Chapter 5

Rule-Following and Failed Execution: “In the Penal Colony”

If the *Philosophical Investigations* ended after §43, Wittgenstein’s brevity would usher in both good news and bad. The good news: Wittgenstein would finally have presented us with a theory of language with no caveats. With the meaning of language now definitively determined through its use, the picture theory of the *Tractatus* would have been discarded and replaced with this viable new theory. The bad news would be that Wittgenstein would be proven a hypocrite, as his insistence that this work does not advance philosophical theses would have been soundly disproven by a philosophical thesis. Fortunately for Wittgenstein’s cohesion, but unfortunately for our desire for a theory of language that works, the paradox of ostensive definition is but the first of the *Investigations*’ many such revelations of the delusions under which language users continually labor. In fact, its unmasking leads directly into a new and more complex unmasking: one that effectively dismantles §43 soon after introducing it. This comes in §201, the culminating remark in what philosophers commonly call the paradox of rule-following. The following chapter will bring this second major paradox together with one of Kafka’s most paradoxical stories, “In der Strafkolonie” (“In the Penal Colony”). In the course of the chapter, I hope to show that this story is not paradoxical in the way we assume it is—that is, because its torture machine fails to do the only thing it is supposed to do—but in an entirely new way that we can only understand in light of Wittgenstein’s rule-following remarks.

This chapter is also an exploration of Kafka’s powerful and resonant failure as a prose narrator—at least when it comes to one major work. The point of departure for a meditation on Kafka as a failure is, undoubtedly, Benjamin’s statement that the “beauty” and “purity” of his work is “die von einem Gescheiterten” (“that of a failure”). Here we will not concern ourselves as much with *why* his work exemplifies the beauty and purity of failure, but rather *how* it accomplishes such a task, and how once again Wittgenstein’s philosophy fits in with this dubious accomplishment.
Narration, Description, and “Murder, Pure and Simple”

Until its penultimate scene, what we have assumed to be the previous narrative progression of “In the Penal Colony” leads us to believe that its “eigentümlicher Apparat” (“remarkable apparatus”), whose adulatory officer opens the story by observing it with a “gewissermaßen bewundernden Blick” (“certain look of wonder”) despite how “doch wohlbekannt” (“nevertheless well known”) it is, works in a certain way. Instead, it apparently malfunctions and impales its greatest adherent in what the visiting explorer terms “unmittelbarer Mord,” murder pure and simple (GW 1:161, 192). Critical interpretation of “In the Penal Colony,” and especially this act, has purported it to mean many things: the terminal existential guilt of man and the struggles with creation of the “true self” (Sokel); Kafka’s own desire to purge himself and his writing of late-Hapsburg decadence and degeneracy (Anderson); cultural hybridity and the violence of the colonized body (Goebel); the writing or artistic-creation process in general (Danielle Allen, Andreas Gailus, Richard Jayne); Kafka’s specific writing process with respect to a single work (Corngold); violent metaphors for the onset of modern technology (Kittler); the act of reading (Koelb); and, in Heinz Politzer’s classic *Parable and Paradox*, no less than “the inevitability of fate” itself. Or, as Anderson describes the situation using a reference to the visiting explorer (Forschungsreisende) and one of his earliest gestures in the face of the colony’s punishing sun, the story has understandably provoked a series of allegorical readings which, like a shielding hand over the eyes or an averted gaze, shifted attention away from the first and literal level of the story, away from its troubling surfaces, in an attempt to get behind or beyond this level and understand its “deeper” significance.

The investigation that follows is a journey filled with denial and disappointment, foremost but not solely for those seeking to understand the story’s “deeper” significance. For just as Gregor’s metaphorical collapse upon himself in *The Metamorphosis* transformed our conceptions of static metaphorical meaning and solid metaphorical structure, so will the theory of meaning in use, also portrayed in the world of Kafka’s prose narration, ultimately fail. This will take place in Kafka’s penal colony, in the form of a fascinating conflation of Kafka’s exquisitely failed narrative execution of the officer’s failed literal execution:

Die Egge schrieb nicht, sie stach nur, und das Bett wälzte den Körper nicht, sondern hob ihn nur zitternd in die Nadeln hinein. Der Reisende wollte eingreifen, möglicherweise das Ganze zum Stehen bringen, das war ja keine Folter, wie sie der Offizier erreichen wollte, das war unmittelbarer Mord. (1:192)
The Harrow didn’t write at all, but rather just stabbed, and the Bed
did not turn his body over, but rather just lifted it, shaking, up closer
to the needles. The explorer wanted to reach in, possibly to bring the
whole thing to a stop—this was not the torture the officer so wanted
to bring about; it was murder, pure and simple.

As many astute voices in literary theory have pointed out, among them
Klaus Scherpe in a highly convincing reexamination of Georg Lukács’s theo-
ries of modernism and critical realism, in writing the story’s climax the way
he has, Kafka has triumphantly destabilized the very notion of prose narration
itself, and any meaning we may have mistakenly attempted to derive from it
(or more specifically, as Scherpe writes, any “sense”: “Eine Erzählung, so sagt
man, macht Sinn und dessen sind wir bedürftig” [“A narration, so we say,
makes sense and this is something we need”]). In his “Penal Colony,” however,
Kafka sabotages any chance of an objective progression of a narration—the
presence of which, Lukács has famously argued, makes Thomas Mann’s work
a triumphant work of critical realism, in comparison to Kafka’s decadent
modernism, which contains its own highly unreliable subjectivity. Scherpe
points out that Kafka’s alleged work of prose narration actually fails to unfold
a traditional narrative progression, this being due to the primacy of descrip-
tion, which Scherpe contrasts with the progress of narration as such:

[Die Beschreibung] macht in besonderer Weise aufmerksam auf die
Dinge, die Natur, die Landschaften, die Menschen, auch das Nicht-
Menschliche, das Anorganische. Auch auf Geschichte? Nicht auf
Geschichte im Fortschritt der Zeit, sondern im Räumlichen von
Flachen und Figuren.

[Description] makes us aware, in a particular way, of things—nature,
landscape, humans, also the non-human, the inorganic. Also aware
of (hi)story? Not of history in the sense of the progress of time, but
rather in the sense of space, of planes and forms.

Whereas a successful (or, at any rate, realistic) narrative places events in
a sequence with the express purpose of allowing the reader to make sense of
them, description creates a story (Geschichte) that is not necessarily a history
(also Geschichte), in that it does not make an authoritative claim to objective
time or reality, nor to any sort of explanation. Whereas narration concerns
itself with such preoccupations as “chronology and causality,” description
“studies the things.” The reader of a description, as opposed to a blatant nar-
ration, “erwartet nicht das Ende, sondern die nächsten Worte, im Prinzip ad
infinitum” (“does not expect the end, but rather the next words, in principle
ad infinitum”). As we will soon see in more detail, the officer’s description of
how the machine should work hijacks almost the entire story, and thus also
does away with any reasonable expectations of narrative causality. This has
the confounding effect of making the climactic event more nonsensical rather
than more explicable; and further, Scherpe argues that the officer’s failed exe-
cution is an apt demonstration of this overly descriptive character describing
his way to his own death. Scherpe as well has little concern over “what” this
gesture means, and is instead interested in how it means. While description is
a process (Beschreibungsverfahren), narration is a progress; further, Kafka’s
officer’s descriptive process in the penal colony seems to work in dramatic
rejection of narrative progress: the “substance” of description (“Ihre Sache”) is
“die Zuständlichkeit und nicht der Progress, das Fortschreiten einer Hand-
lung. Die Beschreibung operiert in der ‘Auszeit,’ der Erzählung; sie sistiert
die Handlung” (“the cognizance and not the progress, the progression of a
plot. Description operates in the ‘time-out’ of the narration; it suspends the
plot”). While this does not necessarily take place in the authorial voice in
Kafka’s story, a similar movement does occur as the result of the dominance
of one character’s description, and the responsibility given to that descrip-
tion to “narrate” important past events. The ability of description to eclipse
narration in Kafka’s colony is, then, the embrace of situation—the painful
and protracted awareness of the condition one is presently in—rather than
progress, which as Wittgenstein (via Nestroy) has reminded us, always looks
greater than it really is. The conundrum upon which I would like to focus
here in the Kafka text deals with exactly this: progress, or the delusionary
nature thereof. And while Wittgenstein works through the Investigations
with the coy goal of undermining the conceit of philosophical progress (a tra-
jectory I explore in great detail in chapter 6), the progress that Kafka begins
to dismantle in “In the Penal Colony” is narrative progress.

