




# How to Read *How to Do Things with Words*: On Sbisà's Proof by Contradiction

Jeremy Wanderer<sup>1</sup> · Leo Townsend<sup>2</sup> 

Received: 26 April 2023 / Revised: 21 November 2023 / Accepted: 28 December 2023 /  
Published online: 23 January 2024  
© The Author(s) 2024

## Abstract

Midway through *How to Do Things With Words*, J.L. Austin's announces a "fresh start" in his efforts to characterize the ways in which speech is action, and introduces a new conceptual framework from the one he has been using up to that point. Against a common reading that portrays this move as simply abandoning the framework so far developed, Marina Sbisà contends that the text takes the argumentative form of a proof by contradiction, such that the initial framework plays an instrumental role as part of a proof in favour of the subsequent one. Despite agreeing with Sbisà's broad instrumentalist approach, we argue that her regimentation of Austin's narrative into a proof by contradiction ultimately fails - both as a proof and as an interpretation of Austin. Instead, we suggest that a better way of interpreting the peculiar structure of *How to Do Things With Words* is as a pedagogical exercise whose point is to bring a concealed alternative into view in a manner that also explains its initial concealment, and that this approach provides richer resources for supporting Sbisà's own conventionalist understanding of illocution than that afforded by reading the text as a proof by contradiction.

**Keywords** J.L. Austin · Performative utterances · Illocution · Marina Sbisà

## 1 Introduction

J.L. Austin's *How to Do Things with Words* (1975) [henceforth HTD] is one of the most widely cited texts of twentieth century Anglophone philosophy and yet it is notoriously challenging to read. A central hurdle facing the reader concerns the fact

---

✉ Leo Townsend  
leo.c.townsend@gmail.com

Jeremy Wanderer  
jeremy.wanderer@umb.edu

<sup>1</sup> Department of Philosophy, University of Massachusetts Boston, Boston, MA, USA

<sup>2</sup> Department of Philosophy, University of Reading, Reading, UK

that mid-way through the text Austin abandons his initial efforts at characterizing things done with words in terms of a bipartite distinction between performatives and constatives, making instead “a fresh start” (Austin, 1975, 91) in terms of a tripartite distinction between locutionary, illocutionary and perlocutionary acts. The hurdle for the reader concerns how best to understand this shift. Does starting afresh mean a complete abandonment of the explanatory framework provided in the first part, as suggested by the introduction of new distinctions, terms and canonical examples in the second part? Or does starting afresh involve a mere modification of aspects of the explanatory framework provided in the first part, as suggested by much continuity in the theoretical terminology between the two parts?

A dominant response in subsequent discussion takes the former route, treating the discussion in the second part of *HTD* as a wholesale replacement of the discussion in the first. This approach owes much to Strawson’s (1964) influential discussion of *HTD*, in which he argued for a distinction between those speech act kinds whose character and conditions of performance depend on the speaker’s intentions, and those that are individuated instead on the basis of the conventions governing them. Though Strawson’s distinction is not offered as an attempted exegesis of *HTD*, the fact that the discussion in the first part of *HTD* is centered around ceremonial speech acts that are more suited to a conventionalist treatment and the discussion in the second part is centered around communicative acts more suited to an intentionalist treatment led many to read this into the structure of *HTD* itself.<sup>1</sup> That is, it is common to treat the fresh start midway through *HTD* as a change of mind by Austin involving an abandonment of convention as the fundamental key to understanding speech as action.<sup>2</sup>

---

<sup>1</sup> We are here following what is a common division in the literature regarding the nature of the illocutionary force of an utterance, between so-called ‘intentionalists’ [including: (Strawson (1964); Bach and Harnish (1979); and Harris (2019))] who claim that it is the communicative intention of the speaker that determines illocutionary force, and ‘conventionalists’ [including: Searle (1969) and Lepore and Stone (2014)] who point to the satisfaction of relevant conventions instead. It is worth flagging that the notion of convention in this common set-up is understood ceremonially. i.e., as an institutionally established and socially maintained set of rules that link certain prescribed behaviors with achieving certain effects. An alternate, non-ceremonial, conception of convention, one favored by Sbisà, will be considered towards the end of the paper.

<sup>2</sup> As this formulation suggests, what we are calling the ‘dominant approach’ is one that treats the fresh start as heralding the demise of conventionalism, and not the concomitant claim that Austin’s subsequent discussions supports intentionalism instead. Indeed, the dominance of an intentionalist approach to illocution in recent discussions and its absence in *HTD* has contributed to the relative neglect of explicit discussion of our exegetical concerns in this literature. What we are treating as the dominant approach thus remains largely implicit. For the idea that the fresh start involves error or a change of mind, see, for example, Warnock (1989, 81), Black (1963, 217) and Ferguson (1969, 412). For the idea of banishing conventional acts to the periphery of the illocutionary domain – see, for example, Bach and Harnish (1979) and Harris et al. (2018, 2–3). We have explored this dominant response in more detail in Wanderer & Townsend (2023).

