Lila Wooden

Professor Dong

HONR 105AG Art of Mathematical Proof

5 December 2024

Defying Containment:

Unifying Themes in Dostoevsky's *Notes from Underground* and Gödel's Incompleteness Theorems

In Fyodor Dostoevsky's 1864 novella *Notes from Underground*, the narrator (unnamed but generally referred to as 'the Underground Man') rails against the idea that humans will ever behave rationally. History does not look particularly rational, he argues, and life would be dreadfully boring if everything could be "calculated and tabulated" in such a way that people could only act according to their "real normal interests" (Dostoevsky 15, 19). Besides, the Underground Man continues, any calculation of advantage based solely upon material concerns misses the most "advantageous advantage" (Dostoevsky 17) of all, that is, the advantage of the ability to choose between the advantageous thing or something outrageous instead:

> "…out of sheer ingratitude, sheer spite, man[1] would play you some nasty trick. He would even risk his cakes and desire the most fatal rubbish, the most uneconomical absurdity, simply to introduce into all this positive good sense his fatal fantastic element. It is just his fantastic dreams, his vulgar folly that he will desire to retain, simply in order to prove

[1] When the Underground Man says "man would do such and such" he is speaking of humanity in general, not a specific man. Russian is a heavily grammatically gendered language, keeping the masculine form in this translation is likely closer to the original. I am referencing Constance Garnett's 1918 translation; wording may vary slightly in other editions.

> to himself—as though that were so necessary—that men are still men and not the keys of a piano[2], which the laws of nature threaten to control so completely that soon one will be able to desire nothing but by the calendar" (Dostoevsky 24).

The inspiration for this diatribe was Nikolay Chernyshevsky's *What is to be Done?* (Barstow, Scanlan), a novel about a logical utopia of 'rational egoists.' Rational egoism is a school of thought in which whatever is best for an individual is also best overall, and those who behave according to a rational assessment of their own interest naturally behave nobly and altruistically because that is what will benefit them the most (Scanlan). Thus, Chernyshevsky's utopia is a place where every action can be calculated according to one's best interest, and it is impossible to desire anything unreasonable (Barstow, Scanlan). Barstow claims that "Dostoevsky was so enraged by the simplistic solutions to complex social and human problems this work preached, that instead of writing a literary review, he delivered a 'bitter artistic response'" (Barstow).

I did not know when I first read *Notes from Underground* that it was a response to anything in particular—I simply liked the contrast of the despicable narrator (the book begins with the narrator explaining all of the ways in which he is despicable, see Dostoevsky 1-3) and the elegant way in which he presents his argument, as well as the contrast between the reasonable argument and its subject: that humanity will never be reasonable.

Each time the Underground Man concedes that some of human nature could be calculated, he introduces the "fatal fantastic element" that is the innate willfulness of humanity. The emphasis on calculation continues throughout, possibly to highlight the absurdity of expecting humans to consistently behave in an orderly manner: "… even if a man really were

[2] This metaphor is introduced on page 18: "…for what is man without desires, without choice and without free will if not a stop in an organ?"

nothing but a piano key, even if this were proved to him by natural science and mathematics, even then he would not become reasonable…" (Dostoevsky 24). Without the ability to control for obstinateness as the final rogue variable, it seems, one can never create a rational utopia: all the advantages of humanity could be accounted for, but this one most "advantageous advantage" would always remain one step ahead of the calculation, uncontrollable despite its predictability: "If you say that all this, too, can be calculated… man would purposely go mad in order to be rid of reason and gain his point!" (Dostoevsky 24).

With his continuous references to mathematics as an area of absolute certainty, it may have pleased the Underground Man to know that 67 years after he was written into existence by Dostoevsky, it would be proven that even mathematics cannot provide absolute certainty: mathematical systems will always be either inconsistent or incomplete. Math, like human nature as described by the Underground Man, contains a "fatal fantastic element" that cannot be avoided by being anticipated.