As we will see, Kafka’s mastery in crafting “In the Penal Colony” lies in
appearing to tell a story but really doing something else, much like in crafting
Gregor Samsa’s transformation, he appeared to create a metaphor but really
did something else. This time, he has created something that takes the shape
of a narration and creates the expectation of one. But the narration of “In the
Penal Colony” does not actually make sense of anything, but rather makes
things more obscure, offers static and arresting descriptions of things and
actions that seem to be floating in space. We know Wittgenstein’s problem
with philosophical progress; now I venture that for Kafka, or at the very least
for “In the Penal Colony,” the problem with narrative progress is that it also,
contrary to Lukács, seems greater than it really is. All of this comes to light
when we view Kafka’s story alongside the paradox of rule-following.

Revisiting the Paradox of Ostension: The Case of the Zeichner

In order to see why the paradox of rule-following even arises as in Kaf-
ka’s penal colony, we must first revisit the paradox that precedes it—that of
ostensive language, which itself leads to the “meaning in use” argument of §43. As in *The Castle*, “In the Penal Colony” offers a remarkably astute take-down of ostension, with which I believe Wittgenstein would be impressed despite his professed dislike of Kafka’s fiction. In addition to being prescient, “In the Penal Colony’s” dismantling of ostension could not be more apt, for it comes in the breakdown of the apparatus’s component known in English as the “Designer,” and in German, *Zeichner*.

I will continue to use the German here, as the English fails to betray the orthographic relationship between the verb *zeichnen* (to sketch, to draw) and *Zeichen* (a sign or symbol). Indeed, the alleged literal purpose of the *Zeichner* is to be a “sign maker,” to create the shape of the text that is to be written upon the body of the condemned—to de-sign (*zeichnen*), to create the *Zeichen* that is meant somehow to signify (*be-zeichnen*) the law the condemned has transgressed. As, following Wittgenstein’s interpretation of Augustine (see chapter 4), the first users of human language gave primacy to ostension, so does the officer give primacy to the ostensive purpose of the machine—the *Zeichner* must work first in order for the rest of the execution to be carried out. Further, the officer’s description of the *Zeichner* must succeed in order for the narrative to proceed.

The problem that arises here is that either the *Zeichner* does not seem to work the way the explorer expects it to, or the officer has not described its working satisfactorily; either way, progress has been hampered (either narrative, cognitive on the part of the explorer, or both). This is because the condemned’s sentence remains (at least in the explorer’s understanding) incomprehensible to him. For it is de-signed/gezeichnet in the “special” language of the deceased old commandant; the explorer cannot even make out the characters that the machine is actually supposed to carve into the condemned’s body, which he recognizes only as “labyrinthartige, einander vielfach kreuzende Linien, die so dicht das Papier bedeckten, daß man nur mit Mühe die weißen Zwischenräume erkannte” (“lines that crisscrossed each other labyrinth-like, and covered the paper so thickly that one could only see the white space between them if one looked very closely”) (*GW* 1:172).

It is therefore not difficult to understand why the officer’s description of the *Zeichner* might not sit right with the explorer (who, throughout the acts of description, remains “nicht befriedigt” or “unsatisfied” [1:169]). As with the word *Landvermesser* (or *Schloß*, or *Dorf*) in *The Castle*, some sort of alternative understanding would predicate a definition (or at least use) of the concepts of *Zeichen*, *zeichnen*, and *bezeichnen* with which the explorer seems unfamiliar. Again, in a result Wittgenstein would deem inevitable, it is impossible for the explorer to discern what the primary act of the machine does because he is not already familiar with its system, with how it does what it does. And yet, in Kafka’s officer’s lengthy and passionate description of the machine, at no time does he truly succeed in showing how the machine works.
For the machine’s actions to make sense, they cannot simply be described: they must be removed from the officer’s descriptive control and instead be the fictional equivalent of “shown in action.” This has, at this point in the story, necessarily not occurred, and thus the true purpose and workings of the machine remain mysterious to both the explorer and the reader. In other words: the real “meaning” the machine offers—despite the efforts of the officer to describe it or the Zeichner to write or communicate a sentence, both grammatical and penal—comes only in its use. And this the officer admits when he explains that the condemned “experiences” the sentence “on his body”—but again, even this “experience” is related in an act of description rather than a narrative experience, and thus still remains obscure:


“Does he know his sentence?” “No,” said the officer, wanting to continue right along with his explanation, but the explorer interrupted him: “He doesn’t know his own sentence?” “No,” said the officer again, and paused a moment, as if he were requesting from the explorer a more detailed reason for his question, before saying: “it would be useless to avail him of that information. He will experience it on his body.” The explorer thought it best to stay quiet.

“Er erfährt es ja auf seinem Leib”—how exactly? Many in the theoretical community posit that the old commandant’s language and its method of “communication” are simply a new and different kind of writing. Kittler argues this “neue Art zu schreiben erzeugt ein anderes System der Macht” (“new way of writing creates a different system of power”), which to him is a system of technology-based recording of language, a technology-driven power that makes the new kind of writing possible. Gailus, on the other hand, places primacy on the prisoner’s sentence being communicated not onto the body but rather “through the body,” and that justice or the law can then be “deciphered through the pain of a body subjected to mechanical torture.” For Allen’s silence thesis, however, the machine’s writing causes non-language, a silence by way of succumbing to the machine’s might through both the gagging of the prisoner and the content of the old commandant’s “opaque script,” which Allen sees as being “as good as silence” because the explorer claims he cannot read it.
Yet all of these theories must remain purely in the esoteric realm, for as the notion of experiencing a sentence on the prisoner’s body is never actually experienced and only described, we never see a prisoner experience the law in any of the ways he “should.” There is no actual narrated intersection of the “symbolic and somatic,” nor a sacralization of silent acquiescence, nor an epiphany of the “master signifier” of justice, because we never see a prisoner executed. All we see of the prisoner are the pre-execution and post-release versions of him, and in his pre-execution stage he appears ignorant of both the content and form of his execution—evident by his eerie mimicry of the officer’s movements: “Er beugte sich hierhin und dorthin. Immer wieder lief er mit den Augen das Glas ab” (“He bent himself this way and that way, the whole time following the movements of the glass with his eyes”) (1:171).

What the prisoner seems to be doing here is a continuation of his daily life (the saluting of the captain’s door on the hour), which consists of training. Training is, of course, the method, rather than ostension, whereby Wittgenstein claims human beings learn a first language (PI §5). But the trouble is that there is an obvious potential disjunction between simply being trained to do something and understanding what that gesture means.

So how, then, does the prisoner “understand” something on his body? By “Er erfährt es ja auf seinem Leib” the officer seems to be claiming that a penal sentence receives meaning in its use, not unlike Wittgenstein argues of a grammatical sentence in §43. In the case of Kafka’s colony, the (penal and also allegedly grammatical) sentence the officer so painstakingly and yet obscurely describes only receives meaning in its use in its context: that is, not its use in language (“sein Gebrauch in der Sprache”) but rather, so to speak, its Gebrauch in der Strafe, or use in the punishment. And this is dependent upon the bringing about of a specifically prescribed dose of pain, which the victim is supposed to feel in a specific way and, in this specific use of the old commandant’s language, “understand” and “decipher” its meaning:

Verstand geht dem blödesten auf. Um die Augen beginnt es. Von hier aus verbreitet es sich. Ein Anblick, der einen verführen könnte, sich mit unter die Egge zu legen. Es geschieht ja nichts weiter, der Mann fängt bloß an, die Schrift zu entziffern. Er spitzt den Mund, als horchte er. Sie haben gesehen, es ist nicht leicht, die Schrift mit den Augen zu entziffern; unser Mann entziffert sie aber mit seinen Wunden. Es ist allerdings viel Arbeit; er braucht sechs Stunden zu ihrer Vollendung, (1:173)

Understanding dawns on even the dumbest ones. It begins around the eyes. From here it spreads. A look that could tempt one to lie down on the Harrow beside him. Nothing else happens—it’s just that the man begins to decipher the script. He purses his lips as if he were
listening. You have seen that it’s not easy to decipher the script with your eyes: our man deciphers it with his wounds. It’s obviously a lot of work; it requires six hours to complete.

It is as if the officer expects that his description will itself cause an understanding of the machine’s working, a “deciphering” of its seemingly altered semiotic function, to dawn upon the explorer, who surely must be brighter than the dumbest of prisoners. And yet the secondhand experience of the machine, as relayed by its most passionate and possibly unreliable adherent, is not sufficient to grant the explorer “understanding,” nor does he seem to be pressured into feigning it under implied threat of being considered dull-witted by the same dubious source. It seems instead that any sort of meaning—traditional, altered, as an object of description—offered by either the Zeichner or the officer’s description of it and its surrounding components is useless without the prior experience of the machine in use, which is nominally impossible (after all, the officer insists on describing the machine first [1:164]). Allegedly there has been prior experience of the machine in correct use. However, a major problem in accepting this prior use as evidence of correct use is that said correct use takes place entirely outside the narrative space, and evidence of its existence comes from a remarkably dubious source: the officer. The experience of the machine in use during a “successful” execution (that is, one that takes place in the manner the officer describes them) would itself constitute an act of what Scherpe has called successful narration (events that take place over a period of narrated time and that allow the reader to make sense of what is happening).