Some readers of Austin stand wary of this narrative of abandonment, and, of these, Marina Sbisà is among the most prominent.<sup>3</sup> In a series of rich and insightful papers, she sketches a detailed picture of Austinian speech acts that (a) provides an alternative conventionalist understanding of the nature of illocutionary force to the still dominant intentionalist approach, (b) argues for greater continuity between the two parts of HTD than is commonly understood, and (c) suggests an intriguing approach to understanding Austin's reasons for presenting the text in this challenging manner, viz. that HTD as whole is best read as a complex argument that takes the form of a proof by contradiction.<sup>4</sup>

We are generally sympathetic to both (a) and (b); our primary focus here is on Sbisà's suggestion in (c). In the discussion to follow, we argue that approaching the structure of HTD as a proof by contradiction is not a fruitful way of reading the text, and offer an alternative to (c) that better supports the position that Sbisà herself articulates in (a) and (b).

## 2 The proof by contradiction

According to Sbisà's proof by contradiction, the text of HTD as a whole functions as an extended argument in favor of the conclusion that "all speech should be considered as action, and, more specifically, that speech can be described as the performing of action of the same kind as those performed by means of performative utterances" (Sbisà, 2007, 462-3). The argument works by beginning with the opposing thesis, that not all of our utterances perform actions of the same kind as those performed by means of performative utterances, and it is by refuting this thesis we are given a proof (by contradiction) of the conclusion. According to this suggestion, then, the point of beginning where Austin does, with the bipartite division between performative and constative utterances, is to give support to the ultimate lesson of HTD, which is that all speech is action. As a result, a reader who simply begins reading HTD at the fresh start in lecture VII would miss out on this proof, as well as the insight it provides into the distinctive character of speech acts reached in Austin's conclusion.

Sbisà presents this proof by contradiction in schematic form:

"In HTD we have a Hypothesis (**P**) "Some utterances are performative", a Thesis (**A**) "All of our utterances perform actions of the same kind as those performed by performative utterances" and a proof by contradiction, which goes approximately like this:

<sup>3</sup> The paper to follow largely concerns itself with Sbisà's interpretation of Austin and its implications for understanding the nature of illocution. The importance of Sbisà's work is gaining overdue recognition within Anglophone philosophical circles, as reflected in the recent publication of Sbisà's collected papers (Sbisà, 2023) and the appearance of a collection of critical essays on her work (Caponetto & Labinaz, 2023).

<sup>4</sup> The proof by contradiction is found in Sbisà (2007) and reprinted in Sbisà (2023: 181–194). References below are to the original text.

Suppose that *A* does not hold (*not A*): since some utterances are performative (*P*), it should be possible to distinguish them sharply from constative utterances (if *P and not A*, then *P/C*);  
 But the performative/constative distinction is flawed, and therefore *not A* does not hold;  
 So the supposition that *A* does not hold is false, and therefore, *A* holds.”  
 (Sbisà, 2007, 463, emphasis in original).

Sbisà describes this schema as “approximate”. Here is a plausible initial reconstruction of the argument with commentary that provides some of the working knowledge of HTD assumed in the schema:

Premise 1: Some utterances are performative [*P*] & not all utterances perform actions of the same kind as those performed by means of performatives [*~A*].

This premise is simply the conjunction of what Sbisà calls the hypothesis of the argument (*Some utterances are performative*) and the starting assumption, which is negation of the conclusion to be derived by the proof by contradiction (*Not all utterances perform actions of the same kind as those performed by means of performatives*).

When Sbisà says that some utterances are performative she means that there is a class of utterances that can be called *performative utterances*. These are “utterances” in the sense of *utterata*, i.e., things uttered, as opposed to utterings (which would be the *utteratio*, not the *utteratum*).<sup>5</sup> Austin’s early examples of performative utterances include the sentences “I do” (as issued in the course of a marriage ceremony) and “I name this ship the *Queen Elizabeth*” (as uttered when smashing the bottle against the stem of the boat). As these examples suggest, such utterances (i.e., the sentences “I do” and “I name this ship the *Queen Elizabeth*”), are the *means* for performing certain actions: when these utterances are issued, they are *used* to perform the actions of marrying or ship-naming. To put it another way, the issuance of these utterances in the relevant context *is* the performance of certain actions (namely, marrying and ship-naming). This helps to better understand the somewhat inelegant phrasing in the second part of Premise 1, specifically the mention of utterances “performing actions”, and actions “performed by means of performative utterances.” To say that utterances perform actions is to say that utterances, when issued, are actions of a certain sort, and similarly, the actions “performed by means of performatives” are those actions performed when performative utterances are issued – as exemplified by the actions of marrying and ship-naming. These actions are not themselves “performatives” but are actions that are performed by the use or issuance of performative utterances.

