

A Linguistic Reason for Truthfulness

Michael Rescorla

§1. Truth and assertion

Morality, propriety, and legality can impel us to speak truthfully. But such factors seem incidental to assertion's fundamental nature. Many philosophers try to isolate a more specifically linguistic reason for speaking truthfully, a reason that does not depend upon extra-linguistic concerns like ethics, etiquette, or the law. According to these philosophers, the essential nature of language or assertion generates a rational bias towards truthfulness. Dr. Johnson memorably expressed this perspective when he suggested that even the devils in hell would speak truthfully to one another.¹

One popular implementation of these ideas builds truthfulness directly into the nature of assertion. This approach admits a variety of formulations: assertion *aims* for the truth; truth is a *goal* of assertion; truth is a *norm* of assertion; truthfulness is *intrinsic* to assertion; and so on. Proponents include William Alston (2000), Michael Dummett (1978, 1981), John Searle (1969, 2001), Timothy Williamson (2000: 238-269), Crispin Wright (1992), and many other philosophers. If such an approach is correct, then truthfulness is somehow constitutive of assertion.²

I will argue against this approach while simultaneously attempting to ground truthfulness in assertion's essential features. On my approach, truthfulness is not constitutive of assertion, but it emerges from more explanatorily fundamental constitutive features of assertion. Thus, we must distinguish between two theses. The *Default Thesis*

holds that constitutive features of assertion render truthfulness the rational default policy. The *Constitutive Thesis* holds that truthfulness is itself constitutive of assertion. Although the two theses are often combined, I endorse the Default Thesis while rejecting the Constitutive Thesis.³

Following Wilfrid Sellars (1963) and Robert Brandom (1994), I hold that assertion is individuated by its role in *the game of giving and asking for reasons*, or *reasoned discourse*. Assertion essentially involves a commitment to defend what one asserts against challenges and counter-arguments. The constitutive elements of reasoned discourse articulate this dialectical commitment. On my view, the constitutive elements of reasoned discourse do *not* mandate truthfulness. A sufficiently skillful liar could engage impeccably in reasoned discourse while offering an elaborate series of lies. Nevertheless, I will argue that truthfulness is the rational default strategy for executing reasoned discourse in accord with its constitutive elements. Roughly, I will argue that truthfulness is the rational default strategy for asserting claims that one need not subsequently retract. Although truthfulness is not constitutive of assertion, it is the *prima facie* best way to achieve a goal that *is* constitutive of assertion: avoiding decisive counter-arguments against what one says.

§2. Constitutive norms and constitutive goals

Philosophers who espouse the Constitutive Thesis almost invariably analogize assertion with a game. Timothy Williamson writes that “someone who knowingly asserts a falsehood has thereby broken a rule of assertion, much as if he had broken a rule of a game; he has cheated” (2000: 489). Williamson explicitly aims “to identify the

constitutive rule(s) of assertion, conceived by analogy with the rules of a game” (2000: 489). On this picture, assertion, just like a game, is a practice partially constituted by rules that dictate how to engage in it correctly. Someone who breaks the rules may still engage in the practice, but she does not engage in it correctly. This picture also informs the writings of William Alston and John Searle. If one adopts the picture, one must say precisely which norms constitute the practice of assertion. Plausible candidates include:

The Honesty Norm: One should assert only propositions that one believes.

The Warrant Norm: One should assert only propositions that one believes with “sufficient” warrant.

The Truth Norm: One should assert only true propositions.

The Knowledge Norm: One should assert only propositions that one knows.

Each of these suggestions entails that someone who knowingly asserts falsehoods thereby violates the rules of assertion.

Another version of the Constitutive Thesis, associated primarily with Michael Dummett and developed more recently by Michael Glanzberg (2004), deploys the analogy with a game to somewhat different effect, as in the following passage from Dummett: “It is part of the concept of winning a game that a player plays to win... Likewise, it is part of the concept of truth that we aim at making true statements” (1978: 2). Imagine someone who learns the rules of chess without realizing that the point of the game is to checkmate one’s opponent. This person has not yet mastered what it is to play chess. In general, the injunction that one try to win when playing a game seems constitutive of competitive game-playing. But this injunction occupies an intuitively different status than, say, the injunction that one should not examine other players’ cards

while playing poker. Alston, Searle, and Williamson assimilate truthfulness to the latter kind of injunction, whereas Dummett and Glanzberg assimilate it to the former.

How significant is the distinction between these two kinds of injunction, and hence between our two versions of the Constitutive Thesis? Searle writes that “it is a matter of rule of competitive games that each side is committed to trying to win. Notice in this connection that our attitude to the team or player who deliberately throws the game is the same as that towards the team or player who cheats. Both violate rules, although the rules are of quite different sorts” (1969: 34). As the final sentence of this quotation demonstrates, Searle acknowledges a theoretically significant difference between the injunction that one try to win and the injunction that one not examine other players’ cards. Yet the passage as a whole minimizes important differences. If a chess master plays chess with an amateur, then the master might play below full capacity to ensure a more balanced game. She might even prolong the game by foregoing some devastating but obscure gambit. Although her conduct seems somehow non-paradigmatic, it does not strike us as *incorrect*, at least not in the same way it would be incorrect to rearrange the chess board while diverting her opponent’s attention. As Joseph Raz puts it, “[t]here are no mandatory norms which require players to win; and if they fail to win, or if they lose, they do not thereby break any rule. They simply fail to achieve their goal” (1990: 117).

To capture such distinctions, Raz distinguishes between a practice’s *rules* and its *values*. In Raz’s terminology, winning is a value of competitive games, but it is not a rule. Employing what I find somewhat more suggestive terminology, I propose that we distinguish between a practice’s *constitutive norms* and its *constitutive goals*. Constitutive

norms dictate how to engage in the practice correctly. Constitutive goals are ends that agents pursue, *qua* participants in the practice. To engage in a practice, one must pay lip service to its constitutive goals. A tennis player who blatantly loses, visibly making no effort to win, ceases to play tennis. Thus, a practice's constitutive goals are in some sense normative for the practice. But they are not constitutive norms in my technical sense, for one can abandon them while executing the practice correctly.⁴

Several marks distinguish constitutive goals from constitutive norms. (1) As Raz observes, constitutive goals typically guide action in more complicated ways than constitutive norms (1990: 118). For instance, the injunction that one try to win at chess guides players in selecting among various *strategies*. The choice among these strategies is usually very complex. (2) The constitutive norms of a well-designed practice possess a certain coherent character. They do not conflict with one another. Constitutive goals need not cohere. Raz notes that competitive games possess at least two constitutive goals: *winning* and *avoiding defeat* (1990: 118). Although the first goal entails the second, the two goals may generate rational conflicts, since the strategy most likely to win is not necessarily the strategy most likely to avoid defeat. (3) Participants in a practice typically pursue its constitutive goals only *to some extent*. One can make some effort towards winning at chess without exhaustively surveying all possible strategies. Indeed, there would be something pathological about investing too much energy in surveying alternative strategies, since doing so could prolong the game indefinitely. In contrast, correct participation in a practice requires strict adherence to its constitutive norms.

We must sharply distinguish between a practice's constitutive elements and the goals that a given participant harbors while engaging in the practice. The latter may

provide reason to pursue the former, as when a desire to humiliate one's opponent impels one to win a game. They may also militate in the opposing direction, as when a father avoids winning at tennis to bolster his son's self-confidence or when desperation for money impels a gambler to cheat. One can rationally engage in a practice without obeying its constitutive norms or pursuing its constitutive goals. Nevertheless, a practice's constitutive norms and goals plausibly occupy *some* privileged role within the practical reasoning of the practice's participants. I will return to this point in §4 and in §7.

The distinction between constitutive norms and constitutive goals is rather murky. How we should define these two notions, and how we should construe the differences between them, deserves much more extended inquiry. Still, it seems likely that activities such as games, dances, musical performances, and religious ceremonies require a framework along the lines sketched above. Working within this framework, we may say that Alston, Searle, and Williamson treat truthfulness more like a constitutive norm of assertion, while Dummett and Glanzberg treat it more like a constitutive goal.

It is hardly obvious that assertion involves either constitutive norms or goals. Many activities do not. Like any action, assertion falls under general norms of morality, legality, practical reason, etiquette, and so on. But this does not show that assertion involves *proprietary* norms that constitute what it is to make an assertion, in the same way that the rules of a tennis constitute what it is to play tennis. As Davidson observes, "in explaining what it is to eat no mention of rules or conventions needs to be made" (1984: 264). So why suspect that, in explaining what it is to assert a proposition, we must adduce constitutive norms or goals? Davidson urges that we need not: "what constitutes the making of an assertion is not governed by agreed rules or conventions," and,

furthermore, “nothing in language corresponds in relevant ways to winning a game” (1984: 267-8). A thorough discussion would engage Davidson’s worries in considerable depth. In this paper, however, my primary objective is to compare various theories of assertion’s constitutive norms and goals, not to argue that assertion *has* constitutive norms or goals.⁵

§3. The dialectical model of assertion

I now want to introduce my own favored conception of assertion’s constitutive elements. I advocate *the dialectical model of assertion*, according to which assertion is individuated partially by its role within what Sellars and Brandom call “the game of giving and asking for reasons.” I will refer to this activity variously, and equivalently, as “reasoned discourse” and “rational dialectic.” Philosophical study of this activity stretches back to Aristotle, continuing through medieval discussion of the so-called Obligation Game, with notable recent contributions by Alvin Goldman (1994), C. L. Hamblin (1970), Paul Lorenzen (1987), Nicholas Rescher (1977), and Stephen Toulmin (1958). Many philosophers would agree that reasoned discourse occupies a central role in linguistic intercourse. The dialectical model furthermore claims that reasoned discourse helps delineate *what it is* to assert a proposition. Assertion is a commitment to defend what one asserts within reasoned discourse. Advocates of the dialectical model include Robert Brandom (1994), John MacFarlane (2003, 2005), Wilfrid Sellars (1963), Douglas Walton and Erik Krabbe (1995), and Gary Watson (2004).