But in Kafka’s story no such execution takes place. Neither, then, does the execution of a narrative, in the sense of “precise completion.” The only execution that takes place is the metaphorical execution of the act of narration; in carrying out this execution instead of the others, Kafka has brought about several potential conflicts while simultaneously avoiding the successful bestowing of meaning or sense upon the machine destined instead to remain opaque, disjunctive, seemingly nonsensical, and thus potentially meaningless. In *The Metamorphosis* we discovered a metaphor that had the shape, trajectory, and necessity of a metaphor but no actual metaphorical content. So, too, does the execution (penal, narrative) in Kafka’s colony reveal itself to be starkly different from the form it appears to take. In replacing the narration of an actual execution with the description of an alleged execution (and thus the “execution” of narration), Kafka has brought the following seemingly stable elements of his story into contention: first, despite the officer’s painstaking description of the machine, we still have no idea whether or not an execution means in the way the officer claims it means, whether or not a prisoner can truly experience the understanding on his body. This is, further, a problem neither solved (on a nonmetaphorical level) by literary criticism, nor dissolved in a Wittgensteinian manner. But the Wittgenstein text is far
from irrelevant, because as we are about to see, the lack of narrated evidence of the machine in prior successful use—seen through the lens of the paradox of rule-following—allows us to question as forcefully as possible whether such an execution has ever even taken place at all.

**Rule-Following and the Execution of Sense**

In connecting the officer’s explanation that the condemned experiences his sentence on his body with Wittgenstein’s assertion in §43 that for a large class of words the meaning of a word is *its use* in language, an incompatibility or contradiction appeared when it became apparent that any “use” of the penal colony’s machine that has the potential to grant it meaning takes place outside the narrative space and is thus useless for our purposes. Or rather, this “use” takes place outside the narrative space *if it takes place at all*. Here we will bring in another of Wittgenstein’s paradoxes to shed light on the fascinating implications of this very real uncertainty.

Let us return to the previously discussed moment in which the officer describes a penal colony execution to the explorer (1:167–69). He begins by claiming that the method of execution does not sound particularly harsh, and then elaborates:


> “The command that the condemned has transgressed will be written by the Harrow upon his body. On this one, for example”—the officer pointed to the man, “will be written: Honor thy Superiors!”

Much of what appears in this interaction consists of the explorer voicing his expectations to the officer, and then reacting with dissatisfaction, a dynamic that becomes apparent when the officer answers the explorer’s follow-up question (“Aber daß er überhaupt verurteilt wurde, das weiß er doch?” [“But he knows that he’s been sentenced in the first place, doesn’t he?”]):

“No, he doesn’t know that either,” said the officer and smiled at the explorer, as if he expected from him yet more special revelations. “No?” said the explorer, rubbing his forehead, “so the man also doesn’t know how his defense was received?” “He has had no opportunity to defend himself,” said the officer.

This results in the moment I have already referenced: the explorer is unsatisfied, and he continues not to be satisfied when the officer shows him the transparent Harrow with its long needles that “write” and shorter ones that squirt liquid to clean away the blood, so that the sentence always remains “klar zu erhalten” (“clear to make out”) (1:170). The trouble here is that the old commandant’s “scripture” is not—given what the explorer seems to expect from the word “klar” and the word “erhalten”—“klar zu erhalten.”

Scherpe’s notwithstanding, much analysis of this moment tends to bypass the fairly obvious presence here of a description that makes for quite a dissatisfactory explanation. Instead, many critics understandably concentrate on the perceived juridical impropriety described, best exemplified in the officer’s proclamation that “Die Schuld ist immer zweifellos” (“Guilt is never to be doubted”) (1:168). Koelb, for example, sees the disconnect or lack of clarity in both the old commandant’s writing and its assumed manifestation on a prisoner’s body as itself the point of the story, which is the “act of reading” rather than writing. He sees the central act of the story as the “scene of reading” (after Derrida); the “successful” application of the penal sentence is actually “the embodiment of the intellectual reproduction that we ordinarily call ‘reading.’ ” And by “reading” Koelb actually means a struggle between what Derrida has characterized as two types of reading, the first being “deciphering,” which “escapes play” and “the order of the sign,” and the other kind, which is “no longer turned toward the origin, affirms play and tries to pass beyond man and humanism” by way of Nietzsche and the “joyous affirmation of the world and of the innocence of becoming.” Koelb argues, then, that while the explorer seems dissatisfied because he cannot “decipher” the old commandant’s writing nor can he figure out how the prisoner is supposed to experience an epiphany “with his wounds,” his problem is not that he lacks the experience that would make such deciphering possible, but instead that the kind of “reading” a user or beholder of the machine must undertake is actually this second, Derridean type of reading.

All this seems to be criticism applied to the machine in its state of working correctly—but again, this is a state that is only described to us secondhand; it is a moment of action the explorer does not witness and an act of narration we do not witness. Where, then, is the proof that this state of working correctly exists at all? The sole instance during which we and the explorer actually see the machine in action is its alleged malfunction. As we now explore the officer’s suicide—one of the story’s few moments that resembles traditional
narration—uncertainty about whether the machine ever has worked in the manner in which the officer describes should emerge.

One of the first indications that the execution may not go according to plan comes when the officer complains that he overheard the new commandant’s words of invitation to the explorer and that he understood “sofort, was er mit der Einladung bezweckte” (“immediately what was meant by that invitation”) (1:179–80). In this moment the officer takes great care to emphasize that his act of reaching out is not an institutional or “community” act. Rather, it complicates the community dynamic more than anything else:

> “Wenn ich eine Meinung aussprechen würde, so wäre es die Meinung eines Privatmannes, um nichts bedeutender als die Meinung eines beliebigen anderen, und jedenfalls viel bedeutungsloser als die Meinung des Kommandanten.” (1:181)

> “If I share an opinion here, it’s the opinion of a private citizen, no more significant than anyone else’s, and in any case far less significant than that of the Commandant.”

Although the officer insists that there are indeed adherents to the old commandant throughout the colony (they are simply silent), there is really only a “private citizen” self-described to legitimize the antiquated method of execution. This moment in the narrative will become crucial after the officer has died and the explorer stumbles upon the sole piece of evidence that there had potentially been a community with a shared “understanding” of the old method of execution, a particular “form of life” (Lebensform), as Wittgenstein will call it (§§19, 23, 241).

But for now, let us take the officer at his word as a “Privatmann.” The officer’s status as a lone “private citizen” is what some critics, Sokel and autobiographical adherents foremost among them, argue makes him “special” and thus also have a strong affinity to his creator, “creator” being used in a loose sense to describe the old commandant whom, it goes almost without saying, we also never truly encounter in the textual space. And when this “specialness” is challenged? First, the officer suggests to the explorer that, should anyone in the outside world ask, he shouldn’t lie about his opinions, if indeed he does believe what he has witnessed is barbaric, but could he not simply omit his judgment (1:183)? Since the explorer seems not to “understand” the machine, would it really be too much trouble for him to grant the officer what Allen terms the silence of assent, a sort of perverted version of TLP 7, “Wovon man nicht sprechen kann, darüber muss man schweigen”? This the explorer will not do, and thus the officer apparently decides to sacrifice himself to the machine. Here we have what is arguably the most easily understandable gesture in the story—a gesture which also happens to
be one of the most strongly narrated rather than described, as the sense of objective time, of moving forward, of action suddenly becomes utterly sharp, if not also utterly devoid of context: the officer, who has until now shown himself to be a barbaric and irrational person searching for approval, makes a desperate overture and reacts to rejection of that overture in a proportionately barbaric manner. In fact, this seems to be a sole provoked payoff in an otherwise causality-bereft narrative universe. To be sure, this payoff is bitter and bloody: the punisher becomes the punished—but it is easily understood, given this context, where the officer’s love affair with torture was bound to lead to some sort of violent consummation.

Thus it is only surprising in the way alleged twist endings in all narrative (i.e., sense-making) fiction are surprising when the explorer tells the officer he cannot comply with the odd request to further the old commandant’s agenda through silence, and then the officer says that “now it’s time” (“Dann ist es auch Zeit”), and proceeds to free the prisoner and prepare to execute himself (1:186). The officer rearranges the designer to “spell” out a new sentence, which the officer must “translate” for the explorer’s benefit: “‘Sei Gerecht!’ heißt es,” sagte er, ‘jetzt können Sie es doch lesen’” (“‘Be just!’ it says,” he said, “‘surely now you can read it’”) (1:187). But, unfortunately, the officer’s twelve hours of epiphany do not occur. Instead, as we have seen before, the Harrow fails to write, the officer is simply impaled in a gruesome act the explorer terms “murder, pure and simple” (1:192).