<sup>5</sup> Austin (1975, 92n1) himself makes this distinction, clarifying that his use of “utterance” is to be read as meaning *utteratum*, and that to refer to an *utteratio* he uses “the issuing of an utterance”. Though this clarification comes only in Lecture VII, it is clearly meant to apply to his early references to “performative utterances” and “constative utterances”, not least because the first label he proffers for the former is “performative sentences” (Austin, 1975, 6).

As is well known, the notion of a performative utterance is introduced by Austin into HTD not only by way of examples, but also by way of contrast with another familiar class of utterance: constative utterances. Constative utterances and performative utterances differ in function (i.e., what they are used for) and standards of assessment. The function of a constative utterance is to say that something is or is not the case, and so issuances of constative utterances are to be assessed solely in terms of truth or falsity. The function of a performative utterance is to do something in saying something, and issuances of performative utterances are to be assessed as to whether or not they felicitously meet the conventional conditions and procedures for exemplary performance. This explains the “not all” in the second part of Premise 1: given their different function and standards of assessment, *constative utterances* do not perform actions of the same kind as those performed by performative utterances. Premise 1 thus rehearses the opening claims of HTD that [P] some utterances are performatives and [ $\sim$ A] not all utterances perform actions of the same kind as performatives – namely those utterances that have the function of constatives.<sup>6</sup>

Premise 2: [P& $\sim$ A] entails that it should be possible to sharply distinguish utterances that are performative from utterances that are constative [P/C].

Our commentary on Premise 1 introduced the term “constative” into the discussion, and this is justified by its explicitly featuring in Premise 2. The claim here is that positing that some utterances are performative and that some utterances are constative entails that it should be possible to “sharply distinguish” one from the other [P/C]. Sbisà’s language here echoes that of Austin, who – having introduced the distinction – hankers after “some precise way in which we can definitely distinguish the performative from the constative utterance” (Austin, 1975, 55) in light of the realization that considerations of the type of truth and falsity “infect” many utterances that seem to be paradigm performatives, and considerations of felicity may infect utterances that seem to be paradigm constatives. It thus seems that the desideratum established in Premise 2 (a “sharp” distinction) is that which Austin himself considers: some kind of grammatical or lexicographical feature of an utterance (again, where the utterance is to be thought of as an *utteratum* and not an *utteratio*) that allows us (theorists) to discern whether or not that utterance, when issued, performs an action of the same kind as the actions performed by the issuance of performative utterances.

Premise 3: it is not possible to sharply distinguish utterances that are performative from utterances that are constative. [ $\sim$ P/C]

Throughout much of lectures V and VI of HTD, Austin considers a range of grammatical or lexicographic features of utterances that are possible candidates for providing the distinguishing mark of the performative required by [P/C]. These include the possibility that the utterance – or the utterance when transformed into an explicit form – contains some special vocabulary (such as the inclusion of special

<sup>6</sup> We consider an ambiguity that remains in this premise in the next section.

“performative words”) and/or a particular mood or tense (such as deploying verbs in the first person singular present indicative active) that could help isolate the performative. After considering this broad array of possibilities, Austin abandons the attempt to isolate the performative in this manner, concluding that “it is often not easy to be sure that, even when it is apparently in explicit form, an utterance is performative or that it is not” (Austin, 1975, 91). Premise 3 [ $\sim$ P/C], then, is the conclusion reached by Austin just before he starts afresh.

Therefore: All utterances perform actions of the same kind as those actions performed by means of performatives. [A].

[A] is the basis for Austin’s fresh start. Constatives are revealed to have “no unique position” (Austin, 1975, 148) in terms of affording insight into the character of language over that afforded by performatives, since it is shown that all utterances share the features we have taken to be characteristic of performatives. We have arrived at [A] via a proof by contradiction. Specifically, the failure to sharply distinguish utterances that are performative from utterances that are constative [ $\sim$ P/C] provides a refutation of the opposing thesis [ $\sim$ A], thereby proving [A] by contradiction.

[A] is not enough to undermine the narrative of abandonment alone, though it does suggest a continuity between the two parts of HTD that runs counter to its central motif. Arriving at [A] in the manner of a proof by contradiction furthers this conclusion, for it shows that Austin’s fresh start does not correspond to a change of mind; rather, the first part of HTD is a deliberate attempt by its author to support the picture of speech arrived at in its second part. If successful, then, Sbisà’s suggested reading of the structure of HTD as a proof by contradiction goes a long way towards helping the struggling reader make sense of the peculiar structure of this challenging text.

### 3 Problems with the proof

Sbisà’s overarching concern, like our own, is to proffer an account of how to read *How to Do Things with Words*. Such an account is needed both because of the peculiar structure of the text of HTD itself noted above and because Austin himself gives his audience little overt guidance as how best to proceed in engaging with this structure. In suggesting that the text be read as a proof by contradiction, Sbisà is going beyond anything mentioned explicitly in HTD itself. The plausibility of her suggestion is thus best evaluated in terms of charitability: Does treating the text as providing a proof by contradiction deliver a way of reading the text that both renders it compelling and that coheres with its structure and fine detail? Our goal in this section is to raise concerns about the charitability of Sbisà’s suggestion.