The dialectical model encompasses many different theories. My preferred version centers around the following two constitutive norms of reasoned discourse:

The Defense Norm: When challenged to defend an asserted proposition, one must provide a non-circular, cogent argument for the proposition or else retract it.

The Retraction Norm: When faced with a counter-argument against an asserted proposition, one must rebut the counter-argument or else retract the proposition.

By asserting a proposition, I commit myself to defending the proposition against challenges and counter-arguments. By retracting the proposition, I cancel this commitment.

Even if one accepts something like the Defense and Retraction Norms, many parameters of variation remain. Most notably, one might replace the Defense Norm with

The Default-Challenge Norm: When faced with a *legitimate challenge* to defend an asserted proposition, one must either provide a non-circular argument for the proposition or else retract it.

For instance, a challenge to the assertion “I have hands” might count as legitimate only if the challenger sketches a compelling skeptical scenario in which I do not have hands. Robert Brandom and John MacFarlane adopt a weakened view along these lines, as do Michael Williams (1999) and Adam Leite (2005). Although the weakened view seems fairly plausible, I believe that the Defense Norm is far more attractive than it initially appears. The view I develop below is neutral between the Defense Norm and the Defense-Challenge Norm.

Working independently, Lorenzen and Hamblin initiated a program of developing norms like the three just mentioned into formal mathematical models of rational dialectic. I will not engage with this program here, although I find it illuminating. I will continue to operate at a fairly informal level of description.

The Defense, Retraction, and Default-Challenge Norms constrain how one must react *after* another speaker challenges one's assertion. They do not constrain what one should assert in the first place. One can develop the dialectical model to include additional norms, such as the Honesty, Truth, Warrant, or Knowledge Norms. Thus, Brandom regards the Warrant Norm as constitutive of reasoned discourse: "making a claim one is not entitled to... is a kind of impropriety, a violation of a norm" (1994: 179). In contrast, I doubt that assertion falls under constitutive norms constraining which propositions one may assert. Morality, legality, propriety, or prudential rationality might direct one to assert only propositions belonging to some favored category. But no specifically linguistic norm supplements these extra-linguistic directives. A lie violates no constitutive assertoric norm, provided that one adequately justifies the lie when challenged.

Consider Iago. Throughout *Othello*, Iago repeatedly asserts sentences he believes false. Nevertheless, I think that he is an exemplary participant in reasoned discourse.⁶ When challenged to provide evidence for Desdemona's infidelity, he offers compelling arguments. He claims that he heard Cassio utter incriminating remarks while asleep; he claims that he saw Cassio with Desdemona's handkerchief; he claims that he saw Desdemona and Cassio flirt with one another. These arguments employ false premises, but they satisfy the Defense norm. Moreover, when evidence finally surfaces against Iago's lies, he retracts the claim that Desdemona was unfaithful, thereby obeying the Retraction norm.⁷ Iago behaves abominably. He violates numerous moral and social norms. Othello rightly reproaches him upon discovering his mendacity. Nevertheless, on my view, there is nothing specifically linguistic about Iago's transgressions. We should

criticize Iago for exploiting assertion to achieve nefarious ends, but we should not criticize him for flouting any intrinsic norm of assertion.

My view is less extreme than it initially sounds, because I additionally recognize *constitutive goals* that guide what one asserts. The first, and most fundamental, constitutive goal is what I will call *rapprochement*: isolating relevant, mutually acceptable premises. This constitutive goal is so fundamental because, if speakers abandon it, they cease to engage one another rationally. Two people who cannot agree on any relevant premises cannot fruitfully reason with each another. As an extreme example, suppose that I *stonewall*, refusing to accept any premises adduced by my interlocutor, no matter how obvious. Stonewalling violates no norm of reasoned discourse, but it flouts the intrinsic *purpose* of reasoned discourse: rational engagement. By stonewalling, I forego genuine rational engagement with my interlocutor.

I believe that reasoned discourse involves two additional constitutive goals. Suppose that speaker *A* and speaker *B* debate some proposition. If *A* provides a cogent argument employing premises that *B* accepts, and if *B* cannot provide additional considerations that override or undermine this argument, then I say that the argument is *decisive* against *B*. On this usage, decisiveness is relative to a speaker. When context makes clear the relevant speaker, I will describe an argument as decisive *simpliciter*. On my view, a constitutive goal of reasoned discourse is to avoid counter-arguments against what one asserts that are decisive against oneself. Another constitutive goal is to provide arguments in favor of what one asserts that are decisive against one's opponent. These two constitutive goals correspond roughly to the two principal constitutive goals underlying competitive games: that one *not lose* the game and that one *win* the game.

The intuitive idea here is that, in themselves, the Defense and Retraction Norms do not privilege vindicating one's assertions over retracting them. The Retraction Norm requires that a speaker retract claims in light of decisive counter-arguments. A speaker could scrupulously obey this norm while irresponsibly asserting anything he pleased, as long as he retracted his assertions when faced with counter-arguments. Such a speaker has not fully mastered the practice of assertion. Some kind of investment in what one says is intrinsic to assertion. As Dummett puts it, "an assertion is a kind of gamble that the speaker will not be proved wrong" (1993: 84).⁸ Thus, taken by themselves, the Defense and Retraction norms do not successfully capture our intuitive starting point: that assertion is a commitment to vindicating what one says. Serious conversation presupposes that people are serious about what they say. It occurs paradigmatically between speakers who articulate defensible positions and defend those positions with cogent arguments. We cannot reason with someone who dismisses all objections, yet neither can we reason with someone who collapses before all objections.

Hamblin's analysis of rational dialectic does not accord such intuitions sufficient weight. Hamblin proposes a formal model whose rules, like the Defense and Retraction Norms, allow speakers to retract assertions whenever they choose (1970: 267). On this basis, Walton and Krabbe critique Hamblin's account. As they emphasize, though, the challenge is to discourage promiscuous retraction without encouraging a dogmatic refusal to retract anything at all: "[i]s there some way of making commitment 'sticky' or 'binding' enough to prevent such an escape from commitment, while still allowing enough freedom to make sensible retractions possible?" (Walton and Krabbe 1995: 10). To meet this challenge, Walton and Krabbe introduce an intricate set of context-sensitive

norms governing when one may retract one's assertions. Their account hinges upon the notion of a speaker's *dark-side commitments*, which "are central to an arguer's underlying position, even if he (or the other party) is not aware of them as explicit commitments he has made" (1995: 11). A speaker cannot simply retract her dark-side commitments. For instance, Walton and Krabbe urge that a life-long communist engaged in a political argument cannot simply retract her commitment to Marxist doctrine.⁹

I find the notion of a "dark-side commitment" rather obscure. More importantly, I think that any proposal along these lines is both too narrow and too stringent. The proposal is too narrow because it does not cover propositions lying outside one's "dark-side," so that it generates no bias against capriciously asserting and then retracting such propositions. The proposal is too stringent because it castigates perfectly legitimate changes in assertoric commitments. A devout communist might reasonably retract her political views in light of economic, sociological, or philosophical arguments. The norms of reasoned discourse should not discourage such radical conversions. There is nothing wrong with abandoning a fundamental conversational commitment, as long as the abandonment occurs against a more general background of serious assertion. A good account should secure a *pervasive* bias against retraction without condemning *any given retraction* as illegitimate or deviant.

I propose that we supplement the Defense and Retraction Norms with the two constitutive goals mentioned above: avoiding decisive counter-arguments against one's assertions, and providing decisive arguments for those assertions.¹⁰ These goals are quite distinct from the practice's norms, which are indifferent between defending one's assertions and retracting them. The goals induce an asymmetry between defense and

retraction. But they do not mandate tenacious defense of all one's assertions, because they must be balanced against the more fundamental constitutive goal of rapprochement. In this way, the constitutive elements of reasoned discourse yield a genuine commitment to defending what one asserts, but they do not condemn any given retraction as illegitimate, nor do they counsel unwavering adherence to one's assertions.