This was demonstrated by Kurt Gödel with his incompleteness theorems, published in 1931(Dawson, Doxiadis and Papadimitriou, Zach). Gödel's first incompleteness theorem showed that any axiomatic mathematical system sophisticated enough to perform basic arithmetic would be able to make mathematical statements that could be neither proven nor disproven within the system (Dawson, Zach). Gödel's second incompleteness theorem showed that an axiomatic mathematical system can never prove itself to be consistent (Dawson, Zach). These discoveries came as a shock to the mathematics community, which was hard at work trying to create a foundation for mathematics that would allow for every mathematical statement to be proven either true or false. (Dawson, Doxiadis and Papadimitriou, Zach).

In the early 1900s, mathematician David Hilbert set out to ground all of mathematics firmly in a set of axioms (Morris, Zach). The goal was to establish a mathematical system that never contradicted itself, where every mathematical statement that could be made from the parts of the system could be proven true or false using only other elements of the system (Morris, Zach). Drawing upon his understanding of algebra and geometry, Hilbert created a list of requirements that an ideal set of axioms ought to fulfill (Morris, Zach). According to Hilbert, an axiomatic system must be consistent: it should not be possible to prove both a statement and its opposite. An axiomatic system should be complete: capable of either proving or disproving any theorem that rests upon the axioms. The axioms should also be independent, that is, not superfluous or repetitive: the removal of an axiom should render the system incomplete (Morris, Zach). Hilbert succeeded in devising a set of axioms but ran into difficulties when he tried to prove their consistency without relying on another theory (Zach). For his system to be airtight, it would need to be able to prove itself, which resulted in circular arguments. One result of Godel's second incompleteness theorem was to prove that this requirement can never be met (Morris, Zach).

Bertrand Russell, a contemporary of Hilbert's, was also working on the foundation of mathematics (Doxiadis and Papadimitriou, Irvine). His approach was to ground mathematics in logic using set theory (Doxiadis and Papadimitriou). Set theory may be defined as the study of groups or sets of things that go together (Bagaria, Doxiadis and Papadimitriou, Irvine). Thus, one may have a set of plates, a set of bowls, and a set of cups, and these sets may each be a member or *element* of a set of dishes. In set theory, Russell ran straight into a paradox when he considered the set of all sets that do not contain themselves (Bagaria, Doxiadis and Papadimitriou, Irvine). If the set of all sets that do not contain themselves does not contain itself, it becomes a member of the set of all sets that do not contain themselves (Bagaria, Doxiadis and Papadimitriou, Irvine). Then, by being a member of the set of all sets that do not contain themselves, it is automatically disqualified from being a member of the set of all sets that do not contain themselves (Bagaria, Doxiadis and Papadimitriou, Irvine). Therefore, the set is a member of itself if and only if it is not a member of itself (Doxiadis and Papadimitriou). This particular contradiction was resolved by expanding the axioms of set theory, but it exposed the issue of self-reference that would later form the heart of Gödel's theorems (Bagaria, Doxiadis and Papadimitriou, Irvine).

To investigate the completeness of mathematical systems, Gödel created an elementary set of axioms and devised a way to assign unique numbers to mathematical statements created from the axioms (Dawson). This mathematical language allowed him to make mathematics talk about itself in a much more sophisticated way than Russell's self-referential sets. By encoding mathematical statements in this way, Gödel translated two statements into numeric code: "*This statement is unprovable,*" and "*the axioms of this theory do not contradict each other*" (Dawson). Expressing these statements as theorems in numeric code allowed him to check their provability and definitively determine that neither could be proven. These two statements are the basis of Gödel's first and second completeness theorems, respectively (Dawson).