To bring these concerns into view, let us first explore an unclarity in the argument just reconstructed. Consider again the first premise: *Some utterances are performative [P] & not all utterances perform actions of the same kind as those performed by means of performatives [ $\sim$ A]*. As it stands, there is an ambiguity in the formulation of [ $\sim$ A]: in saying that not all utterances perform actions of the same kind

as those performed by means of performatives, are we assuming that the relevant *non-performative* utterances (i.e., constatives) perform actions of a *different kind* to those performed by performatives, or are we assuming that they do not perform actions at all?

Although the text of HTD itself is not fully clear on this, it seems that latter interpretation of [ $\sim A$ ] fits more comfortably with the early moves made by Austin in setting up the performative-constative contrast. Applying the term “performative” to an utterance, we are told, “indicates that the issuing of the utterance is the performing of an action—it is not normally thought of as just saying something.” (Austin, 1975, 7). So, “just saying something”, which is issuing a constative utterance, is to be contrasted with the performing of an action. Thus, the purported differences between performatives and constatives in terms of function and criteria of assessment, and the search for their distinctive grammatical and lexicographic hallmarks, are part of an attempt to distinguish between those utterances whose issuance is an action and those whose issuance is just saying something.<sup>7</sup> So, in light of this reading, Premise 1 should be read as: *the issuing of some [P], but not all [ $\sim A$ ], utterances is the performing of an action.*

This reading of Premise 1 makes better sense of the entailment relation to Premise 2 than the alternative reading. According to the alternative reading, the first premise claims that there are actions performed by the issuance of constatives, and there are actions performed by the issuance of performatives, and these are two different kinds of actions. The distinction we are after in [P/C] on this reading is thus between two action-kinds (the kind of action performed by the use of constatives, and the kind performed by the use of performatives) and it is hard to see why a distinction of this kind must be sharp. First, it is often the case that the precise contours of an action-kind are not precisely drawn but may contain a cluster of overlapping features without the need for delineating any clear edge. Second, if the goal is to distinguish action kinds, it is somewhat strange that Austin devotes so much effort to seeking out grammatical and lexicographic hallmarks of performative utterances, since these are features that stand independent of thinking of performative in terms of the actions they are used to perform. The earlier reading of Premise 1 fares better on both scores. First, the distinction between an utterance that, when issued, performs an action and one that does not is a categorial one, for which there should in principle be a sharp distinction, even if it may not be clear how to apply the distinction to a token utterance. Second, since we are focused on distinguishing

<sup>7</sup> It was not, of course, lost on Austin that to say something is to perform a certain kind of action. Yet he clearly wished to maintain a sense of “action” – of things done with words – that could be meaningfully contrasted with (just) saying something. This is evident in many of Austin’s remarks throughout the earlier part of HTD. For example, just before the fresh start, Austin attempts to respond to the worry that any time “we issue any utterance whatsoever, are we not ‘doing something’?” In response, he notes that “the ways in which we talk about ‘action’ are liable here, as elsewhere, to be confusing. For example, we may contrast men of words with men of action, we may say they did nothing, only talked or said things: yet again, we may contrast only thinking something with actually saying it (out loud), in which context saying it is doing something” (Austin, 1975, 92). It is thus clear that, for Austin, not every instance of issuing an utterance (not every *utteratio*) must *ipso facto* be treated as an action in the sense of action at issue.

an utterance whose issuance performs an action from one that does not, it makes sense to focus on those features of an issued utterance, such as grammar or lexicography, that are available even when the utterance is not considered to be an action.

Suppose that however much we turn over the issued utterance we cannot find the mark of a performative (i.e.  $[\sim P/C]$ ). According to the proof by contradiction, [A] follows: *All utterances perform actions of the same kind as those performed by performatives*. On the interpretation just given, this means that all utterances, when issued, perform actions. So, what we thought was a distinct category of utterance (performatives) added on to a more basic category (constatives) turns out to subsume all utterances including the latter. As a result, it emerges that no utterance-kind is more basic; what there is the “total speech act in the total speech situation” (Austin, 1975, 147), out of which we can abstract distinct action-kinds as needed. We thus have a plausible reading of the proof by contradiction that both follows the narrative structure of HTD and arrives at a conclusion [A] that is indeed the point of departure for Austin’s fresh start.

Our reservations with Sbisà’s reading of HTD are not with this interpretation of conclusion [A], but rather with her suggestion that [A] is reached in HTD via a *proof*. Interpreting the peculiar structure of HTD along the lines of a proof by contradiction raises the bar for each stage of the narrative beyond that which it can clear.