The constitutive goals I have posited exhibit the three general marks highlighted in §2. (1) They guide action fairly complex ways. How do I avoid asserting propositions that will encounter decisive counter-arguments? Depending upon what I know about my interlocutor, I will choose different conversational strategies. Such complexities will be our main topic in §§5-6. (2) They may conflict. Providing decisive argument against my interlocutor requires asserting additional propositions. By asserting these propositions, I may increase my odds of encountering decisive counter-arguments. Similarly, the pursuit of rapprochement may lead me to accept various premises asserted by my interlocutor, thereby leaving me more vulnerable to decisive counter-arguments. (3) Normal participation in reasoned discourse requires only that I pursue these goals *to some degree*. For instance, it would be pathological to invest too much energy in avoiding decisive counter-arguments, since I may thereby prolong the conversation indefinitely or maneuver myself into inhabiting a blatantly false position.

To some extent, my emphasis upon these various constitutive goals diminishes the contrast between my version of the dialectical model and Brandom's. Several crucial differences remain. First, whereas Brandom posits a constitutive *norm* that restricts what one initially asserts, I posit only constitutive *goals* that guide what one initially asserts. This contrast will not impress philosophers, such as Searle, wary of the distinction

between constitutive norms and goals. But a second contrast does not depend upon that distinction. On Brandom's view, constitutive elements of reasoned discourse enjoin one to assert only propositions one is justified in believing. On my view, constitutive elements of reasoned discourse enjoin one to assert only propositions one can vindicate within rational dialectic. It is neither necessary nor sufficient that I be warranted in believing these propositions. If my interlocutors are sufficiently ignorant or stupid, then I might successfully defend a proposition I am not warranted in believing. If my interlocutors are sufficiently nimble or recalcitrant, then I may fail to vindicate a proposition I am warranted in believing.

A natural objection to my view is that, in many contexts, we hardly care about avoiding decisive counter-arguments against our assertions, let alone defending those assertions with compelling arguments. When we chat about the weather, or about what movie to see, I may feel little personal investment in defending what I say. I would just as soon retract as defend. I can also play "devil's advocate," endorsing positions I do not believe in the hope that my interlocutor will refute them. So it seems misguided to attribute constitutive status to vindicating what one asserts.

To keep this objection in proper perspective, we must recall the crucial distinction between a practice's constitutive elements and the goals that a given person harbors while engaging in the practice. I can also play chess without caring whether I win, or even when hoping to lose. That does not show that winning is incidental to chess's essential nature. On the contrary, someone who regarded winning as completely on a par with losing would simply not have grasped what it is to play chess. To play chess is to exploit a pre-existing practice biased towards winning rather than losing. I can instantiate the

practice without adopting its constitutive goals as my own. But then my activity enjoys a derivative, non-paradigmatic status. If the foregoing version of the dialectical model is correct, a similar analysis holds for assertion. Assertoric practice enshrines an inherent bias towards vindicating what one says. I can engage in the practice without adopting this constitutive goal as my own. But then my linguistic activity is parasitic upon paradigmatic cases where speakers do adopt that goal.

Another natural criticism of my account is that it omits crucial features of assertion, such as: people often use assertion to influence one another's beliefs; under certain circumstances, one acquires knowledge through other people's assertions; when I assert that p , I thereby represent myself as believing or knowing p ; assertion aims towards truthfulness. My view must either explain these phenomena or else explain them away as illusory. I believe that, in many cases, we can *derive* the relevant phenomena from the constitutive elements sketched above. Thus, there is no need to build the phenomena directly into assertion's essential nature. They emerge naturally from the dialectical considerations articulated by my account.

In what follows, I will illustrate this explanatory strategy with respect to truthfulness. At first blush, it might seem that my approach yields no rational bias towards speaking truthfully. As I have emphasized, a congenital liar might engage quite unimpeachably in reasoned discourse, so long as he defended his lies with sufficiently skillful arguments. Nevertheless, I will argue that my account generates *prima facie*, defeasible reason for aiming to assert true propositions. The basic idea is that truthfulness is the rational default strategy for achieving the constitutive goals of reasoned discourse. Crudely: the safest way to win an argument is to speak the truth in first place.

§4. Reasons and norms

Before developing these ideas, I address a preliminary issue: when one participates in a practice, why should one obey that practice's constitutive norms and pursue its constitutive goals? To what extent, and in what way, do the constitutive norms and goals of a practice provide reasons for action? An adequate answer to these questions would require a much more intensive investigation than I have undertaken into the concepts of constitutive norm and constitutive goal, not to mention the concept of a reason itself. I will offer a few superficial remarks.

If a participant in a practice desires to obey its constitutive norms or pursue its constitutive goals, the desire can yield a reason for action. On a conception of practical reason often associated with Hume, this is essentially the *only* way constitutive norms and goals provide reasons for conforming to them. Davidson articulates the Humean conception, writing that “[w]henever someone does something for a reason, therefore, he can be characterized as (a) having some sort of pro attitude towards action of a certain kind, and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind” (1980: 4-5). Bernard Williams defends something like the Humean conception under the rubric “internalism” (1997). The conception entails that participants in a practice have reason to obey its constitutive norms and pursue its constitutive goals only insofar as they bear some pro-attitude towards those norms and goals.

An opposing conception, vigorously defended by Searle, holds that, at least in certain circumstances, no such pro-attitude is required. In Searle's terminology, when I voluntarily enter into a practice governed by certain constitutive rules, I thereby generate a *desire-independent reason* for me to obey those rules. Let us call this view the Desire-

Independence Thesis. Brandom (1994: 243-253) and Raz (1990: 117-123) defend versions of the Desire-Independence Thesis. John Rawls (1955) also defends the thesis, although in his later work he restricts it to practices that are “just” in his technical sense (1971: 111-112, 344-347).

Assume that the Desire-Independence Thesis is correct. Then, once I voluntarily engage in some practice, the rational default strategy is to obey its constitutive norms and pursue its constitutive goals, whether or not I bear any pro-attitude towards those norms and goals. For instance, once I voluntarily decide to play a game, I need some special reason to violate the rules of the game, and I need some special reason not to try to win. More relevantly, once I voluntarily opt into the practice of assertion, then, assuming that truthfulness is a constitutive norm or goal of assertion, I have *prima facie*, defeasible reason to speak truthfully. Or, assuming that §3’s version of the dialectical model is correct, then I have *prima facie* reason to assert claims that I expect I can vindicate with cogent arguments. The further question remains why I should enter into assertoric practice in the first place. But, under the proposed view, once I *do* enter into it, I immediately provide myself with *prima facie* reason to engage in it correctly and to pursue its constitutive goals.

R. M. Hare attacks the Desire-Independence Thesis by citing *fox-hunting*. Someone who loathes the practice of fox-hunting might participate in it precisely so as to subvert it, secretly freeing foxes whenever possible. According to Hare, “it may be that to try to help foxes escape is contrary to the constitutive rules for fox-hunting,” but our opponent of fox-hunting has no reason to obey these rules (1969: 154). In response to this objection, Searle can respond that the agent’s dislike for fox-hunting provides a *defeating*

reason that overrides or undermines his *prima facie* reason to engage in the practice correctly. Hence, the counter-example merely exhibits the *prima facie* character of the reason posited by the Desire-Independence Thesis.

Hare's objection is instructive, illustrating both the resilience and the elusiveness of the Desire-Independence Thesis. Like all views that posit *prima facie* reasons, the thesis is difficult to refute through counter-examples, because one can usually rebut putative counter-examples by saying they involve special defeating factors that override or undermine the alleged *prima facie* reason. Yet, precisely because of this extreme resistance to counter-examples, the Desire-Independence Thesis strikes many philosophers as suspect or even empty.

Although I find the Desire-Independence Thesis attractive, I am not aware of any entirely satisfactory argument for it. Many such arguments suffer from insufficient attention to the fundamental point that one can engage in a practice without obeying its constitutive norms or pursuing its constitutive goals. In his initial treatment, Rawls wrote that, "[t]o engage in a practice, to perform those actions specified by a practice, means to follow the appropriate rules. If one wants to do an action which a certain practice specifies then there is no way to do it except to follow the rules which define it" (1955: 26). But a chess player who cheats by covertly rearranging chess pieces is still playing chess. Thus, it is simply not true that participation in practice requires adherence to its constitutive norms. Perhaps for this reason, Rawls later adopted a version of the Desire-Independence Thesis restricted to "just" practices, a version which he defended by drawing heavily upon considerations peculiar to his moral philosophy. In contrast, Searle defends a fairly sweeping version of the Desire-Independence Thesis, and he does so

without invoking moral considerations. But Searle's arguments have not commanded widespread assent.

Whether or not the Desire-Independence Thesis is correct, the constitutive norms and goals of a practice surely occupy *some* kind of privileged role in the practical reasoning of voluntary participants in the practice. We might say that the practice's constitutive norms and goals characterize how I act insofar as I act within my guise as a participant in the practice. The Desire-Independence Thesis is one attempt to capture this intuitive idea more rigorously. The attempt may fail, but any convincing account must find some way of preserving these intuitions.

Leaving these preliminary issues unresolved, I turn to my central task: deriving truthfulness from the dialectical model of assertion.

§5. Deriving truthfulness: epistemic and cognitive asymmetries

In §3, I urged that reasoned discourse involves at least three constitutive goals: rapprochement (isolating relevant, mutually acceptable premises); avoiding decisive counter-arguments against what one says; and providing decisive arguments for what one says. I focus upon the second goal of avoiding decisive counter-arguments. I will first present my argument rather schematically and then fill in some details.

