

## 1. Self-Knowledge and Self-Blindness

Two Moorean Principles:

**Assertion Principle:** It is irrational to assert  $\langle p, \text{ but I don't believe that } p \rangle$ .

**Belief Principle:** It is irrational to believe  $\langle p, \text{ but I don't believe that } p \rangle$ .

Plausible corollary of the Belief Principle:

**Shoemaker's Thesis:** Self-blind agents are impossible.

(A self-blind agent is an agent with ordinary rationality, intelligence, and conceptual sophistication, but who lacks introspective access to her own belief.)

In defense of *empiricist* theories of self-knowledge, I will argue that the Assertion Principle fails to support the Belief Principle, and thus fails to support Shoemaker's Thesis.

## 2. First Route from Assertion to Belief: Inference to the Best Explanation

Accounts of the Assertion Principle:

**epistemic account:** an account that presupposes that an agent cannot rationally believe propositions of the form  $\langle p, \text{ but I don't believe that } p \rangle$

Counterexample 1: Self-blind Sadie's non-introspective evidence supports that it will rain, but that she does not believe that it will rain. Sadie's philosopher friend is looking for examples of agents who are ignorant of meteorological facts, and he asks Sadie for help. Sadie has the aims of a normal speaker, and she knows that she does. Even so, Sadie refrains from asserting 'It will rain, but I don't believe that it will rain.'

**pragmatic account:** an account that appeals to aims associated with assertion/communication

Counterexample 2: Self-blind Sadie's third-person evidence supports that it will rain, but that she does not believe that it will rain. Sadie is taking a true/false exam on which wrong answers incur a penalty. The first statement on the exam is 'It will rain, but I, Sadie, don't believe that it will rain.' Sadie aims to earn as high a score on the exam as possible, and she knows that she has this aim. Even so, Sadie refrains from marking this statement is true.

**alethic account:** an account that appeals to Sadie's aims to speak the truth, the whole truth, and nothing but the truth

Must alethic accounts be epistemic? They must only under the mistaken assumption:

**Bridge Principle:** If a rational agent believes a proposition, then she will mark it as true if her only aims are alethic.

## 2. Second Route from Assertion to Belief: Shoemaker's "Zany Argument"

Shoemaker's "zany argument":

(Moorean Premise) A self-blind agent would act like an agent who is not self-blind.

(Behaviorist Premise) If an agent acts like one who is not self-blind, then that agent is not self-blind.

Therefore, (Shoemaker's Thesis) Self-blind agents are impossible.

But to be plausible, the premises must equivocate on ‘acting like an agent who is not self-blind’. If it means acting like an agent who is not self-blind *in all circumstances*, then the Moorean Premise is implausible.

**Multiple Choice Exam:** Self-knowing Steve is taking a multiple choice exam. Steve’s evidence supports that it will rain, and he believes that it will rain. But Steve’s non-introspective evidence also supports even more strongly that he does not believe that it will rain. He comes to a multiple choice problem asking him to circle one of the following:

- (a) It will rain, but I don’t believe that it will rain.
- (b) It will rain, and I believe that it will rain.
- (c) It won’t rain, but I don’t believe that it won’t rain.
- (d) It won’t rain, and I believe that it won’t rain.

Steve not only avoids (a) and (c), but moreover selects (b) rather than (d).

If it means acting like an agent who is not self-blind *in cases like Counterexamples 1 and 2*, then the Behaviorist Premise is implausible. Consider:

**Egan’s Psychopath Button:** You find a ‘kill all psychopaths’ button. You would prefer to live in a world without psychopaths to one with psychopaths, but you’d rather live in a world with psychopaths than die. You are in fact a psychopath, but your third-person evidence supports that you are not a psychopath. You know that only a psychopath would press the button.

Upshot: You can be self-blind about character traits yet still act like an agent who has introspective access to his own character traits. Why not belief?

#### 4. Third Route from Assertion to Belief: Inner Assertion as a Kind of Belief

Generalization argument from public assertion to belief:

- (1) If it is irrational to publicly assert  $\langle p, \text{ but I don't believe that } p \rangle$ , then it is irrational to privately affirm  $\langle p, \text{ but I don't believe that } p \rangle$ .
- (2) If it is irrational to privately affirm  $\langle p, \text{ but I don't believe that } p \rangle$ , then it is irrational to believe  $\langle p, \text{ but I don't believe that } p \rangle$ .
- (3) If it is irrational to believe  $\langle p, \text{ but I don't believe that } p \rangle$ , then it is irrational to jointly believe both that  $p$  and that one does not believe that  $p$ .
- (4) Therefore, if it is irrational to publicly assert  $\langle p, \text{ but I don't believe that } p \rangle$ , then it is irrational to jointly believe both that  $p$  and that one does not believe that  $p$ .

Re (1): We should accept (1) if we reject pragmatic accounts in favor of alethic accounts of Moore’s paradox.

Re (2): We can reject (2) if we one of the following:

- (i) judgment = private affirmation
- (ii) judgment = ‘conscious belief’

Re (3): While there are other problems with *general* multi-premise closure principles (e.g., preface and lottery cases), anyone who accepts (2) must also grant more germane counterexamples to (3).