As a first hurdle, note that even when we grant all three premises it is doubtful that the conclusion follows. If  $[P \& \sim A]$  entails  $[P/C]$ , then establishing  $[\sim P/C]$  shows that one must deny  $[P \& \sim A]$ . But this still leaves open whether it should be  $[P]$  or  $[\sim A]$  that should be abandoned, or indeed whether both should be abandoned. So, as we have formulated the argument, the conclusion could either be  $[\sim P]$  *No utterances are performative* AND/OR [A] *All utterances perform actions of the same kind as those performed by means of performatives*. In her formulation, Sbisà forestalls the possibility of concluding that no utterances are performative by calling  $[P]$  a “hypothesis,” a status that serves to shield it from the arrow of modus tollens. But it is not clear what motivates this special protection. Without any special justification, it seems as though, given the role of  $[P]$  in the rest of the argument, it should be equally vulnerable to being overturned when the consequent in Premise 2 is shown to be false.

A second hurdle concerns the relation between the assumption made for the purposes of the proof by contradiction  $[\sim A]$  and the proof’s conclusion [A]. In order for the proof to work, these must be directly contradictory, and this requires that the meaning of key terms must be consistent across both (i.e., there should be no equivocation). However, our discussion has revealed that the term “performative” has a different meaning in the starting assumption  $[\sim A]$  as compared to the conclusion [A]. To see this, consider Austin’s “preliminary isolation of the performative” which explicitly defines the notion of the performative utterance in contrast to that of the constative utterance:

“[performative utterances] do not ‘describe’ or ‘report’ or constatae anything at all [...] [their being uttered] is, or is a part of, the doing of an action, which again would not normally be described as, or as ‘just’, saying something” (Austin, 1975, 5).



Insofar as Austin can plausibly be characterized as starting with the assumption that not all utterances perform actions of the kind performed by means of performative utterances, the relevant notion of *performative utterance* should be understood in the way that he initially construes it. But this initial construal expressly excludes the possibility that the things done by the issuance of performative utterances are of a kind with what is done by means of constatives. So, the conclusion Sbisà attributes to Austin - that all utterances (*including constatives*) perform actions of the same as those performed by means of performative utterances – must be invoking a different sense of “performative utterance” from the sense with which Austin begins. And this means that the proof by contradiction Sbisà ascribes to Austin is not strictly valid, since there is equivocation between a key term featured in the assumed premise and the conclusion, with the result that these are not directly contradictory.

A final hurdle is exegetical: the tentativeness with which Austin concludes his attempt to isolate the performative (“it is often not easy to be sure ...”) falls short of the confidence required by [ $\sim$ P/C] as employed in the course of a proof. Indeed, Austin’s modest characterization of his overall achievement in the first part (“... I could do no more than explode a few hopeful fireworks” (Austin, 1975, 148) does not sit easily with the contention that the argument has provided a proof.

These three hurdles stand in the way of casting Austin’s narrative as a proof by contradiction. They also point towards an alternative.

#### 4 The pedagogical reading

In arranging Austin’s narrative as a proof, the link between each stage is provided by the content of the claim made. An alternative is to focus on the activity undertaken by the reader in the course of making the claim. Put differently, an alternative is to treat the text as a *pedagogical exercise* in which the reader is invited to render intuitively compelling claims about speech into explicit form, and the very attempt at doing this leads the reader to reframe their initial understanding. Let us call this a ‘pedagogical reading’ of HTD.

Like Sbisà’s proof, the pedagogical reading treats the abandonment of the binary distinction between constatives and performatives in the first part of HTD as teaching the reader something important with regards the ternary distinction that emerges. They differ in how this lesson is taught. For Sbisà, the lessons to be learned by the reader take an argumentative form: its goal is to persuade the reader of the veracity of a conclusion through reasoned argumentation in its support. In this form of learning, both opening claim and its contrary are available and intelligible to the reader at the outset of the argument, and we are led from one to other by a series of steps that constitute a proof. Not all philosophical learning takes an argumentative form. What we are here calling a pedagogical exercise has the goal of aiding the reader see a hitherto unnoticed theoretical possibility by undertaking a guided process designed to dispel probable impediments that obscure it from coming into view. In this form of learning, the desired conclusion is precisely not antecedently intelligible to the reader; if it were, there would be no need to undertake the practical exercise at

all.<sup>8</sup> The purpose of starting with the binary distinction according to the pedagogical reading is thus not for it to function as a premise in a proof concluding that all utterances perform actions of the same kind as those actions performed by means of performatives, but as part of a practical exercise needed to undermine a pervasive orientation to speech that prevents this conclusion being available as a possibility to be considered at all.

The first part of HTD also has three stages on the pedagogical reading.<sup>9</sup> The first stage in the narrative attempts to capture a pervasive intuition about language, viz. that much of what we think to be central when we think about language, such as its describing state of affairs, can be captured by thinking about sentences alone and their propositional content, independent of their actually being issued in a particular context.<sup>10</sup> Given this intuition, the act of issuing an utterance in context (to an audience, on an occasion, with a purpose) is not treated as central to its linguistic significance; it is “just saying something” whose importance for explicating the phenomena of language does not lie in treating it as the performing of an action. Put differently, the intuition is that, at least as far as central uses of language go, the linguistic significance of an *utteratio* (the issuing of an utterance) is exhausted by the *utteratum* (the thing uttered). Guided by this intuition, utterances for which the significance of their issuance does lie in recognizing them as the means to perform actions (performative utterances) thus appear as different from, and less basic to understanding speech itself, than cases where the issuing of the utterance is “saying something” (constative utterances).