Assume that I seek to avoid decisive counter-arguments. Suppose I assert p and my interlocutor asserts premises q_1, \dots, q_n supporting $\neg p$ over p . Suppose I accept q_1, \dots, q_n . Then I must assert additional premises r_1, \dots, r_m , where $q_1, \dots, q_n, r_1, \dots, r_m$ no longer support $\neg p$ over p . If I cannot isolate suitable premises r_1, \dots, r_m , then q_1, \dots, q_n are decisive against p . Thus, when deciding whether to assert p , I speculate about possible

premises q_1, \dots, q_n that my interlocutor might assert. If I could not provide premises r_1, \dots, r_m counter-balancing q_1, \dots, q_n , I do not assert p .

A well-informed and intelligent interlocutor will doubtless offer challenging counter-arguments. An ill-informed or incompetent interlocutor will doubtless offer rather feeble counter-arguments. If I know my interlocutor fairly well, then I know what quality of counter-argument to expect. But sometimes I know virtually nothing about my interlocutor's epistemic and cognitive standing. In this state of ignorance, how do I speculate about what counter-arguments I will encounter? I have little idea what premises q_1, \dots, q_n my interlocutor will assert, so I have little idea whether I will be able to adduce premises r_1, \dots, r_m sufficient to evade decisive counter-arguments. Under these circumstances, I must reflect upon how *I myself* would argue against my position. I reflect upon the premises q_1, \dots, q_n I would advance against p . I examine whether I could answer my own best counter-arguments.

Numerous factors might lead to me to expect weaker counter-arguments than I myself would offer. Perhaps I know that my interlocutor is very poorly informed. Perhaps I systematically destroyed all evidence against my position. But, lacking relevant information, I should presume that I enjoy no significant, relevant epistemic or cognitive advantage. The default assumption is that other speakers will offer counter-arguments at least as powerful as those I would offer.

When I operate under this default assumption, I decide whether to assert some proposition by subjecting the proposition to vigorous rational scrutiny. I reflect upon what might be said against the proposition, and I reflect upon what might be said in favor of it. I decide to assert it only if I could defeat my own best counter-arguments. I reflect

upon how I would defend the proposition against *myself*. And that means that I attempt to defend the proposition *to* myself. Any evidence that militates against believing some proposition could also fuel counter-arguments against that proposition. To evaluate whether a position will withstand the rational scrutiny of other speakers, I examine whether it withstands my own rational scrutiny. Thus, the process whereby I evaluate some proposition's dialectical strength is the same process whereby I assess whether the proposition seems sensible to believe. I examine various pieces of evidence, such as my perceptual experiences and my other beliefs, evaluating whether they rationally support the relevant proposition. I assert the proposition only if it seems justified in light of my overall evidence. In other words, I aim to speak truthfully.

The preceding four paragraphs present my argument in schematic form. I now want to develop the argument by considering various objections. In this section, I focus on a crucial "lemma" introduced in the middle of the argument: the rational default presumption is that I enjoy no significant, relevant epistemic or cognitive advantages over my interlocutors. Call this the *No Asymmetries Thesis*. The No Asymmetries Thesis may recall the *principle of charity*, espoused by Quine and Davidson in their discussions of translation and interpretation. It may also recall the assumption of *perfect information*, which plays an important role in economics and game theory. Since detailed comparison with these two doctrines would carry us too far afield, I confine myself to defending the thesis on its own terms.

The No Asymmetries Thesis does *not* hold that, lacking any relevant evidence, I am *prima facie* justified in believing that no significant, relevant epistemic or cognitive asymmetries obtain. The thesis holds only that the rational conversational strategy is to

proceed *as if* no such asymmetries obtain. As an analogy, suppose I play chess with an opponent about whom I know virtually nothing. Initially, I am not epistemically justified in holding any beliefs about the opponent's chess skills. Nevertheless, the rational default strategy is to proceed as if he is at least as skilled as I am. For instance, if I see that some contemplated move opens me to checkmate, it would be *prima facie* irrational to make the move while telling myself, "I'm sure that *he* won't see the possibility of checkmate."

Obviously, speakers often exhibit relevant epistemic or cognitive asymmetries. These cases are not devastating counter-examples to the No Asymmetries Thesis, for one can handle them rather mechanically as follows: the assumption of no asymmetries is only a *default*, so it can be overridden by evidence about some particular interlocutor. This mechanical response strikes me as somewhat effective. But it requires supplementation by more detailed consideration of particular cases. I briefly discuss three such cases: perception, rational reflection, and introspection.

Perception of the external world can induce epistemic asymmetries. I may know that I was the sole witness to some transitory past event, or that my interlocutor suffers from perceptual biases, frailties, or constraints. Nevertheless, there is a clear sense in which such cases are deviations from a baseline parity among perceivers. The external physical world is not my private domain, accessible only to me. It is in some sense objective, existing independently of my actions, thoughts, and perceptual experiences. Moreover, there is nothing special about me. I am a cognitive creature much like any other, equipped with the same basic perceptual and cognitive apparatus. We inhabit the same world, and we experience that world in basically the same way. These truisms are grounded in our conception of what it is to be a perceiver. They support the No

Asymmetries Thesis, as applied to perceptual evidence: the rational default presumption is that no significant, relevant perceptual asymmetries obtain.

Similar remarks apply to knowledge gained through rational reflection, as in mathematical proof, logical inference, abductive inference, and generation of novel scientific hypotheses. Normal humans exhibit wide variation in their ability to execute such reasoning. Nevertheless, chains of reasoning are in some sense accessible to all. A given chain of reasoning is not valid for me and invalid for someone else. If I can appreciate its validity, then other speakers can also do so. I may recognize that some reasoning is especially difficult or obscure, as in a newly discovered mathematical proof. But, lacking evidence that the reasoning is notably difficult or obscure, the rational presumption is that my fellow conversationalist grasps it as well.

Introspection provides the most serious challenge to the No Asymmetries Thesis. Whereas perception and reason are inherently transpersonal, most philosophers regard knowledge gained through introspection as somehow inherently private. Thus, introspection arguably exhibits constitutive self-other asymmetries that militate decisively against the No Asymmetries Thesis.

These worries are serious, but ultimately I think we can finesse them. Although any good philosophical account of self-knowledge must preserve *some* self-other asymmetries, a good account must also acknowledge that, in a vast range of cases, we are extraordinarily adept at divining one another's thoughts through voluntary and involuntary bodily manifestations.¹¹ My interlocutors can cite these manifestations, although with other relevant considerations, such as my internal physiology, general constraints imposed by theoretical and practical reason, and so on. These are not

phenomena regarding which I possess any special authority. Thus, even when we debate my own mental states, I can expect counter-arguments regarding whose premises I enjoy no special epistemic or dialectical advantages. I can dogmatically dismiss such arguments by saying something like “Well, I know how I feel,” or “Just trust me.” But then I have not avoided decisive counter-arguments. I have simply failed to engage with them.

Let us henceforth assume the No Asymmetries Thesis. Even so, an important worry persists. In practice, I always possess considerable evidence regarding my interlocutor’s dialectical resources. So the No Asymmetries Thesis may appear otiose. Either my current evidence already entails that no relevant asymmetries obtain; or else my evidence outweighs any default presumption that no asymmetries obtain.

To address this objection, consider *modest foundationalism* regarding perception, according to which one is *prima facie*, defeasibly entitled to accept perceptual experiences at face value.¹² In practice, one always possesses supplementary evidence either confirming or infirming one’s perceptual faculties. The inevitability of such supplementary evidence does not impugn modest foundationalism about perception. The supplementary evidence is irrelevant to whether perceptual experience, taken by itself, furnishes warrant for belief. Similarly, the inevitability of supplementary evidence about one’s interlocutors is irrelevant to whether constitutive features of assertion, taken by themselves, generate a rational bias towards truthfulness.

A less generic response to the objection hinges upon the following observation: the constitutive goal of reasoned discourse is to avoid decisive counter-arguments *in general*, not simply decisive counter-arguments offered by one’s present interlocutors. Once I assert a proposition, other speakers may legitimately challenge my assertion, even

if they were not present at the time. The constitutive norms of assertion apply to these belated challenges, albeit perhaps with diminished force. I may know quite a bit about my present interlocutors, but I generally know much less about what interlocutors I will eventually encounter. Thus, the No Asymmetries Thesis remains relevant even when I possess extensive information about my current interlocutors.

§6. Deriving truthfulness: alternative conversational strategies

I turn now to the final stage of the schematic argument from §5. The final stage maintains that, assuming no relevant epistemic or cognitive asymmetries obtain, truthfulness is the rational default strategy for avoiding decisive counter-arguments. An underlying assumption here is that my interlocutors will offer any counter-arguments at all. If this assumption lapses, then I can say whatever I want without fear of retraction. The assumption seems reasonable. Admittedly, in many contexts, I am unlikely to encounter counter-arguments against what I say. I may know that my interlocutor already agrees with me or that he is too polite to voice his disagreement. However, I am unlikely to know that *future* interlocutors will similarly abstain from disputation. If I seek to avoid future decisive counter-arguments, the rational default strategy is to presume that other speakers will dispute what I say.