Gödel's incompleteness theorems destroyed any hope of achieving the goal of mathematical perfection by showing that in mathematics, there will always be true statements that cannot be proven, and changing the rules to allow unprovable statements to be proven will simply result in more unprovable statements stemming from the new rules (Dawson, Doxiadis and Papadimitriou, Morris). Like the Underground Man's "most advantageous advantage" that resists being quantified in idealized models of human nature (Barstow, Dostoyevsky 17), Gödel's unprovable mathematical statements crop right back up even after attempting to account for them by changing the rules (Dawson, Morris). Where the Underground Man claims that attempting to rationalize human behavior would encourage irrational behavior simply for the sake of rebellion (Dostoyevsky 24), Gödel's proofs show that attempting to avoid unprovable statements by adding more axioms results in new and different unprovable statements (Dawson). Perfect self-contained systems, it seems, are an unreachable goal in both human social systems (Barstow, Dostoyevsky, Scanlan) and mathematics (Dawson, Morris). In either case, humanity and mathematics break free from their confines, humanity by being unreasonable and mathematics by being perfectly reasonable.

Though the Underground Man's angry rant and Gödel's incompleteness theorems both respond to proposals for idealized rational systems, the method and spirit of attack is different for each. The Underground Man seems intent on 'gaining his point' and exposing Chernyshevsky's literary logical utopia as absurd (Barstow, Dostoyevsky). His argument is thorough but expressive and emotional, filled with imagery and anecdotes (Dostoyevsky 16-25). It seems that in his mind he accepts that Chernyshevsky's utopia is theoretically a lovely idea, but in his heart, he knows that it is incompatible with the reality of human nature, so he calls upon his listeners to remove their idealistic heads from the clouds and look around at the way that real humans actually behave (Dostoyevsky 16 & 17). Gödel's incompleteness theorems, on the other hand, show no emotionality, just a brilliantly creative use of logic to settle an open question in mathematics (Dawson, Morris).

Thus, Dostoevsky through the Underground Man and Gödel through mathematics expose their respective perfect systems as impossible ideals. But perhaps perfection would be dreadfully boring and leave the philosopher-novelists and mathematicians alike with nothing to do.

> “Man likes to make roads and create, that is a fact beyond dispute. But why has he such a passionate love for destruction and chaos also? Tell me that!... May it not be that he loves chaos and destruction… because he is instinctively afraid of attaining his object and completing the edifice he is constructing? Who knows, perhaps he only loves that edifice from a distance, and is by no means in love with it at close quarters, perhaps he only loves building it and does not want to live in it…” (Dostoevsky 25).

What possibility would be left for literary social commentary, if a society of perfect rationality could be attained? And what would a mathematician do, if all of math could be solved? Luckily, we will never have to find out.

Works Cited

Bagaria, Joan, “Set Theory”, The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), edited by Edward N. Zalta & Uri Nodelman. https://plato.stanford.edu/archives/spr2023/entries/set-theory. Accessed 6 December 2024.

Barstow, Jane. “Dostoevsky’s ‘Notes from Underground’ versus Chernyshevsky’s ‘What Is to Be Done?’” *College Literature*, vol. 5, no. 1, 1978, pp. 24–33. JSTOR, http://www.jstor.org/stable/25111196. Accessed 4 December 2024.

Dawson, John W. “Gödel and the Limits of Logic.” *Plus Maths,* 2006. https://plus.maths.org/content/godel-and-limits-logic. Originally published in *Scientific American,* 1999. Accessed 1 December 2024.

Dostoyevsky, Fyodor. *Notes from Underground*, 1864. Translated by Constance Garnett, 1918.

Doxiadis, Apostolos and Christos Papadimitriou. *Logicomix.* Bloomsbury USA, 2009.

Morris, Rebecca. “We must know, we will know.” *Plus Maths,* 2006. https://plus.maths.org/issue41/features/morris/index.html. Accessed 4 December 2024.

Irvine, Andrew David and Harry Deutsch, “Russell’s Paradox.” *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2021/entries/russell-paradox. Accessed 5 December 2024.

Scanlan, James P. “The Case against Rational Egoism in Dostoevsky’s ‘Notes from

Underground.'" *Journal of the History of Ideas*, vol. 60, no. 3, 1999, pp. 549–67. JSTOR, https://doi.org/10.2307/3654018. Accessed 5 Dec. 2024.

Zach, Richard, "Hilbert's Program." *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), edited by Edward N. Zalta & Uri Nodelman. https://plato.stanford.edu/archives/win2023/entries/hilbert-program. Accessed 4 December 2024.