In the second stage, the exercise of searching for sharp criteria to distinguish between performative and constative does not just end empty-handed, but also reveals in practice the paucity of this way of thinking about speech to start off with.<sup>11</sup> This is because thinking of utterances in terms of grammar and lexicography independent of the significance of their issuance in a particular context is revealed by the search as an impoverished way of thinking about what we understand to be central to the linguistic significance of speech. What had been taken for granted by the reader, the foundational standing of the *utteratum* over the *utteratio* as the most basic feature of language, now comes into view through this exercise, and increasingly appears as a matter of artifice.<sup>12</sup> The exercise of searching itself thus sets the stage for the third stage of the narrative in which grip of the pervasive intuition is

<sup>8</sup> Our characterization of a pedagogical exercise here is narrower than what is perhaps a more ordinary sense of “pedagogy” that incorporates any process of teaching and learning. We concede that even a proof by contradiction could fall under the banner of pedagogy in this broader sense, in that – unlike dominant approaches – it treats the first part of HTD as teaching the reader something important. Thanks to an anonymous reviewer for pointing this out.

<sup>9</sup> These three stages closely follow the general structure of HTD: Stage 1 takes up Lectures I–III; Stage 2 takes up Lectures IV–VII; Stage 3 begins in earnest with Lecture VIII.

<sup>10</sup> This is an aspect of the ‘descriptive fallacy’ mentioned in (Austin, 1975: 3, 100).

<sup>11</sup> “[I]t seemed that we were going to find it not always easy to distinguish performative utterances from constative, and it therefore seemed expedient to go farther back for a while to fundamentals”. (Austin, 1975, 94).

<sup>12</sup> Cf. (Austin, 1975, 147) for talk of ‘mere abstractions’ or (Austin, 1975, 149) for “artificial abstractions”.

broken by having an alternative come into view, one in which the object of what we study when trying to make sense of language is action: “[t]he total speech act in the total speech situation is the only actual phenomenon which, in the last resort, we are engaged in elucidating” (Austin, 1975, 148). The very idea of an utteratum, much as the idea of an illocutionary speech act, is an ‘abstraction’ from this whole.<sup>13</sup> The three stages in this pedagogical reading are not premises in a proof but are moments in a practical exercise needed to expunge a pervasive intuition about language.

The goal of HTD’s narrative on this reading is not to prove a conclusion but to bring a concealed alternative into view in a manner that also explains its initial concealment. The taking for granted of certain assumptions and the changing meaning of key terms that were challenges facing the attempt to treat the narrative as a proof emerge as expected features of such a pedagogical exercise. It operates not through the mandatory persuasiveness of a proof, but through the broader view of the terrain afforded by the light cast by a “few hopeful fireworks”. Thus, the three hurdles that got in the way of regimenting Austin’s narrative into a proof by contradiction are cast aside by the pedagogical reading. We contend that the pedagogical reading is thus a more charitable reading of HTD than that suggested by Sbisà: while not explicitly flagged in the text itself, it delivers a more compelling and exegetically coherent way of approaching the challenges of reading HTD than does treating it as a proof by contradiction.<sup>14</sup>

Let us take stock. A striking feature of HTD is that the initial route taken to account for how we do things with words, based on a binary distinction between constatives and performatives, is abandoned mid-way. One common reading of the text treats this original route as having led to a dead-end; despite initial promise, the journey proved to be unsuccessful, and the reader is asked to begin afresh. On this reading, little is missed by a reader not venturing down the path provided by the initial route at all. A second reading of the text is provided by Sbisà, according to which the initial detour through the binary distinction paves a secure path for arriving at the final destination by way of proof. On this reading, the detour is an unmissable feature of the philosophical route to the conclusion. Exegetical and conceptual difficulties with Sbisà’s proof by contradiction have led us to champion a third reading. According to the pedagogical reading, the function of the initial route is to break open a path to another route that is likely to have been obscured by prior

<sup>13</sup> “Furthermore, in general the locutionary act as much as the illocutionary is an abstraction only: every genuine speech act is both” (Austin, 1975, 147).