Even assuming that I will eventually encounter formidable opposition, various alternative conversational strategies may seem at least as effective as truthfulness at avoiding decisive counter-arguments.

One such strategy is *stonewalling*. When my interlocutor offers an argument against my position, I can refuse to concede the argument's premises. This will place my

interlocutor on the defensive, potentially initiating a regress of justifications. Since decisive counter-arguments arise only when I accept my interlocutor's premises, this conversational strategy stymies decisive counter-arguments. Stonewalling will often lead me to challenge propositions that no normal person would doubt, such as obvious propositions about my observable surroundings. But such conduct, while very strange, seems perfectly consistent with the Defense and Retraction Norms, as well as the constitutive goal of avoiding decisive counter-arguments.

In response to this objection, I concede that this constitutive goal does not itself favor truthfulness over stonewalling. However, as we saw in §3, *another* constitutive goal of reasoned discourse renders stonewalling highly deviant: rapprochement.

Conversationalists who pursue this goal seek mutually acceptable premises, which serve as a neutral evidentiary base for evaluating disputed claims. If I challenge relatively unproblematic claims without offering reasons to doubt them, or if I refuse to concede *anything* my interlocutor says, I abandon this goal. I forgo genuine rational engagement.

Stonewalling subverts the most central constitutive goal of reasoned discourse: convergence upon neutral argumentative common ground. In contrast, truthfulness is an excellent way to achieve rapprochement, given the No Asymmetries Thesis. Thus, truthfulness is the *prima facie* superior strategy for evading decisive counter-arguments.

This stage of the argument illustrates the more general point, emphasized in §2, that the constitutive goals of a practice may conflict with one another. If I sought only to avoid decisive counter-arguments, the safest strategy would be to stonewall. But normal participation in reasoned discourse requires avoiding decisive counter-arguments while pursuing the more fundamental goal of achieving rapprochement.

To strengthen this stage of the argument, we might additionally invoke our final constitutive goal: furnishing decisive arguments for one's position. Stonewalling subverts this goal, since it prevents speakers from attaining argumentative common ground. However, such maneuvers are superfluous, since the constitutive goal of rapprochement already renders stonewalling highly deviant.

Another alternative conversational strategy is to *dissemble*. With enough ingenuity, I can concoct an elaborate scenario supporting virtually any non-contradictory proposition. I can defend my lies with further lies. If I am sufficiently skillful, I may even convince my interlocutors to accept these lies. *Othello* vividly showcases this alternative conversational strategy. In effect, Iago constructs an internally consistent alternative reality so plausible that he persuades Othello of its verisimilitude.

Crucially, Iago's conversational strategy consumes substantial cognitive resources. The strategy demands major expenditures of time, energy, and ingenuity. Iago must *fabricate* reasons for believing claims he himself disbelieves. When I defend my lies with further lies, I construct an elaborate network of deceptive justifications. As Sir Walter Scott observes, "O what a tangled web we weave/When first we practice to deceive!"¹³ In contrast, when I speak the truth, I defend what I say by reproducing my internal edifice of justifications. I publicize *my own* reasons for believing what I say. Since truthfulness offers such an effortless strategy for evading decisive counter-arguments, we require special reason to deviate from it. We require special reason to invest cognitive resources in alternative strategies.

Quite aside from the excessive demands lying places upon one's cognitive resources, it is an inherently *riskier* conversational strategy than truthfulness. Even if I

successfully fool my present interlocutor, how can I know I will fool whatever future interlocutors I encounter? The truth is out there, waiting for others to discover it. The most convincing liar may be exposed, and the most plausible falsehood may stand refuted. Even the brilliant Iago eventually encounters defeat. Conclusive evidence surfaces against his lies, and he retracts his position. In general, although not invariably, *truth will out*. The risks of lying generate further rational bias against choosing it over truthfulness, lacking special reasons for such a choice.

At this point, we should distinguish various meanings of phrases such as “truthful,” “truthfulness,” and “speak truthfully.” Two possible meanings are: aiming to assert true propositions; and aiming to assert propositions supported by one’s overall present evidence. A scientist who aims to assert true propositions may conduct additional experiments before endorsing some hypothesis. She need not conduct these experiments if she seeks only to assert propositions she is presently warranted in believing. Thus, aiming to assert true propositions is not the same as aiming to assert warranted propositions. The arguments developed above do not distinguish between these two aims. The arguments secure a reason for asserting warranted propositions, but I am less confident that they also secure a reason for asserting true propositions. Since securing a reason to assert warranted propositions already seems like enough to establish an intimate connection between truth and assertion, I will not investigate whether arguments similar to mine can additionally secure a reason to assert true propositions.

As a heuristic application of my argument, we can imagine an amoral misanthrope who revels in the unique thrill of well-conducted rational conversation. The misanthrope does not care about other people. He wants nothing from them except to

engage them in reasoned discourse, and he demands nothing from them except that they engage in reasoned discourse correctly. He does not even care about convincing others that he is right. He simply wants to argue. Still, if he wants to avoid losing the argument, he has *prima facie* reason to speak truthfully.

Some philosophers might deny the possibility of a rational agent who rejects moral, altruistic, and social factors when deciding what to do. They might claim that such a specimen is so deviant that it makes no sense to talk about his reasons for acting: he does not really act at all. I disagree, but my argument in this section does not turn upon my disagreement. My main goal here is to illuminate *our own* reasons for truthfulness. I have tried to isolate a specifically linguistic reason for truthfulness, one that emerges from considerations internal to reasoned discourse, without any input from morality, legality, propriety, or even prudential rationality.

Many factors might defeat this dialectical reason for truthfulness. We may classify the factors into two broad categories: *overriding* and *undermining* defeaters. Overriding defeaters outweigh the dialectical reason for truthfulness by providing reason to pursue an alternative conversational strategy. An extreme example would be a mother who lies to protect her son from arrest. Undermining defeaters subvert the dialectical reason without providing any reason to pursue an alternative conversational strategy. For instance, suppose John knows he has destroyed all evidence against *p*. Then the goal of avoiding decisive counter-arguments no longer provides any reason for John not to assert *p*. He knows that he enjoys a relevant epistemic advantage over other speakers, and this knowledge undermines the dialectical reason for truthfulness. The knowledge provides

no reason for him to stonewall, to dissemble, or to pursue any other conversational strategy. It simply nullifies the dialectical reason.

Any plausible account must make room for overriding defeaters. They arise with just as much prevalence and force for an account like Searle's, which treats the Honesty Norm as constitutive of assertion and deploys the Desire-Independence Thesis to secure a *prima facie* reason for truthfulness. A distinctive feature of my approach is the prominent role it accords *undermining* defeaters. Such defeaters do not arise very naturally on Searle's view, which claims that correct participation in assertoric practice *is* truthful assertion. My account allows undermining defeaters to proliferate. Any evidence of a relevant epistemic advantage over one's present and future interlocutors might serve.

Thus, my account isolates an extremely fragile reason for truthfulness. Virtually any reason for lying can outweigh it, and virtually any evidence of relevant epistemic or cognitive advantages can undermine it. I think that this extreme fragility is a virtue, rather a defect. We should not posit an overly strong rational bias towards truthfulness, on pain of depicting ordinary linguistic commerce, so frequently tinged by deceit, as overly irrational. A satisfactory theory of assertion must honor the rationality of quotidian mendacity. My goal here is to understand how truthfulness figures in our reasons for speaking as we do, not to extol truthfulness as the *sine qua non* of rational conversation.

§7. Comparing the accounts

In the previous three sections, I developed a model of assertion's constitutive elements, and I argued that those elements generate a rational impetus towards

truthfulness. I will now compare my account with the more straightforward Constitutive Thesis, which treats truthfulness as constitutive of assertion.¹⁴

When investigating constitutive features of assertion, a fundamental difficulty arises. Ordinary speakers may harbor a bewildering variety of goals in asserting propositions. On what basis can we promote certain of these goals as constitutive while denigrating the rest as non-constitutive? The difficulty becomes particularly acute when we acknowledge, as any plausible account must, that individual speakers may not seek to achieve assertion's constitutive goals or obey its constitutive norms.

As I have emphasized, a practice's constitutive norms and goals must occupy a privileged role in the practical reasoning of the practice's voluntary participants. How to characterize this privileged role remains unclear. The Desire-Independence Thesis is one attractive but controversial attempt. What seems clear, though, is that a community whose inhabitants assign some practice's constitutive norms and goals *no* special weight do not instantiate the practice. For instance, a community whose members move pieces around a chess board while assigning no special weight to chess's rules do not play chess. These people do not display appropriate sensitivity to the rules to count as playing a game constituted by those rules. (Cf. Dummett 1991: 88-89.) Similarly, to adapt an example due to Raz, imagine a community whose members move pieces on a chess board in scrupulous conformity to chess's rules while helping one another achieve checkmate. These people still do not play chess, because they do not assign sufficient priority to winning (Raz 1990: 118).

Such examples suggest the following bipartite methodology:

(a) If there is a possible group whose members execute a practice P while assigning no special weight to some norm or goal, then the norm or goal is not constitutive of P .

(b) Conversely, if groups that assign no special weight to some norm or goal do not instantiate P , then a plausible explanation is that the norm or goal is constitutive of P .