What is at first apparent is that the twist in the plot has twisted again, this time not quite so understandably. The officer’s execution fails—according to all prescriptions by the officer himself of what would in this context constitute “success”—to result in him “deciphering” “sei gerecht” on his body. So, then, what has happened must be dramatic irony: the officer has failed in the very gesture of sacrificing himself to his own machine, because the machine didn’t work in the way he claimed it would. However, dramatic irony only itself works if the machine did indeed malfunction—but did it? In fact, to reiterate the earlier problem of “The Judgment,” most possible interpretations of “In the Penal Colony” only seem to “work” when the act of the officer’s suicide is judgeable, and thus judged—nonsensical, ironic, symbolic, fateful, anything. Here I would like to offer a counter-argument: that the story “works,” thanks in part to its clever subversion of the narrative medium, largely allowing narration to be eclipsed with the officer’s lengthy descriptions up until the moment he decides to end his own life, by disallowing that gesture any sort of interpretive judgment, any sort of sense-making privilege the successfully executed narrative might normally bring. Any judgment of the action as ironic, symbolic, fateful, or otherwise is predicated on the recognition of the final “twist” in the plot: that the machine failed or malfunctioned, that it did not work in the way it was supposed to work. The problem with this is that there is actually no irrefutable textual evidence that the machine has ever worked the way it was “supposed” to work. Instead, it
is possible to argue that the machine simply did what it was “supposed” to do when its most passionate adherent turned it on himself.

This discovery—that it is impossible to prove the validity of a prescription for how something is “supposed” to go, most commonly referred to as a “rule”—is at the heart of Wittgenstein’s paradox of rule-following, which comprises the approximately one hundred remarks up to and including §201, and further complicates development on the previously discussed paradox of ostension (and that of “explanation” or logic, which we have not discussed here). As such the paradox unfolds in a similar way: through demonstration in a complex language game, it is shown that this particular construct of language (the concept that one might “follow a rule” set forth in language) does not actually work—because in order to understand how it works, one must already know how it works:


Our paradox was this: no behavior could be determined by a rule, because all behavior can be made to agree with the rule. The answer was: if everything can be made to agree with the rule, so can it also to contradict it. Thus there would be here neither agreement nor contradiction.

This is a remarkably complex and prolifically debated passage, itself following a remarkably complex and interwoven series of remarks, several of which we must examine in more detail before we can begin to see what §201 actually might mean. As the entire methodology of the Investigations necessarily eschews platitudes and propositions in favor of examples, “samples” (Muster), and “games,” Wittgenstein must enact the fallibility of the superconcept; in order to do this, he comes up with examples that need their own examples, which also need their own examples. The examples that can make the rule-following paradox clearest are, inspired by David Stern’s reading, those of “fit,” of learning the meaning of a word “in a flash” (which is borne out of “fit”), of learning to count by twos (which will come into play again when we work with Kripke’s Wittgenstein), and, finally, of “reading,” paying particular attention to Wittgenstein’s example of a human who is a “reading machine” (“Lesemaschine”), a term that harkens back to Koelb’s Derridean interpretation. In each of these cases, the superconcepts of “fit” (“Passen”), “understanding” (“Verstehen”), and “ability” (“Können”) will turn out to need governing superconcepts of their own in order to “work,” thus highlighting the fallibility of rule following.
Wittgenstein begins his discussion of “fit” by dismantling the assumptions the “author of the Tractatus” had to have been acting under in order to come up with the general form of all propositions. The form (“die Allgemeine Form des Satzes”) claims that what all propositions have in common is that they all, in one form or another, say “such and such is the case,” and they do this by connecting a group of simples or atomic propositions by some combination of the “neither-nor” operator, and as such they are truth-functional (TLP 6). In Investigations §§134–37 Wittgenstein seems to undermine this bipolarity thesis, which is essentially the culminating systematic idea of the Tractatus. In §136, he takes his former self to task by pointing out that in order to judge a proposition, one must already have an extra-propositional conception of “truth” that exists apart from the proposition but “fits” together with it:

Im Grunde ist die Angabe von “Es verhält sich so und so” als allgemeine Form des Satzes das gleiche, wie die Erklärung: ein Satz sei alles, was wahr oder falsch sein könne. Denn, statt “Es verhält sich . . .” hätte ich auch sagen können: “Das und das ist wahr.” (Aber auch: “Das und das ist falsch.”) Nun ist aber ‘p’ ist wahr = p
‘p’ ist falsch = nicht-p.

Und zu sagen, ein Satz sei alles, was wahr oder falsch sein könne, kommt darauf hinaus: Einen Satz nennen wir das, worauf wir in unsere Sprache den Kalkül der Wahrheitsfunktionen anwenden . . .

Was in den Begriff der Wahrheit eingreift (wie in ein Zahnrad), das ist ein Satz.

Aber das ist ein schlechtes Bild. (§136b–d)

Ultimately, the designation of “Such and such is the case” as General Form of Proposition is the same as the declaration: a proposition is something that can be true or false. Because instead of “such and such . . .” I could also have said “such and such is true.” (But also: “Such and such is false.”) Now let us say:

‘p’ is true = p
‘p’ is false = ¬p.

And to say a proposition were everything that could be true or false comes down to this: what we call a proposition is that to which we apply the calculus of truth-functions in our language . . .

What engages with the concept of truth (like in a cogwheel) is the proposition. But this is a poor picture.

So here the problem of “fit” is introduced: “true” and “proposition” must fit together, and to do this “true” must exist independently of “proposition” and, somehow, also “fit” with it like the teeth of a gear wheel with the gaps of another gear wheel. The “true” has to exist independently and
simultaneously be dependent on context—which leads us, understandably, to ask: which is it? (or at any rate to think that the gear wheels are a “schlechtes Bild”). This leads Wittgenstein into one of many instances wherein he links the idea of understanding something “in a flash” with the idea of “fit” (that is, to understand whether a proposition is true or false, you also have to have understood the concept “true” on its own and yet in this context, all at once, which would require some sort of in-a-flash development). Wittgenstein’s main example of the problems of this phenomenon come in §139, when he talks about hearing or reading the word “cube”—and concludes that if we do happen to get a picture of a cube in our minds as a result, this is merely because the word has suggested a certain use to us, not that the word is inextricably linked to the picture by a governing force of referentiality. This finding, however, far from being conclusive, merely opens the proverbial Russian doll to three new examples, one of which is the aforementioned “fit” and the others of which are Verstehen, Passen, and Können:

Die Kriterien, die wir für das “Passen,” “Können,” “Verstehen” gelten lassen, sind viel kompliziertere, als es auf den ersten Blick scheinen möchte. D.h., das Spiel mit diesen Worten, ihre Verwendung im sprachlichen Verkehr, dessen Mittel sie sind, ist verwickelter—die Rolle dieser Wörter in unserer Sprache eine andere, als wir versucht sind, zu glauben. (§182b)

The criteria we allow for “fitting,” “being able to,” and “understanding” are much more complicated than they might appear at first glance. That is, the game with these words, their use in linguistic circulation that is carried on by their means, is more involved—the role of these words in our language other than what we are tempted to think.

Thus “to fit” creates the problem of “to understand,” which itself creates the problem of “to be able to” do something, such as reading. Is it possible that in order to understand the general form of proposition, we have to somehow “understand” the idea of “truth” instantly, or “in a flash”? This is how the problem of “fit,” instead of being at all resolved, reveals the problem of Verstehen, that is, instant understanding, the idea of grasping a word “in a flash,” the Investigations’ major argument on behalf of a superconcept.