<sup>14</sup> Earlier we noted that, despite the obvious challenges facing the reader of HTD, Austin himself gives no explicit guidance to the reader with regards to how to read ‘How to Do Things with Words’. Our pedagogical reading – much like Sbisà’s reading – is an attempt to offer this guidance. Our goal is not to discern Austin’s own intentions in this regard: circumstances surrounding the genesis of the text (edited and published posthumously based on lectures and notes) make it hard to arrive at a definitive conclusion about these intentions, and it is possible that Austin himself had no clear or single intention on the matter. Rather, our target is the published version of the text – specifically, the 2nd edition (Austin, 1975) – and our aim is to provide an answer to the question of how to read HTD that delivers the most charitable reading of this text. Given these aims, our central claim is that our pedagogical reading is a more charitable reading of HTD than that proffered by Sbisà.

philosophical baggage. On this reading, taking the initial route is necessary for most readers who simply could not arrive at the final destination without it.

## 5 In support of conventionalism

In rejecting Sbisà's reading of HTD as delivering a proof by contradiction, we are not thereby rejecting her conventionalist understanding of the book's central lesson. Far from it. We think that the pedagogical reading of the text as outlined provides a firmer foundation for Sbisà's conventionalism than that afforded by reading HTD as a proof by contradiction.

Sbisà takes illocutionary acts to be conventional.<sup>15</sup> By conventional, she does not mean that they must be performed through conventional *means*, but that they bring about conventional *effects*. A speech act that is performed through conventional *means* is an act for which the success of its performance on an occasion depends on some institutionally established and socially maintained set of rules that link certain prescribed behaviors with achieving certain effects. A speech act that brings about conventional *effects* is one whose successful performance alters the deontic properties (such as being entitled and/or obligated to do something) of the participants in a discursive interaction. Unlike the effects that emerge from a chain of causes that cannot be retracted or annulled once they have obtained, conventional effects depend on intersubjective agreement for their existence, and thus can be retracted or annulled if found to be inappropriate in context. This intersubjective agreement can be tacit and operate by default or be explicit through negotiation between relevant participants whose normative standing is altered, including that which results from the conversational context in which the acts take place. According to Sbisà's conventionalism, *all* illocutionary acts – ranging from ceremonial acts like ship-naming to conversational acts like promising – are such as to “take effect” through the imposition of altered normative statuses on the parties involved. The character and success of a given illocutionary act thus depends on the dynamic interaction between interlocutors, rather than just the communicative intention of the speaker (and the grasp of this by the hearer) that is highlighted in intentionalist understandings of illocution.

That illocutionary acts are conventional in the sense of bringing about conventional effects is a central lesson that Sbisà draws from HTD itself. An illocutionary act, Austin insists,

“takes effect” in certain ways, as distinguished from producing consequences in the sense of bringing about states of affairs in the “normal” way, i.e. changes in the natural course of events.... (Austin, 1975, 117).

Conventional effects are thus not the product of natural causation; rather, they correspond to a distinctive way that the illocutionary act “takes effect”, that is, by making certain subsequent acts “out of order” (and presumably making others “in order”). For example, Austin points out that:

<sup>15</sup> This summary draws heavily on the essays collected in Sbisà (2023), especially essays 7, 10 and 11.

“I name this ship the *Queen Elizabeth*” has the effect of naming or christening the ship; then subsequent acts such as referring to it as the *Generalissimo Stalin* will be out of order’ (Austin, 1975, 117).

The role of conventional effects is mentioned by Austin in both parts of HTD (i.e., both before and after the “fresh start”): they are first mentioned in connection with the ways that performative utterances can be infelicitous,<sup>16</sup> and later invoked as one the ways that illocutionary acts are connected with certain kinds of effects.<sup>17</sup> Sbisà’s conventionalism is thus a development of, and draws some of its authority from, these Austinian claims.

There is little doubt that when the notion of convention is introduced in making the performative / constative distinction in the first part of HTD it is the sense of conventionality of means that dominates. This is apparent in the ceremonial character of the examples of performatives (such as marrying and ship-naming) provided and made explicit through talk of “accepted conventional procedure” in describing them.<sup>18</sup> If Sbisà is right that *conventional effects* are the key to understanding HTD, why does Austin begin the text with a discussion that highlights *conventionality of means*?

Sbisà’s proof by contradiction reading of HTD struggles to answer this question. The proof lays out two contradictory theses: that some but not all utterances perform actions of the kind performed by means of performatives, and that all utterances perform actions of the kind performed by means of performatives. We can now see that the notion of the relevant kind of action (viz., the kind of action performed by means of performatives) differs between the two. At the outset, the relevant idea of action is characterized via the idea of conventionality of means. By the end, however, the relevant idea of an action performed by means of an utterance is construed in terms of conventionality of effects. The falsity of the first thesis, now understood as the claim that some but not all utterances perform actions that operate via conventional means, does not therefore prove the truth of the second thesis, now understood as the claim that all utterances perform actions having conventional effects.<sup>19</sup> This means that the introduction of a ceremonial context in the first premise does little to advance the conclusion and only provokes in the unsavvy reader an overly narrowed conception of convention.

In contrast, the pedagogical reading of HTD can explain why Austin starts with an emphasis on conventionality of means, even if his ultimate purpose is to highlight conventionality of effects. We have seen that the starting point of the pedagogical

<sup>16</sup> Rule A1 of the “doctrine of the infelicities” states, “There must exist an accepted conventional procedure having a conventional effect...” (Austin, 1975, 14).