Principle (b) is far more tenuous than principle (a), since someone might propose *alternative* explanations that do not ascribe constitutive norms or goals to P .

Let us apply this bipartite methodology to assertion.

Imagine a group G_1 of isolated individuals, inhabiting a vast, sparsely populated wilderness. Each individual is economically self-subsistent. Members of G_1 do not seek out interaction with one another, but they occasionally meet by accident in the wilderness. I stipulate that these individuals possess innate knowledge of a common primitive language through which they can express and understand propositional contents. The language includes a central speech act ψ . Members of G_1 generally try to ψ only true propositions, a proclivity which is common knowledge within G_1 . When special overriding considerations arise, they may deviate from this pattern of truthful ψ ing. But the considerations must be very weighty to override the default predilection towards truthfulness. I furthermore stipulate that members of G_1 do *not* participate in anything resembling reasoned discourse, a fact which is also common knowledge. For instance, if A ψ s that p and B ψ s the proposition expressed by “ p cannot be true, because of q ,” where q is some highly credible proposition that militates against p , then A is likely to shrug his shoulders and walk away, a reaction which B would not find anomalous or defective. In general, members of G_1 make no effort to defend or vindicate propositions that they ψ .

They regard it as very important to ψ only true propositions, but they attach no importance to establishing subsequently that they ϕ ed true propositions.

Before examining the thought experiment's significance, we must ask whether it is coherent. If speakers assiduously seek to ψ true propositions, don't they also have an interest in establishing that they ϕ ed true propositions? And doesn't this show that my description attributes pervasive rational deficiencies to members of G_1 ?

I do not think so. Someone who seeks to ψ true propositions need not have any reason to ψ additional propositions supporting the truth of propositions previously ϕ ed. The former goal might reflect an underlying concern, such as maintaining a reputation for truthful ψ ing, that also supports the latter. But it need not. Members of G_1 might regard ψ ing true propositions as a cultural, religious, moral, or imperative. They might regard it as a constitutive norm of their linguistic practice. Such reasons are intelligible, and they need not provide any reason for vindicating what one says. Thus, my stipulations regarding G_1 seem legitimate. This legitimacy would persist if we altered the thought experiment in various ways, such as positing that members of G_1 strive to ψ only propositions that they know. I will not explore such alterations here.

Taking our description of G_1 as coherent, I think speech act ψ is distinct from assertion. More generally, I do not think that G_1 instantiates anything like assertoric practice, so long as G_1 involves no relevant linguistic activity besides what I have stipulated. By methodological principle (b) above, one natural explanation is that G_1 omits constitutive features surrounding assertion's role in rational dialectic. To adopt this explanation is to embrace the dialectical model. As always, alternative explanations are

possible. But the fact that the dialectical model so effortlessly handles our intuitions about this thought experiment surely counts in its favor.

We must proceed carefully here, because intuitions about whether a speech act counts as assertoric are highly theory-laden. An opponent of the dialectical model might reply to my argument by insisting that ψ , as employed by G_1 , is assertion. Alternatively, an opponent could respond that, whether or not G_1 instantiates assertion, it instantiates some speech act assertion* that shares philosophically relevant features with assertion.

These maneuvers become less attractive when we reflect upon G_1 's sheer strangeness. There is a natural sense in which these people simply talk past one another without achieving rational engagement. They do not employ ψ to resolve disagreements. They do not reason with one another at all. For instance, if A and B ψ conflicting propositions, then neither speaker regards the conflict as something to be resolved through further speech acts. The conflict may precipitate individual rational reflection by A or by B , but it does not precipitate joint rational deliberation between A and B . A practice with these features is not necessarily useless. It might serve to transmit justified beliefs between speakers, although I think the justification would be quite different than what *we* acquire through assertion. But the thought experiment highlights that our use of assertion involves much more than the transmission of justified beliefs. We regard our assertoric performances as participating in a broader practice of giving and asking for reasons. Once detached from this practice, the resulting speech act assumes a fundamentally different character. Just imagine living among the people of G_1 . Although I might learn much from what they say, I could not really converse with them. Conversation presupposes an overt sensitivity to reasons that these people lack.

Let us consider a second thought experiment. Like the members of G_1 , the members of G_2 are isolated and self-subsistent. Their language involves a central speech act ϕ and a dual speech act $\underline{\phi}$ corresponding roughly to retraction. Members of G_2 are highly amoral and asocial. They recognize no moral, religious, or cultural imperative towards truthfulness. They do not regard ϕ ing true propositions as inherently desirable or worthwhile. However, they attach great importance to the constitutive elements of reasoned discourse, as described in §3. For instance, in conformity with the Retraction Norm, if $A \phi$ s p and $B \phi$ s some highly credible proposition that militates against p , then A attempts to ϕ additional propositions that undermine or override B 's argument. If A cannot locate such propositions, then $A \underline{\phi}$ s p . Thus, G_2 is basically a community-scaled version of the “amoral misanthrope” posited in §6. I stipulate that the foregoing facts are common knowledge among members of G_2 .

As I argued in §§5-6, truthfulness is the rational default strategy for avoiding decisive counter-arguments without stonewalling. Generalizing this argument, and assuming that members of G_2 are rational, it follows that truthfulness serves as the default conversational strategy within G_2 . Lacking any relevant defeating reasons, members of G_2 aim to ϕ true propositions. For instance, they do not generally ϕ blatant falsehoods about their observable surroundings, because the only way to avoid $\underline{\phi}$ ing these falsehoods is through deviant strategies like stonewalling or constructing elaborate skeptical scenarios. However, when relevant undermining or overriding factors arise, the misanthropes may rationally lie. If they expect to avoid detection, or if lying would advance their interests, then they ϕ propositions they believe false. Because the

misanthropes are untrustworthy, they mistrust one another. They carefully scrutinize one another's ϕ ings, assessing whether to trust a given speaker regarding a given topic.

Despite their moral failings, the misanthropes strike me as *linguistically* quite impeccable. While operating against a background of cynicism and malfeasance, they succeed in asserting, defending, challenging, and retracting claims. Yet they do not assign special weight to truthfulness in their practical reasoning. They pursue truthfulness as a default strategy, but they abandon it as soon as undermining defeaters arise. Truthfulness figures solely as a means to the intrinsically worthwhile ends of achieving rapprochement and avoiding decisive counter-arguments. By principle (a), we may infer that truthfulness is not constitutive of assertion.

To render the example especially vivid, we can stipulate that evidential states and prudential interests align within G_2 so that untruthful ϕ ings *statistically far outnumber* truthful ϕ ings. Under this stipulation, the misanthropes almost never try to ϕ true propositions. But my argument does not depend upon such statistical assumptions. We might equally assume that evidential states and prudential interests yield a statistically overwhelming predilection towards truthfulness. Even then, G_2 would provide reason for doubting the Constitutive Thesis. The main point here is not *statistical* but *rational*. No matter how frequently they speak the truth, the misanthropes do not assign proper weight to truthfulness within their practical reasoning. (I must admit that, although my argument does not depend upon statistical assumptions, the thought experiment seems far more intuitively compelling if we stipulate that truthful ϕ ings are statistically rare.)

As in our discussion of G_1 , intuitions about what counts as an assertion are highly theory-laden. An opponent of my argument might simply insist that G_2 does not

instantiate genuine assertion. I think this reaction seems very implausible once we recognize that, in many crucial respects, G_2 resembles our own linguistic activity far more than does G_1 . If I found myself living among the people of G_2 , I might not *trust* them, but I could talk to them. I would find them repugnant, but I would recognize their linguistic activity as relatively familiar, the result of divesting our own linguistic activity of familiar extra-linguistic accoutrements like morality, propriety, legality, and decency. If I found myself living among the people from G_1 , I would find their linguistic activity quite alien. I could not talk to them in anything resembling a normal way. I can reason with people who routinely lie, but I cannot reason with people who exhibit no inclination to respond appropriately to challenges and counter-arguments.

The contrast between G_1 and G_2 suggests that assertion is constituted not by truthfulness but by dialectical factors. Our own linguistic practice involves both elements. Yet the two elements do not seem equally essential. When we downgrade the second while retaining the first, the resulting practice is so alien that rational conversation becomes impossible. When we downgrade the first while retaining the second, the resulting practice is a horrifying but recognizable caricature of our own. I submit that, while truthfulness may be a virtue, but it is not a virtue built into assertion's intrinsic nature. It is not a goal one must entertain, *qua* participant in assertoric practice. Members of G_1 overlook a fundamental aspect of assertion. Members of G_2 do not.

Ideally, we would conduct a finer-grained analysis by comparing various subtly different thought experiments along the lines of G_1 and G_2 . I suspect that such comparisons help discriminate between rival versions of the dialectical model while

casting further doubt upon the Constitutive Thesis. Rather than pursue this line of inquiry, I conclude with some global reflections about thought experiments like G_2 .

Many philosophers find something defective or even incoherent in a putative linguistic community whose members exhibit no tendency towards truthfulness. From this perspective, G_2 may seem impossible, incoherent, or, at the very least, plagued by systemic rational defects. Within this general perspective, I distinguish four more specific analyses: a tendency towards truthfulness is necessary for the existence of society; or for linguistic expressions and speech acts to possess propositional content; or for language to be learnable by children; or for language to transmit justified beliefs.