In order to demonstrate the problem with “being able to do something” (and thus also, implicitly, with “understanding” and then, were something to be able to be understood, the idea of “fit”) Wittgenstein turns in §§156–58 to the problem of “reading,” which he characterizes as the act of reproducing language to oneself as a result of having the eyes register marks on paper. The problem as Wittgenstein sees it is that the human being has the potential to be simply a “reading machine” who does not actually “understand” anything he
reads, just as much as he has the potential to be a human being reading in the manner we expect. To help us understand this point, the “reading machine” example digresses into yet another example of reading without really reading, this one of a schoolchild who is ostensibly “learning to read,” and in his first attempt, happens to reproduce a word correctly by chance, though he does not actually know how to read yet, and this is followed by many incorrect guesses, and so his teacher says, “No, that’s not reading,” it was just luck. Now we see that several months later, when his reading has become consistent, the teacher says, “Well, now he can read.” Was the first word he read actually a correct act of reading? Or was it really an accident? When did the child “really” start reading? (§157). It is, Wittgenstein argues, a matter of the student’s behavior (“Verhalten”) in context, and not at all the superconcept of “to read correctly.”17 This brings Wittgenstein back to the example of the reading machine:

Im Falle aber der lebenden Lesemaschine hieß “lesen”: so und so auf Schriftzeichen reagieren. Dieser Begriff war also ganz unabhängig von dem eines seelischen, oder anderem Mechanismus.—Der Lehrer kann hier auch vom Abgerichteten nicht sagen: “Vielleicht hat er dieses Wort schon gelesen.” Denn es ist ja kein Zweifel über das, was er getan hat.—Die Veränderung, als der Schüler zu lesen anfing, war eine Veränderung seines Verhaltens; und von einem “ersten Wort im neuen Zustand” zu reden, hat hier keinen Sinn. (§157d)

In the case of a living reading machine “reading” meant: reacting to written signs in this way and that. This concept was thus quite independent from that of a mental or other mechanism.—Nor can the teacher here say of the pupil: “Perhaps he has already read this word.” For there is no doubt as to what he did.—The change, when the pupil began to read, was a change of his behavior; and to speak of a “first word in his new state” makes no sense.

This, as Stern argues, leads us to the realization that “fit” (such as the “fit” of “reading correctly” with “knowing how to” or “proposition” with “truth”) describes a situation, or, in less Tractarian terms, a behavior and not a rule,18 and thus the full scope of the contentiousness of the idea of a superconcept has come full circle by the time we get to the paradox of §201, which Stern paraphrases as follows:

No occurrent act of meaning or intending can give a rule the power to determine our future actions because there is always the question of how that act is to be interpreted. As a result, the idea that a rule, taken in isolation, can determine all its future applications turns out to be misguided . . .
Nothing taken in isolation from its context can determine how we go on, as all determination is dependent on our proceeding in the usual way.19

As we said before: in order for a rule to “work,” there would have to be a rule governing how that rule worked; in order for that rule to work . . . and so on. The only thing that could stop that infinite chain would be a superconcept, which Wittgenstein begins dismantling with the “in a flash” discussion, weakens further as each example further chips away at “fit,” and to which he seems to deliver a death blow not long after §201:

“A thing is identical with itself.”—There is no finer example of a useless sentence, which is somehow still connected with a play of the imagination. It is as if in imagination we put a thing into its own shape, and saw that it fit.

We could also say: “Every thing fits into itself.”—Or otherwise: “Every thing fits into its own shape.” At this we look at a thing, and imagine that there was a spot made for it, into which it now exactly fits.

Does this spot “fit” into its white surrounding?—But that is exactly how it would look, if there had at first been a hole in its place, and it then fit into that hole. But when we say “it fits,” we are not just describing this appearance, not simply this situation.

“A thing is identical with itself.”—There is no finer example of a useless sentence, which is somehow still connected with a play of the imagination. It is as if in imagination we put a thing into its own shape, and saw that it fit.

We could also say: “Every thing fits into itself.”—Or otherwise: “Every thing fits into its own shape.” At this we look at a thing, and imagine that there was a spot made for it, into which it now exactly fits.

Does this spot “fit” into its white surrounding?—But that is exactly how it would look, if there had at first been a hole in its place, and it then fit into that hole. But when we say “it fits,” we are not just describing this appearance, not simply this situation.

“Every colored spot fits exactly into its surroundings” is a somewhat specialized form of the law of identity.

Again, now we see here that by recognizing the absurdity of saying that a black ink stain “fits” perfectly in its background, we are recognizing the
absurdity of the idea that “fit” can be a superconcept, and are instead coming to recognize that “fit” instead describes a specific use or situation, as do all instances that attempt to claim they are following rules.

So, to place the paradox of rule-following in similar terms as we have placed that of ostensive definition, we can say that the problem with “learning” to follow a rule is that when we follow a rule, we already have to know what it means to follow a rule, and thus must be following a rule (Wittgenstein mulls on this very dilemma even as far back as §84, when his interlocutory voice asks: “—Können wir uns nicht eine Regel denken, die die Anwendungen der Regel regelt? Und einen Zweifel, den jene Regel behebt,— und so fort?” (“—can we not think of a rule that regulates the application of this rule? And a doubt that eliminates that rule—and so on?”) (§84a).

In appropriate fashion, we can illustrate the relevance of the rule-following paradox to the officer’s suicide in Kafka’s penal colony using an example not of its “Gebrauch in der Sprache,” but of its use in the failed execution. Looking once again at the moment in which the officer expresses disbelief that the explorer cannot “read” the old commandant’s writing, we must also look again at Wittgenstein’s problem with “reading,” which was actually an example of his problem with “rule-following” altogether. And again back to Wittgenstein’s example of the reading machine and the schoolchild who “accidentally” reads correctly in an isolated incident. Through these language games, Wittgenstein has demonstrated that there is no way to know what it means “to know how to read,” that there is thus no such thing as following an actual “rule” for “learning to read”—that would require a super-rule for knowing how to follow a rule, and Wittgenstein has rejected the superconcept solution.

The question then arises: (how) is the paradox of rule-following relevant to Kafka’s penal colony? In the alleged malfunction of the machine, in the moment of the officer’s suicide, when the Harrow does not write and instead simply impales the officer in an act of what the explorer dismisses as “murder, pure and simple” (1:192)? What, exactly, has happened in this moment? Is it, as Goebel has argued, a symbolic suicide of colonial power? Has the machine, as Politzer argues, taken on human characteristics or “acquire[d] human life”? Or is the malfunction simply the completion of the story’s “pornological fantasy,” as Margot Norris argues, and a necessary plot twist, given that, in her reading, Kafka saw to it that the officer was “robbed of a transcendence that was always fraudulent, and [his] carcass . . . therefore disposed of with the unceremonious dispatch of animal burial”? Is the machine’s malfunction necessary for the story itself to function, specifically as a “blow at reason,” at “the ‘rationalization’ of suffering,” the transcendence of pain? If Norris is correct, then the point of the machine has always been to malfunction, its function being simply as a narrative (or, rather, descriptive) construct created to show the futility (and incorrectness) of the sacralization of pain. Koelb also seems to argue that the machine’s malfunction is, if not the only
important aspect of its function, at least immanent, for after all, “[a] machine that did run by itself—what would that be? No longer a pure representation, it would deconstruct the very notion of a ‘machine.’ A machine running by itself, especially a reading/writing machine, would deconstruct itself. Perhaps it would also self-destruct.”

This is because, according to Derrida, a “pure representation, a machine, never runs by itself.” And thus further:

The machine, liberated by the traveler’s disbelief, destroys itself by rebelling against the condition of its existence, which stipulates that such a machine may not run by itself. The apparatus transcends its own nature by its act of self-destruction, for in doing so it ceases simply to represent the system of reading and writing the law.

The purpose of a machine that ostensibly runs by itself must be to self-destruct, because inherent in the concept “machine” is “cannot run by itself.”

Like Norris or Koelb, we can also argue that the “purpose” of the machine may indeed be to malfunction—but that this purpose may also carry this property without a “higher” (that is: metanarrative or structural) purpose (i.e., something more intrinsic to the story than “it had to happen for the story to ‘work’”). Indeed, constructions of pornological fantasy aside, what appears to have happened here is that the machine has malfunctioned. But I would like to argue that the error is not the machine’s; it is everyone else’s, for laboring under the illusion that there was a set way in which the machine was to “function” at all. In §143, Wittgenstein argues that it is never really possible to tell the difference between a “random” and a “systematic” mistake, and this argument should resonate when we ask, regarding the “mistake” of Kafka’s machine: has the machine “malfonctioned” for the first time? There is, in the context of the narration, no way to tell. Further, the only evidence that the machine had any prior human contact at all, the saliva-soaked wool gag that causes the prisoner to lose his meager lunch after he bites down on it, could easily have gotten into its condition in countless other ways (“Wie kann man ohne Ekel diesen Filz in den Mund nehmen, an dem mehr als hundert Männer im Sterben gesaugt und gebissen habe?” [“How can one not take this felt into his mouth without revulsion, knowing that more than a hundred men before him have sucked and bitten on it?”] [1:176]). To understand on a purely literal level the officer’s suicide in this story—much less to afford oneself the luxury of interpreting said act as “meaning” any number of its metaphorical equivalents—we must accept the moment of surprise and fulfilled dramatic irony that occurs when the machine, instead of allowing its greatest admirer to “sacrifice” himself to it in “Folter,” malfunctions. But because there is no textual evidence of the prior “correct” functioning that would determine a malfunction other than that provided by a remarkably unreliable source.
This unreliable source—the officer—is self-proclaimed engineer, judge, jury, and executioner of the colony (1:168): the architect of all of its present rules (or at any rate the mouthpiece of the architect of its past rules). We can see using simple common sense that these “rules” are arbitrary at best and unintelligible at worst, and that there is no objectively narrated evidence to support them. But what is truly fascinating about this story’s connection with the Wittgenstein text is that—just as Kafka’s characterization of failed metaphor helped elucidate the problem with all metaphor, which echoed structurally Wittgenstein’s problem with all philosophical language—Kafka’s own characterization of this particular example of untrustworthy narrated rule-making and sabotaged narrative execution demonstrates rather incredibly Wittgenstein’s contention with all rule-making.