<sup>17</sup> Austin (1975, 116–7) claims there are three special kinds of effect that illocutionary acts are bound up with: uptake, conventional effects, and illocutionary sequels.

<sup>18</sup> Such as in the first rule of the ‘Doctrine of Infelicities’ in Austin (1975: 14), especially when understood alongside the examples.

<sup>19</sup> This challenge differs from the second hurdle noted in Section 2. Here the equivocation can be solved by substituting ‘effects’ for ‘means’ in the proof. The challenge here is that the approach cannot explain why the need for this substitution should arise in the first place.

reading is the intuition that the linguistic significance of an utterance in central cases is exhausted by what is uttered (the *utteratum*). Performative utterances, once our attention is drawn to them, do little to unsettle this intuition, since they appear as specialized utterances whose significance is dependent on extra-linguistic ceremonial institutions. Since the notion of ceremonial institutions is not itself internal to the very idea of language, bringing performative utterances into view in this way suggests not only that they are “special” but also that they have a secondary standing in terms of affording insight into language as compared to the constative. Once the failed exercise of finding a sharp distinction between constatives and performatives reveals the paucity of this starting point, the possibility of an alternative explanatory framework for thinking about how to do things with words comes into view. More specifically, the idea of performances having conventional effects is viewed as internal to the very idea of speech as an activity, especially in the illocutionary cases that are Austin’s main focus, including the ceremonial cases with which we began.

We have seen that the meaning of the key term ‘convention’ as it relates to action changes between the two parts of HTD, from being understood primarily in terms of conventional means to conventional effects. While changes in the meaning of key terms as one progresses is fatal in an argumentative structure, it provides the very lifeblood of a pedagogical encounter. The encounter does not take place in abstraction. Its starting point is the understanding of a phenomenon tacitly held and it aims to develop this understanding through a process of explication, challenge and refinement, so that by the end we arrive at an enriched understanding of the relevant terrain, including the pedagogical journey itself. Treating HTD as a pedagogical exercise involving an evolving conception of convention as it relates to speech action better serves Sbisà’s own conventionalist treatment of illocution than treating it as a proof by contradiction.

## 6 Conclusion

Austin’s “fresh start” midway through HTD has frustrated many readers. The dominant reading of HTD, often motivated by a broader rejection of conventionalism about illocution in favour of intentionalism, portrays this fresh start as an abandonment of the initial explanatory framework. In contrast, we have here followed Sbisà’s lead in claiming that the initial framework is better understood as playing an instrumental role in the text and in using HTD as the source for a more viable version of conventionalism (invoking conventional effects, not means). We differ from Sbisà in the way we understand the instrumental role of the initial explanatory framework. For Sbisà, this takes the argumentative form of a proof by contradiction: the first part of HTD involves an assumption, (that not all utterances perform actions), and it is by noticing that this assumption implies something false (that we can sharply distinguish those utterances that do perform actions from those utterances that don’t), that we arrive at the central insight of the second part of HTD (that all utterances perform actions). On our account, in contrast, the instrumental role takes the form of a pedagogical exercise whose point is to bring a concealed alternative into view in a manner that also explains its initial concealment. We have argued that regimentation

of Austin's narrative into a proof by contradiction ultimately fails, both as a proof and as a charitable interpretation of Austin, and that the pedagogical alternative provides both a better reading of the text and a firmer foundation for Sbisà's conventionalism than that afforded by reading HTD as a proof by contradiction.

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Austin, J. L. (1975). *How to do things with words*. Oxford University Press.
- Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. MIT Press.
- Black, M. (1963). Austin on Performatives. *Philosophy*, 38, 217–263.
- Caponetto, L., & Labinaz, P. (Eds.). (2023). *Sbisà on speech as action*. Springer Nature.
- Forgonson, L. W. (1969). In pursuit of performatives. In K. T. Fann (Ed.), *Symposium on J.L. Austin*. Routledge.
- Harris, D. W. (2019). Intention and commitment in speech acts. *Theoretical Linguistics*, 45(1–2), 53–67.
- Harris, D. W., Fogal, D., & Moss, M. (2018). Speech acts. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New work on speech acts*. Oxford University Press.
- Lepore, E., & Stone, M. (2014). *Imagination and convention: Distinguishing grammar and inference in language*. OUP Oxford.
- Sbisà, M. (2007). How to read Austin. *Pragmatics*, 17(3), 461–473.
- Sbisà, M. (2023). *Essays on speech acts and other topics in pragmatics*. Oxford University Press.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge University Press.
- Strawson, P. F. (1964). Intention and convention in speech acts. *The philosophical review*, 73(4), 439–460.
- Wanderer, J. & Townsend, L. (2023). Illocution by example. *Synthese*, 202(7), 1–22.
- Warnock, G. J. (1989). *J.L. Austin*. Routledge.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.