Peter Winch endorses the first analysis: “general adherence to [a norm of truth-telling] is a feature of any society in which there are conventions, that is, any society *tout court*” (1972: 62-63). Similarly, to explain why devils would speak truthfully to one another, Dr. Johnson urges that “truth is necessary to all societies; nor should the society of hell subsist without it” (1753: 361).¹⁵ C. A. J. Coady (1994), Michael Dummett (1993), and Peter Winch (1972) endorse the second analysis, with Dummett writing that “we can know *a priori* that, among human beings generally, lying is rare. If it were not that most assertions seriously made are made in the belief that they are true, our words could not mean what they mean” (1993: 426-427). The third analysis arguably informs the writings of Quine and Davidson on translation and interpretation. Coady explicitly advocates it, as does Bernard Williams, who writes that “for many sentences, if their meaning is to be learned, they have to be uttered in situations in which they are true” (2002: 284). The fourth analysis, famously expressed by Kant’s verdict that lying so as to deceive is not “universalizable,” is embraced by Coady and possibly also by Dummett.¹⁶

While the first analysis is very striking, we may dismiss it as irrelevant to our discussion. We can concede that members of G_2 do not constitute a society, whatever exactly a “society” is. Since it is hardly obvious that language, assertion, and communication are possible only within a society, as opposed to a “non-social” collection of isolated individuals, the concession does not affect my argument.

What about the remaining three analyses? Perhaps *some* tendency towards the truth is a prerequisite for language to possess propositional content, to be learnable, and to transmit justified beliefs. But is the requisite “tendency” statistical, nomological, normative, rational, or a hybrid of the foregoing, or something else altogether? Scenario G_2 exhibits *some* tendency towards the truth, since truthfulness is the rational default strategy for avoiding decisive counter-arguments. Maybe this is tendency enough to secure propositional content, learnability, and justified belief-transmission.

Consider the purely statistical interpretation of “tendency.” Specifically, consider the thesis that propositional content emerges within a linguistic practice only if truthful assertions statistically outnumber untruthful assertions. Dummett and Winch explicitly advance this statistical thesis.¹⁷ The thesis does not militate very strongly against my use of G_2 , since I argued earlier that G_2 undermines the Constitutive Thesis even if evidential states and prudential interests yield a statistical predilection towards truthfulness. More importantly, though, the statistical thesis seems implausible. A good theory of assertion must acknowledge that we lie quite routinely. This fact, which figures so prominently in recent commentary on the epistemology of testimony, is strangely absent from writings on assertion itself. Many such writings proceed as if it were obvious that we almost always speak honestly. As far as I know, no sociologist has conducted a rigorous

statistical investigation of how frequently people speak honestly or dishonestly. A sociological investigation might well reveal that lies far outnumber honest assertions.

Does there exist a powerful *a priori* argument against this possibility? Dummett provides little argument for the statistical thesis. Winch defends the thesis by considering two highly deviant thought experiments: one in which “what we now call ‘true’ statements [are] always uttered in place of what we now call ‘false’ statements, and vice versa”; the other in which “the incidence of ‘true’ and ‘false’ statement [is] statistically random” (1972: 62). No doubt there is something incoherent or rationally defective about a practice in which people *always* try to assert falsehoods where we assert truths, or one in which they *randomly* assert truths or falsehoods. G_2 does not exhibit such bizarre features. It showcases a well-constructed linguistic practice constituted by clearly defined norms and goals. There is no evident reason why such a practice would collapse if it did not generate a statistical bias towards truthful assertion.

In this connection, Coady offers a particularly revealing discussion. He considers a community of hypothetical Martians who “constantly misinform each other about issues in dispute or unresolved issues of interest and never correct each other’s misinformation on the basis of their own observations” (1994: 233). He argues that the thought experiment is impossible, since the alleged speech acts of this community could not possess the alleged propositional contents. He furthermore argues that, even if the scenario *were* possible, Martian children could not learn the meanings of words. Coady concludes that any viable, learnable linguistic practice must involve a “quite extensive connection... between testimony and reality” (1994: 245).

Like Winch, however, Coady prejudices his case by focusing upon a highly unrepresentative thought experiment. Coady's Martians constantly utter *blatant* falsehoods concerning topics such as the current weather. The Martians never challenge one another's blatantly false assertions. They make no effort to advance defensible claims, to avoid decisive counter-arguments, to avoid stonewalling, or to provide arguments against obvious falsehoods. It seems clear, then, that the Martians do not instantiate a practice remotely resembling reasoned discourse. Perhaps Coady's analysis of his Martians is correct. But a community like G_2 , which instantiates reasoned discourse without enshrining any intrinsically privileged role for truthfulness, falls outside the scope of Coady's critique. As already noted, members of G_2 do not generally utter blatant falsehoods about their surroundings. If one of them did utter such a falsehood, then other speakers, being highly disputatious, would surely challenge her. Thus, G_2 features a solid baseline of true assertions, most notably true assertions about speakers' immediate observable surroundings, even though lies about more recondite matters may statistically outnumber those true assertions. There is no obvious reason to dismiss such a linguistic practice as incoherent or unlearnable.¹⁸

But what about transmitting justified beliefs through assertion? Doesn't that require a more pervasive bias towards truthfulness than what G_2 offers?

The issue is complex, because considerable controversy surrounds *our own* justification for believing what other people say. On a broadly Humean conception, we depend upon inductive or abductive evidence that other speakers are honest and reliable. On a broadly Reidian conception, we are *prima facie*, defeasibly, immediately entitled to accept what other people say at face value, without any mediating inductive or abductive

evidence. Both positions face serious problems.¹⁹ However, G_2 is not obviously incompatible with a suitably sophisticated version of either the Humean *or* the Reidian position. Modern Humeans such as Elizabeth Fricker usually emphasize “local justification” about the honesty and reliability of particular informants on particular occasions regarding particular topics, rather than “global justification” about speakers in general (Fricker 1994). Local justification is available within G_2 even if global justification is not. Sophisticated contemporary Reidians like Tyler Burge eschew implausible appeals to an innate or statistical bias towards truthfulness, instead trying to isolate a *rational* bias (Burge 1993). I have argued that even G_2 features a rational bias towards truthfulness. Thus, a Reidian account might apply just as well to G_2 as it does to us, although members of G_2 are far more likely than us to encounter defeating evidence against trusting one another.

Full development of these ideas would require a separate paper. My point is just that the epistemology of testimony yields no quick dismissal of G_2 as incoherent, impossible, rationally defective, or non-assertoric.

There is surely something correct in the intuitive idea that, if we did not display a minimal investment in the truth of what we say, then assertoric practice would collapse. But it is not clear that this intuitive idea militates against G_2 , or more generally against my version of the dialectical model. On my view, assertoric practice depends upon a baseline of serious investment in *vindicating* the truth of what one says. This is not the same as an investment in speaking the truth, although I have argued that it generates a rational bias in that direction. What would fatally undermine assertion is pervasive *blatant* disregard for the truth of what one says, not *devious* mendacity of the kind

practiced within G_2 . Systematic lying need not undermine assertion, as long as the liars carefully avoid detection. If the devils in hell speak truthfully to one another, then they need do so only insofar as they can catch one another's lies.

The view I have developed probably seems inconsonant with a traditional philosophical conception of humans as noble seekers after truth. I think that much recommends my alternative conception. It isolates a pervasive, if fragile, reason for speaking truthfully. It comports well with our intuitions about various hypothetical linguistic practices. And, although I have not developed the point here, it may well yield a plausible version of Reidianism about testimony. Besides articulating a convincing analysis of testimony, the major challenges facing my view include: addressing Davidson's skepticism regarding whether assertion involves constitutive norms and goals at all; exploring the Desire-Independence Thesis and related doctrines; and combating the existing literature's many additional arguments for the Constitutive Thesis.

Notes

¹ See note 15.

² A related but distinct approach that I will not discuss in this paper is to treat truthfulness as constitutive not just of assertion but of practical reason. Kant's moral philosophy provides the most famous illustration of this explanatory strategy. (Korsgaard 1996a, 1996b) elaborates and defends the Kantian perspective. Tyler Burge also develops this explanatory strategy, but without any explicit emphasis upon morality: "lying for the fun of it is a form of craziness... Lying occasions a disunity among functions of reason. It conflicts with one's reason's transpersonal function of presenting the truth, independently of special personal interest... Reason has a function in providing guidance to truth, in presenting and promoting truth without regard to individual interest" (1993: 475).

³ Bernard Williams rejects the Constitutive Thesis (2002: 66-79), and he offers a multi-faceted "genealogical" explanation of why truthfulness nevertheless occupies a rationally privileged role in

assertoric practice. I am not sure I understand Williams's genealogical explanation, but the basic idea is to embed sincerity within a "structure of other virtues and values... in such a way that the reflective agent can make sense of it as an intrinsic value" (2002: 95). Although Williams stresses that these "other virtues and values" may vary considerably with historical circumstances, he highlights how "mutual respect and the capacity for shame in the face of oneself and others" underwrite our own present interest in cultivating trusting relations with one another, which in turn confers an intrinsic value upon truthfulness (2002: 121). Williams's approach is diametrically opposed to mine. Whereas I try to *detach* assertion from more general ethical and social factors, Williams regards it as hopelessly intertwined with such factors. From Williams's perspective, no doubt, my account draws suspect boundaries between the linguistic and the extra-linguistic.