We will extrapolate on this idea in the forthcoming discussion of Saul Kripke, a philosopher whose interpretation of §201 is both extraordinarily elucidatory for our purposes and highly problematic, but at the moment it remains crucial to understand the rule-following paradox as it relates to Kafka’s story: a “rule” is a misnomer because the one thing it is supposed to do—determine what one is supposed to do—it cannot do without some sort of super-rule to control “how it is to be interpreted,” which Wittgenstein has dismissed. Furthermore, Wittgenstein argues that the only way to tell whether a rule really works or not is to examine its behavior in “normal” cases, the “control experiments,” as it were—with an abnormal case, which Wittgenstein argues every case can be made out to be (a remark which the only “actual” execution in Kafka’s story handily demonstrates), there is no way to tell whether the rule works or not (and further, “je anormaler der Fall, desto Zweifelhafter wird es, was wir nun hier sagen sollen” [“the more abnormal the case, the more doubtful it becomes what we should say here”] [§142]). Wittgenstein argues again and with more force on behalf of this notion later in the rule-following section, shortly before the paradox of §201, when he comes to the conclusion that it is not possible for one person to follow one rule once:

Es kann nicht ein einziges Mal nur ein Mensch einer Regel gefolgt sein. Es kann nicht ein einziges Mal nur eine Mitteilung gemacht, ein Befehl gegeben, oder verstanden worden sein, etc.—Einer Regel folgen, eine Mitteilung machen, einen Befehl geben, eine Schachpartie spielen sind Gepflogenheiten (Gebräuche, Institutionen). Einen Satz verstehen, heißt, eine Sprache verstehen. Eine Sprache verstehen, heißt eine Technik beherrschen. (§199b–c)

It is not possible for there to have been only one occasion on which someone obeyed a rule. It is not possible that there could have been but one occasion on which a report was made, an order given or understood, and so on.—To obey a rule, to make a report, to give
an order, to play a game of chess, are habits (uses, institutions). To understand a sentence means to understand a language. To understand a language is to master a technique.

And this has been, in effect, what the officer has done in Kafka’s story as well: he has described the rule for how the machine works and then attempted to follow it, but because he is but one man attempting to follow one rule once, we have no idea whether the machine’s malfunction was random, systematic, a normal case, a special case, or, in fact, a “malfunction” at all, because all of the officer’s actions have existed in his own personal juridical and punitive microcosm, with parameters of his choosing and his choosing only (or, at any rate, with help only from the deceased old commandant). Thus we realize that it is indeed impossible to tell whether the machine has malfunctioned, because no super-rule governing all of the officer’s other rules was possible, and thus none of his rules are provable as rules.

As Robert Fogelin puts it, the “notion of following a rule” is “a natural source of philosophical illusion. Pressures seem to come from every side to turn this notion into a super-concept,” and, further, “it is a central task of Wittgenstein’s later philosophy to fight this tendency by showing that rules are neither sublime nor are they mysterious.” That is: rules are not “sublime” because there is no super-rule governing the rules for following all rules; they are not “mysterious” because they are taken to be a rejection of the Platonism Norman Finkelstein analyzes, which he claims “makes communication look miraculous.” That is, the Platonist would argue that though there is no super-rule for governing all rules, the idea of “rule” has an abstract, mysterious, but nevertheless existing Platonic truth concept somewhere, so that we can, when, for example, following a recipe for meringue, “come to grasp which of [our] activities would accord, or fail to accord, with a sequence of noises that [we’ve] heard.” A Platonist, Finkelstein suggests, “might say”:

My words and gestures can be interpreted any which way, but the thing behind them—the meaning—needs no interpretation. Now, I can’t convey this item directly to my interlocutor. All that I can do is talk to him, or gesture to him, and all my words and gestures can be interpreted in various ways. But if he’s lucky, he’ll guess what I have in mind and understand me.

Wittgenstein, Fogelin argues, wholly rejects this kind of Platonism (and this should be no surprise), and instead seems to enjoy that interpretation [of a rule] is an obvious candidate to play the role of bridge, but—under the pressure of an insistence that there’s a gulf between any string of words and what it calls for—every interpretation seems inert as well. The paradox of . . . §201 has its roots in the
thought that there is always a gulf between the statement of a rule—a
string of words—and the rule’s execution or application.\textsuperscript{30}

Wittgenstein relishes the struggle to resist, always, not only the existence
of a sublime superconcept, but also that of a prelinguistic abstract truth.
Rather, he simply presents us with the “gulf,” the paradox of §201. And it
is with an acceptance of this “gulf,” rather than the reliance upon an extra-
textual truth easily assumed (that an execution has taken place as planned
before), that we can best appreciate the depth and nuance of the uninterpret-
ability of the officer’s suicide in Kafka’s story.

It is also through an exploration of the rule-following paradox that we
can discover why at least two of the officer’s most crucial directives have not
exactly failed, but have proven themselves impossible to follow “correctly.”
The second and main directive is, of course, that the machine works all by
itself. The first, with which we will begin our analysis of Kripke’s reading, is
this notion of “‘Sei gerecht,’ heißt es” (“‘Be just,’ it says”). To the problems
with “Sei gerecht, heißt es,” we can actually find a hint of the analytic tradition
in Kittler’s work—for he invokes Russell’s theory of nonsense via category
error to argue that the problem with the machine stems from its instructions,
or what it is “programmed” to do. That is, according to Kittler, the machine
is programmed to carry out specific sentences (“Ehre deinen Vorgesetzten!”)
and has never before been asked to carry out a general one (“Sei gerecht!”)—
therefore, Kittler argues that the machine may simply not be programmed to
function with “Be just” and, like a computer, it “crashes.”\textsuperscript{31} However, I will
argue that the problem or potential problem is not with “Sei gerecht,” but
rather with “heißt es.” Using Kripke’s argument, we will be able to see that
by “heißen” it is entirely possible that the officer means something entirely
different than our usual notion of “to be called,” and has, unbeknownst to
the explorer, simply used the word “heißt” to indicate this “other thing,”
because, there being no super-rule to control the rules of how words or other
operators are used in each and every case, there is no way to stop him.

This is an adaptation of the Kripkean term “quaddition,” noted with the
function “quus” rather than “plus.” How can we really say that any action
may be made out to accord with a rule? Because there is no way for us to
prove that the way an operation, designated by a particular mark, worked
in the past will be the way it works in the future. “Quaddition,” Kripke
explains, works just like addition—until it ceases to, and the trouble is that
there is no way for us to know or control when the operation will mutate.
Take, for example, the supposedly easy computation 68 + 57. “I perform the
computation, obtaining, of course, the answer ‘125,’ ” he finds, and, further,
this answer, after careful review, should be correct both in the arithmetical sense that 125 is the sum of 68 and
57, and in the metalinguistic sense that “plus,” as I intended to use
that word in the past, denoted a function which, when applied to the numbers I called “68” and “57,” yields the value 125.

Now suppose I encounter a bizarre skeptic.32

This bizarre skeptic suggests that the answer should instead have been 5, which Kripke’s fall guy first purports to be “obviously insane.” But, after closer consideration of the skeptic’s argument, suddenly all is not so certain: is Kripke sure that by “+” he meant addition in the way most of us who have completed elementary school conceive of addition? After all:

Who is to say what function this was? In the past I gave myself only a finite number of examples instantiating this function. All, we have supposed, involved numbers smaller than 57. So perhaps in the past I used “plus” and “+” to denote a function which I will call “quus.”

