac{\text{Contraction}\ \dfrac{\text{Coord}_2\ \dfrac{\text{(D), C2}\ \dfrac{\dfrac{+D \vdash +D}{+D \vdash +\mathcal{D}(\ulcorner D \urcorner)}}{+D \vdash -D} \qquad +D \vdash +D}{+D, +D \vdash}}{+D \vdash}}{\vdash +\mathcal{D}(\ulcorner D \urcorner)}$$

Result: D is both assertible and deniable.

Assuming that self-reference isn't the culprit, glut theorists who accept the standard structural rules are left but with two uncomfortable options:

- either deny that denial is exclusive,
- or deny that exclusive denial is expressible.

A first reply. Norms of Denial

The norm for denial codified by (C2) asserts that A is correctly deniably iff A is *false only*.

Priest considers two norms of denial, neither of which makes denial exclusive. The first one is the following:

[**Deny(U)**] You may deny A if there is good evidence for A's untruth

A plausible corresponding norm for assertion:

[**Assert(T)**] You may assert A if there is good evidence for A's truth.

But the existence of sentences, such as the Liar sentence, that can be proved to be both true and untrue, **Deny(U)** licenses us to accept and reject the same sentence. And if it licenses us to accept and reject the same sentence it would no longer be guaranteed to express disagreement.

Priest also considers a second norm:

[**Deny(U)***] You may deny A if there is good evidence for A's untruth, *unless there is also good evidence for its truth.*

Problem: This second norm makes denial profoundly *unlike* assertion. Unlike assertion, any denial may later turn out to be incorrect, since any false sentence can in principle be discovered to be a glut.

- Thus, you can disagree with my assertion that $0 \neq 0$, and thus *deny* $0 \neq 0$. But, even if you can prove $0 = 0$, and hence disprove $0 \neq 0$, you can never be fully confident that your denial is correct: a proof of $0 \neq 0$ may always turn up.
- By contrast, if you have proved $0 = 0$ and thereby *assert* it, you can be fully confident that your assertion is correct.

Second Objections and replies

No paradoxes of denial. Following Parsons (1984) and Priest (2006a, b), it may be *objected* that there are no paradoxes of *denial*:

> attempts to formulate distinctive Liar paradoxes in the form of denial fail, since [−] being a force operator, has no interaction with the content of what is uttered (Priest 2006, p. 108).

I (of course) *agree*. The deniability *predicate* is, precisely, *not* a force operator.

Dropping $\mathcal{D}$-E?

$\mathcal{D}$-E must be invalid, on the grounds that there's valuations $v \in V_{LP}$ such that $v(A) = v(\mathcal{D}(\ulcorner A \urcorner)) = 1$. The assertion of both A and $\mathcal{D}(\ulcorner A \urcorner)$ would then be correct, and the rule would not preserve correctness.

The problem with this is that the valuations in question are ruled out by the principle that we may not correctly deny what's true. More precisely, that no such valuation is admissible is a consequence of C2, that the denial of A is correct iff A is false, together with (D), the principle that $v(\mathcal{D}(\ulcorner A \urcorner)) = 1$ only if the denial of A is correct.

Dropping $\mathcal{D}$-I?

- It may be objected that $\mathcal{D}$-I isn't valid in LP*.
- However, if knowing that A entails triviality isn't a *good enough* ground for asserting that A is deniable, one wonders whether there can *ever* be such grounds.

Conclusions: A dilemma

Denial plays a key role in the standard dialetheist account of disagreement: In order for this to work, denial must be exclusive: one may not correctly assert and deny the same proposition.

Glut theorists are faced with a *dilemma*:

- either denial can serve as means to express disagreement, but the notion of exclusive deniability isn't expressible in the glut theorist's language,
- or deniability is expressible, but denial may no longer serve as a means to express disagreement.

References

Beall, J.: 2009, *Spandrels of Truth*, Oxford University Press, Oxford.
Humberstone, L.: 2000, The revival of rejective negation, *Journal of Philosophical Logic* 29, 331–81.
Parsons, T.: 1984, Assertion, denial and the liar paradox, *Journal of Philosophical Logic* 13, 136–52.
Priest, G.: 1979, The logic of paradox, *Journal of Philosophical Logic* 8, 219–241.
Priest, G.: 2006a, *Doubt Truth to be a Liar*, Oxford University Press, Oxford.
Priest, G.: 2006b, *In Contradiction*, Oxford University Press, Oxford.
Priest, G.: 2010, *Hopes Fades For Saving Truth*, Philosophy 85, 109–140.
Rumfitt, I.: 2000, *Yes and No*, Mind 109, 781–824.
Simmons, G. Litmman, K. 2004. A Critique of Dialetheism, in *The Law of Non- Contradiction*, G. Priest, Jc Beall, and B. Armour-Garb, eds., Oxford: Oxford University Press, 314?335.
Smiley, T.: 1996, *Rejection*, Analysis 56(1), 1–9.