⁴ As many authors have emphasized, practices typically involve constitutive elements beyond what I am calling their constitutive norms and constitutive goals. For instance, baseball involves constitutive rules that determine how the score evolves as the game progresses. For discussion of this point, see (Rawls 1955), (Raz 1990), and (Searle 1969).

⁵ Another prominent philosopher to reject constitutive norms of assertion is Paul Grice, who initiated a program of analyzing assertion in terms of speakers' communicative intentions and expectations. Unlike Davidson, Grice does not explicitly deny that assertion falls under constitutive norms and goals; he merely eschews such notions in his theorizing. An additional wrinkle is that several recent authors interpret Grice's celebrated Maxim of Quality as a constitutive norm of assertion. For this interpretation, see (Williamson 2000: 243) and, even more explicitly, (DeRose 2002: 199). I believe this interpretation to be mistaken. The Maxim of Quality has the same *content* as the Truth Norm, but it has a completely different *normative status*, since it results from subsuming assertion under more general extra-linguistic norms of morality and practical reason. Although a complete account would undertake a detailed comparison of my approach with Grice's, I will not undertake such a comparison here.

⁶ Until the play's end, at which point Iago extrudes himself altogether from the practice of reasoned discourse: "From this time forth I never will speak word" (Act V, Scene 2).

⁷ The retraction occurs off-stage. It is reported by Lodovico: "This wretch hath part confess'd his villany" (Act V, Scene 2).

⁸ I am not sure whether this quote represents Dummett's own view or whether he merely intends to describe one possible view of assertion.

⁹ Their account is considerably more complex than this sketch indicates. Walton and Krabbe present a range of dialectical models featuring subtly different norms governing retraction of assertoric and dark-side commitments. Moreover, Walton and Krabbe articulate norms that ensure consistency of retraction, so that, for instance, a speaker cannot simply retract some conclusion while refusing to retract previously asserted premises supporting that conclusion.

¹⁰ Walton and Krabbe endorse a somewhat similar set of constitutive goals (1995: 135). Recognition that a good account of rational dialectic requires something along these lines goes back to Aristotle. Like Aristotle, Walton and Krabbe emphasize the goal of *convincing other people of what one says*. In contrast, my account posits no constitutive link between rational dialectic and persuasion. Although reasoned discourse can serve this purpose, it is not fundamentally a device for influencing what people think. Reasoned discourse aims constitutively at vindication, not persuasion.

¹¹ For extensive discussion, see (Nichols and Stich 2003).

¹² For defense of modest foundationalism about perception, see (Alston 1989), (Audi 1993), (Peacocke 2004), and (Pryor 2000).

¹³ "Marmion," Canto VI, Stanza 17.

¹⁴ Dummett embraces a highly anti-realist notion of truth, on which truth is intimately connected to justification. Given Dummett's anti-realism, his version of the Constitutive Thesis may render it more similar to my position than it initially appears. Indeed, Dummett frequently emphasizes how the speech act of retraction illuminates the relation between truth and assertion; see, for instance, (Dummett 1978: xvii.) Thus, perhaps we should describe the difference between Dummett and myself as follows. Dummett and I both ground the relation between truth and assertion in the phenomenon of "decisive counter-arguments" and the correlated speech act of retraction. Dummett's approach is to explicate the notion of truth in broadly verificationist terms. My approach is to derive a *prima facie* reason for truthfulness from constitutive features of rational dialectic. In effect, then, my position assimilates many Dummettian insights without embracing Dummett's anti-realism.

¹⁵ Johnson attributes this quote to Sir Thomas Brown. The editors of Johnson's *Works* report that it is a paraphrase of a similar passage from Brown's writings.

¹⁶ Many philosophers employ such analyses to isolate a *prima facie* reason for speaking truthfully. The basic idea is that we share an interest in truthfulness, because if *no one* spoke truthfully then society, language, assertion, or communication would become impossible. I will not discuss such arguments here, since they raise various extremely complex issues surrounding free-riders and the paradox of the common.

¹⁷ See also (Lewis 1983: 182).

¹⁸ For additional criticisms of Coady, see (Graham 2000).

¹⁹ For a helpful survey of the literature, see (Adler 2006).

Works Cited

- Adler, J. (2006) "Epistemological Problems of Testimony", in *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2006/entries/testimony-episprob/>
- Alston, W. (1989) *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca: Cornell University Press.
- Audi, R. (1993) *The Structure of Justification*, Cambridge: Cambridge University Press.
- Brandom, R. (1994) *Making it Explicit*, Cambridge, MA: Harvard University Press.
- Burge, T. (1993) "Content Preservation," *Philosophical Review*, 102: 457-488.
- Coady, C. A. J. (1994) "Testimony, Observation, and 'Autonomous Knowledge'," in B. Matilal and A. Chakrabarti (eds.) *Knowing from Words*, Boston: Kluwer.
- Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.
- (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford

University Press.

DeRose, K. (2002) "Assertion, Knowledge, and Context," *Philosophical Review*, 111: 167-203.

Dummett, M. (1978) *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press.

---. (1991) *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press.

---. (1993) *The Seas of Language*, Oxford: Clarendon Press.

Fricker, E. (1994) *Against Gullibility*, in B. Matilal and A. Chakrabarti (eds.) *Knowing from Words*, Boston: Kluwer.

Glanzberg, M. (2004) "Against Truth-Value Gaps," in J. C. Beall (ed.) *Liars and Heaps: New Essay on Paradox*, Oxford: Oxford University Press.

Goldman, A. (1994) "Argumentation and Social Epistemology," *Journal of Philosophy*, 91: 27-49.

Graham, P. (2000) "The Reliability of Testimony," *Philosophy and Phenomenological Research*, 61: 695-709.

Hamblin, C. L. (1970) *Fallacies*, London: Methuen & Co., Ltd.

Hare, R. M. (1969) "The Promising Game," in W. D. Hudson (ed.) *The Is-Ought Question*, London: Macmillan and Co., Ltd.

Johnson, Samuel (1753) "On Lying," *The Adventurer*, 50, rpt. in W. J. Bate, J. Bullitt, and L. F. Powell (eds.) (1963), *The Yale Edition of the Works of Samuel Johnson, Volume II*, New Haven: Yale University Press.

Korsgaard, C. (1996a) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.

-
- . (1996b) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Leite, A. (2005) "A Localist Solution to the Regress of Justification," *Australasian Journal of Philosophy*, forthcoming.
- Lewis, D. (1983) "Languages and Language," in *Philosophical Papers I*, Oxford: Oxford University Press.
- Lorenzen, P. (1987) "Logik und Agon," in P. Lorenzen and K. Lorenz (eds.) *Dialogische Logik*, Darmstadt: Wissenschaftliche Buchgesellschaft.
- MacFarlane, J. (2003) "Future Contingents and Relative Truth," *Philosophical Quarterly*, 53: 250-65.
- . (2005) "Making Sense of Relative Truth," *Proceedings of the Aristotelian Society*, 105: 321-39.
- Peacocke, C. (2004) *The Realm of Reason*, New York: Clarendon Press.
- Pryor, J. (2000) "The Skeptic and the Dogmatist," *Nous*, 34: 517-549.
- Nichols, S. and Stich, S. (2003) *Mindreading*, Oxford: Oxford University Press.
- Rawls, J. (1955) "Two Concepts of Rules," *Philosophical Review*, 64: 3-32.
- . (1971) *A Theory of Justice*, Cambridge: Belknap Press of Harvard University.
- Raz, J. (1990) *Practical Reason and Norms*, 2nd edn, Oxford: Oxford University Press.
- Rescher, N. (1977) *Dialectics*, Albany: State University of New York Press.
- Searle, J. (1969) *Speech Acts*, Cambridge: Cambridge University Press.
- . (2001) *Rationality in Action*, Cambridge: MIT Press.
- Sellars, W. (1963) "Some Reflections on Language Games," rpt. in Sellars's *Science, Perception, and Reality*, New York: Routledge and Kegan Paul.
- Walton, D. and Krabbe, E. (1995) *Commitment in Dialogue*, Albany: State University

Press of New York.

Watson, G. (2004) "Asserting and Promising," *Philosophical Studies*, 117: 55-77.

Williams, B. (1997) "Internal and External Reasons," rpt. in S. Darwall, A. Gibbard, and P. Railton (eds.) *Moral Discourse and Practice*, Oxford: Oxford University Press.

---. (2002) *Truth and Truthfulness*, Princeton: Princeton University Press.

Williams, M. (1999) "Skepticism," J. Greco and E. Sosa (eds.) *The Blackwell Guide to Epistemology*, Malden: Blackwell.

Williamson, T. (2000) *Knowledge and its Limits*, Oxford: Oxford University Press.

Winch, P. (1972) "Nature and Convention," in *Ethics and Action*, London: Routledge & Kegan Paul Ltd.

Wright, C. (1992) *Truth and Objectivity*, Cambridge: Harvard University Press.