X “quus” y, furthermore, works exactly like addition as long as y is less than 57. But if y is over 57? X “quus” y is always 5. “Who is to say,” asks Kripke, that “this is not the function I previously meant by ‘+’”?33 Thus as Kripke conceives it, Wittgenstein’s “new kind of rule skepticism” can be summed up as follows: if we have a rule/operator/category/signifier (counting, addition, color, etc.) and we have, until a certain moment, gone about our business applying it in our “normal” accustomed way (1 + 1 = 2; 3, 6, 9; turquoise), how can we prove that we were ever “really” following the rule like we believed it was written, and not, instead, doing something else—quounting, quaddition, schmolor, etc.—something that can be marked by the word or expression we usually use to mark our operation (“I’m counting,” “+,” “this chair’s color is turquoise”), but which really means something else?34

Conventional wisdom would tell us that no normal person can read the old commandant’s writing and that when it was the officer’s turn to die, the machine didn’t work and thus his death was, indeed, “unmittelbarer Mord.” But, as Wittgenstein’s §201 would have it, there is actually no way to tell. For how are we to know, really, for certain, that when the officer says “‘Sei gerecht,’ heißt es,” by “heißt” he means “says” and not something else, say, “queißt,” which means “says, except in this particular situation, in which case it means causes an extralinguistic epiphany on the body” and which, unfortunately, is designated in ordinary language, unbeknownst to (or forgotten by) the officer, by the word “heißt”? While this would certainly be inconvenient for many of the most compelling metaphorical or allegorical interpretations of “In the Penal Colony,” and may seem a tad contrived, its is meant to convey merely the shadow of a doubt that “heißt es” means what we think it means, to show that it is hasty to assume that this phrase can mean in the way we expect it to in the context of a reliable narrative. Because “In the Penal Colony” is not a reliable nor particularly “successfully” executed narrative, and its climax cannot be assumed to function in the way
a standard narrative climax would. Thus if we are to give credence to at least this part of Kripke’s interpretation of §201 (which I do find quite helpful), we can then see that the most striking element of the officer’s suicide is not simple dramatic irony. The officer’s death is, instead, something far more chilling because there is simply nothing that can be done with it: the officer’s suicide, rather than being the easily judged “unmittelbarer Mord,” is “unmittelbar” but perhaps not as simple as “Mord”: it is uninterpretable.

A Skeptical Conclusion

Here, then, we must examine the consequences—philosophical and literary—of determining the undeterminable. The good news is that the dis/solution of the rule-following paradox is not as clear as that of the paradox of ostension; this is advantageous, in the sense that it therefore cannot itself be dismantled with such ease; the less-good news is that the complexity of the rule-following passages of the *Investigations* cause their very existence to be the source of extensive debate, especially about Saul Kripke. Kripke is the foremost example of philosophers who argue that rule skepticism can be solved the same way referential skepticism can (and thus, on the literary side of things, also implying that the “problems” we encountered in both referential and metaphorical meaning can also be resolved): not simply by language’s use, a so-called rule and its “Gebrauch in der Sprache,” but its patterns and forms of use *within* a community of language users, as part of a “form of life” (*Lebensform*). But problems arise, just as before: first of all, the solution of a language community or *Lebensform* itself could be taken to espouse exactly the kind of superconcept or super-rule the very paradox of rule-following precludes.  

Secondly, and presenting vastly more far-reaching problems, Kripke’s approach (along with countless others) posits that §201 has philosophical consequences because it espouses actual philosophical doctrine—and not just §201, but all of the *Investigations*. This brings us back to our original discussion of Wittgenstein’s text and “the problem with [philosophical] progress,” and radically skeptical approaches: as we saw before, unlike Kripke, a number of “New Wittgenstein” critics believe that the paradoxes we have just examined with great care—ostensive definition, rule-following—and the “conclusions” their three-stage arguments offer are not substantive philosophical theses at all, but rather nicely written nonsense; this is often called a Pyrrhonian view, whereas the belief that the *Investigations* do indeed perform philosophy (just correctly) is often called “non-Pyrrhonian” or “anti-Pyrrhonian.” We will return to this debate shortly, as its importance in the late Wittgenstein canon cannot be overstated.

But first, there are also far-reaching literary consequences of a Kripkean take on the rule-following paradox as it appears in Kafka’s fiction, and in his
penal colony in particular: Kripke’s (philosophical) solution (and he is not alone) to the rule-following paradox brings us back to the idea of a community of speakers, a Lebensform. For, argues Kripke, “of course” Wittgenstein wishes “to solve” the rule-following problem, because “the skeptical conclusion is insane and intolerable.” The way to “solve” the rule-following crisis is first—nominally—to free ourselves from “the grip of the normal presupposition that meaningful declarative sentences must purport to correspond to facts.” Once we have cleared ourselves of the disproven idea that (as the author of the Tractatus may have claimed) words and, more specifically, logical structures, correspond to facts in the world, what we have left is the idea of a language community or Lebensform. That is, a rule “works” or a language is “understood” when “each person who claims to be following a rule can be checked by others.” So by this rationale, the uninterpretability of the officer’s suicide would, in Kafka’s story, have been nullified if the officer had simply had colleagues.

But here is the problem with the language-community solution, which encapsulates once again the problem of philosophical or anti-philosophical doctrine in literature, which is that the clever authors of literature can very easily write a language community that is nonexistent or nonfunctional, and with that creation show exactly how fallible the language-community “solution” is, regardless of its untenability as an esoteric philosophical thesis, but rather because of the unreliability of the human members of a community. For example: what happens when a community does not comprise a Lebensform, a “form of life,” but rather, so to speak, a Todesform, a form of death? The prisoners in “In the Penal Colony,” all of whom are allegedly susceptible (or, at least, were susceptible) to the same “Schuld ist immer zweifellos” commandment of the officer, seem defined more than anything else by the constant threat of death that follows them around; as the officer’s description or depiction of his community would have it, every resident in Kafka’s penal colony (including, it turns out, the officer himself), seems defined by his death, by his method of death, or by the justification of his death; by the lost sacrificial capabilities of his death; or by the imminence of his death. And as such, the community is more than anything else defined by its proximity to death (and more so, even, after the officer has met his, as the death transitions from perceived sacrifice to “unmittelbarer Mord”).

The culmination of the officer’s lengthy description of how the machine works is exactly the kind of singular case of behavior (Verhalten) that would normally constitute a “form of life,” but in this situation actually constitutes a “form of death”—an execution, eine Exekution (at least when the machine’s work is executed properly [1:164]). And just as what the colony has in place of a Lebensform is, as it were, a Todesform, what could have been the officer’s community of speakers is, instead, isolation, claimed/perceived community and, most of all, silence. This is especially evident in the silent directive on what claims to be the old commandant’s gravestone, which
the explorer encounters on his retreat from the colony. The stone is inscribed, and it warns of the old commandant’s impending resurrection—which would introduce an actual, living community that could substantiate the officer’s own claims:


Here lies the Old Commandant. His adherents, who now must remain nameless, have buried him here and set this gravestone. There exists a prophecy that after a set number of years the Commandant will return, and reclaim his adherents from this house and reclaim sovereignty over the Colony. Have faith, and wait!

Leaving aside the probable reality that the officer himself could easily have written that inscription with no “community” of “Anhänger” to speak of, even if said “community” did exist, their uprising has not happened yet, and the only comfort an adherent to the officer’s (and old commandant’s) pseudo-rules may have is that of blind faith: “Glaubet und wartet!” (“Believe and wait!”). So blind faith rather than relying upon a mode of behavior (here the “form of death,” usually a “form of life”) created by a language community may now be to the rule-following paradox—which is akin to the “flat-footed” response to rule skepticism: we know how to follow a rule because we just do. The machine malfunctioned because it obviously did. It was indeed “unmittelbarer Mord” because it was.

This mentality is in fact substantiated by Wittgenstein himself when he posits that “Wenn ich der Regel folge, wähle ich nicht. Ich folge der Regel blind” (“When I follow a rule, I don’t choose. I follow the rule blind”) (§219). And this “flat-footed” response to rule skepticism is put forth eloquently by Finkelstein in his reading of Crispin Wright’s criticisms of Kripke (a terrific example of the forbidding web of secondary criticism about which Stern warned us): Finkelstein quotes Wright as arguing that Kripke’s entire line of reasoning—How do we know that by “plus” Jones meant “plus” and not “quus”? How do we know that by “arbeiten” the officer meant “arbeiten” and not “quarbeiten”?—can and should be met with the response: we just do (Glaubet, in other words). This, however, in Finkelstein’s reading, accounts (despite purported reliance on pattern) for just the kind of Platonism that Finkelstein insists is “not Wittgenstein’s view” (and fealty to “Wittgenstein’s view” is always paramount, though nobody seems to agree what it was). Rather:
The Platonist who figures in Wittgenstein’s texts is someone who first imagines that there’s a gulf between every rule and its application, and only then thinks that somehow, mysteriously, the rule (or its meaning, or something) autonomously manages to call for one activity rather than another. Once we stop thinking of words in isolation from the human lives in which they are embedded—once we give up imagining that there’s a gulf between every rule and its application—we can say, innocently, that a particular rule autonomously called for this or that.\textsuperscript{39}

Thus the action of flat-footedness, however natural it may seem (the machine malfunctioned because \textit{it just did}) carries with it exactly the kind of weddedness to abstract superconcepts (and thus by definition also philosophical theses) so many Wittgenstein critics—the Pyrrhonians especially, but even anti-Pyrrhonians would likely not come to a consensus about Platonism in the \textit{Investigations}—see as exactly the wrong way to go about reading Wittgenstein.

However, while the flat-footed response against Kripke is (arguably) relatively easy to argue against, the more calmly reasoned critiques, Stern’s or Goldfarb’s among them, make a stronger case both against Kripke’s solution and, in a broader sense, for a Pyrrhonian or at least anti-philosophy-tolerant approach. Stern, for example, acknowledges that Kripke’s reading “has certainly succeeded in focusing critical attention on the central importance of Wittgenstein’s treatment of meaning, rules and rule-following. But,” he cautions, “as an exposition of Wittgenstein’s treatment of rules and private language, it must be considered a failure.”\textsuperscript{40} The reason why, Stern reveals, has nothing at all to do with a flat-footed or “it just does” quasi-Platonism. Rather, this brings us full circle, back to the motto of (narrative or philosophical) “progress”: the second, more pressing problem with the language-community “solution” and with many readings of the \textit{Investigations} as a whole is that Kripke (and others), in Stern’s conception, does not correctly situate §201 within the \textit{Investigations}, and that philosophers who espouse that the \textit{Investigations} “advances” philosophical theses (undertakes the goal of philosophical progress), has misunderstood the text entirely.

Instead, Kripke and others have failed to situate the paradox of §201, and instead read it “straight,” that is, as if it were Wittgenstein’s “real” voice speaking, when in reality, Stern argues that

Kripke failed to see that the paradox of section 201 is a problem for the voice the author argues with, the “interlocutor.” The skeptical paradox only arises if one thinks of grasping a rule as a matter of being able to say something that explains the meaning of the words in question, for which Wittgenstein reserves the term “interpretation” (“\textit{Deuten}”). Indeed, that is the main moral of this section of
the Philosophical Investigations: the interlocutor’s view is mistaken precisely because he views interpreting as essential to grasping a rule. Thus, despite the fact that Kripke makes so much of the first paragraph of section 201, where Wittgenstein sums up “our paradox” and his initial reply to it, he never quotes or even refers to the second paragraph, where Wittgenstein replies that the paradox is due to a mistaken conception of understanding or a matter of “interpreting,” providing explanations where one substitutes one expression of a rule for another.41

The problem with Kripke’s interpretation is that despite its revolutionary method and sharp identification of some part of §201’s ability to dismantle a previously taken-for-granted semantic construct, it fails to situate itself correctly in the maze, as it were, of §201 (directly after the rule-following paradox comes the revelation that “Die Sprache ist ein Labyrinth von Wegen. Du kommst von einer Seite und kennst dich aus; du kommst von einer andern zur selben Stelle, und kennst dich nicht mehr aus” [“Language is a labyrinth of paths. You come from one side and don’t know where you are; you come from another side to the same place, and you don’t know where you are”] [§203]). Thus Kripke’s conception insists, despite its best efforts, upon the presentation of a paradox that requires a reliance on philosophical progress. Kripke thus ends up falling into the pattern of so many other Wittgenstein readers, a pattern that “treats Wittgenstein’s opposition to theses as a device that allows Wittgenstein to avoid stating the controversial theses he supposedly really believes—the theses the reader finds in Wittgenstein’s writing.”42

Wittgenstein, Stern argues, “would have regarded Kripke’s skeptical thesis and paradoxical solution as equally nonsensical consequences of a mistaken conception of language and rules.”43 This realization, in turn, has tremendous bearing in Kafka’s literary world as well. To Stern, it seems as if Kripke’s conception of and proposed solution to the rule-following paradox is the closest the mouse of a parallel-universe “Little Fable” could ever have come to successfully switching direction, of heeding the first paradigm shift while never really “grasping” that the problem is and always was that the mouse was a mouse in a maze in the first place (so, in this situation, the problem is that Kripke is a philosopher attempting to philosophize anti-philosophy in the first place). Kripke’s only possible result is to be gobbled up whole, either by critics standing in for what Wittgenstein “really” wanted or, in a more generous conception, Wittgenstein himself. Like Kafka’s fable mouse’s original problem, one of Wittgenstein’s Investigations voices also claimed that a philosophical problem had the form “Ich kenne mich nicht aus” (“I don’t know where I am”) (§123). And that—not just the content of the “problem,” but the form of “Ich kenne mich nicht aus,” the presumption of the need for reorientation in the “right” direction, and the fact that it is a philosophical problem—is what Kripke’s problem may be as well. The bearing of this idea
on Kafka’s “Penal Colony” is grimmer yet: if our problem was, for example, believing that any “progress” could be advanced in the penal colony at all—progress toward interpreting what the officer’s suicide “means,” and above all the luxury of identifying what passes for a sensical *narrative* progress(ion)—then where does that leave the story? What is the point of reading it, if our only valid result is to realize that there is no such thing as a “correct” mode or direction of advancement through it?

**An Actual Conclusion**

On the purely literal, textual level, the story (and what enigmatic sensical narration there remains in it) seems to leave us with a grim choice: either accept, in all of its faults and its complete narrative nonworkability, the language-community “solution” (the old commandant has adherents in the colony who can substantiate the officer’s stories) or operate on blind faith (that the old commandant will somehow rise again): both choices, it turns out, are the same, as they both hinge on allowing the idea that the old commandant’s adherents’ return has any merit at all; they both are simply “Glaubet und wartet!” The only other choice seems to be a non-choice: the acceptance of uninterpretability and simple removal of oneself from the situation—which is, after all, what the explorer in the penal colony does: he simply removes himself from the situation as quickly as possible, sailing away on the boat that brought him there in the first place, shaking a knotted rope at the soldier and the prisoner so that they can’t even entertain the thought of stowing away (1:195).

One glaring issue remains, however, even after the story has ended and the explorer has ceased his explorations (or at least these particular ones): if only the officer had just demonstrated an execution instead of merely describing one, if only the officer had just executed the prisoner as he originally claimed he would, then we would have had an instance of “correct” use and thus the officer’s suicide—provided that the rest of the story went in exactly the same manner as it is actually written—would have the kind of concrete meaning and metaphorical “meaning” so many of us ascribe to it. But let us not forget the most confounding element of Wittgenstein’s rule-following paradox: even a documented precedent, a prior behavior (*Verhalten*) is just that and nothing more. All behavior is merely singular incidents, every case (every *Lebensform*) different. Thus although the glaring absence of a “correct execution”—and we may be tempted to ask why Kafka has done this, to which there is no satisfactory answer—provides ammunition for the theory that the machine may never have worked “correctly,” even a “correct” precedent does nothing, in the end, to guarantee the presence of a rule. Thus even with a small amount of narrative speculation, Kafka’s provocative demonstration of the penal colony’s rule-following paradox remains unsolvable.
But what about coming at both the Kafka and Wittgenstein “problems” via yet another way of looking at Wittgenstein’s motto, “Überhaupt hat der Fortschritt das an sich, daß er viel größer ausschaut, als er wirklich ist” (“But overall, the thing about progress is that it always appears greater than it really is”)? What of the opposite of progress, which is situation? If we remember, Wittgenstein insisted that the *Investigations* were to have been seen as a collection of landscape sketches, whose readers should have no “object” other than to recognize them as such. Using this idea, we could discover just as plausibly that in Kafka’s story, the *Forschungsreisende*’s problem comes in the unfortunate presumptions of his name, which forces upon him a purpose, which we now understand could not have ended well: *forschen* (explore, research, make progress of knowledge) and *reisen* (travel, move in a particular direction). In setting up the events the way Kafka has—by placing the burden of narrative proof within an unreliable, extended description, by making the process of reading and writing of the apparatus impossible to understand without experience (*Erfahrung*), and then making said experience the uninterpretable “unmittelbarer Mord” of the officer and sudden release of the prisoner—Kafka has made sure that the entire nominal purpose of the research trip (*Forschungsreise*) is called into question; after what he has seen it is doubtful that he will have any legitimate “progress” on which to report and instead may be called into a crisis of career. What the explorer seems fated not to have recognized is the misleading pretense of his vocation—and, not coincidentally, his only identifying moniker, his entire identity within the narrative. The problem with his search for (research) “progress” was that it ended up being not nearly so great as it appeared; the problem with our expectation for narrative progress is that we failed to see exactly that expectation as the problem. By playing as he has with the entire assumption of causal narrative “reality,” by sabotaging, in effect, his execution of the narrated execution, Kafka has forced us to confront what turned out to be misguided expectations of narrative “sense” or narrative “meaning.” For if the officer’s suicide precludes interpretation and simply just is, definitive interpretive “progress” of the story must largely halt as well. Not only, it seems, is an “ordinary language” theory of narrative meaning prohibitively problematic, but the attempt to define it is invalid. The discovery of this invalidity is, then, the only progress here.