SYLLABUS
Gödel's Theorems
Phil 93902
TR 3:30
143 Debartolo Hall

Michael Detlefsen (mdetlef1@nd.edu), 408 Malloy Hall
Curtis Franks (cfranks@nd.edu), 411 Malloy Hall

Fall 2007

I. Course Objectives

- To acquaint the student with the basic ideas, methods and techniques behind one of the most striking scientific achievements of the twentieth century, the discovery of Gödel's incompleteness theorems.
- To engage the student in productive thinking concerning the broader significance of these theorems.

II. Texts

- *Kurt Gödel: Collected Works*, vol. I, Feferman *et al.* (eds.), Oxford, Oxford U. Press, 1986.

III. Organization and Projected Schedule

The course will divided into the following units, each estimated to run the indicated number of meetings. We plan to cover all of units A through G. We will cover the additional units according to the interests of the seminar participants.

A. Historical and conceptual background and basic terminology. Two meetings.

B. Basic mathematical background. Proof of the first theorem (G1). Four to five meetings.

C. Modifications of the proof of the first theorem, esp. Rosser's modification. One to two meetings.

D. Proof of the second theorem (G2). Central task is to clarify the proofs of the second and third of the Hilbert-Bernays-Löb Derivability Conditions (HBL Conditions). Two meetings.

E. Alternative proofs of G2. Jeroslow's 1973 proof. Feferman's 1960 proof. Freidman's recent proof. [Note: We may discuss Jeroslow's and Feferman's proofs in inverse historical order because Jeroslow's proof is in some ways the less radical departure from the Hilbert-Bernays-Löb proof.] Three meetings.

F. Gödel's theorems and the representation of metamathematical concepts. Are the HBL Conditions needed for the proper representation of provability and consistency? Are they enough? Two to three meetings.

G. Gödel's theorems and Hilbert's Program. Four to five meetings.

H. Gödel's theorems and the nature of mathematical proof. One meeting.

I. Gödel's theorems and the mechanizability of mind. Three to four meetings.

J. Incompleteness and complexity. Two to three meetings.

K. Incompleteness and length of proofs. Two to three meetings.

IV. Readings

Abbreviations. **CW x**=*Kurt Gödel: Collected Works*, vol. **x**; **U**=*The Undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions*; **G1**=Gödel's first incompleteness theorem; **G2**=Gödel's second incompleteness theorem.

Order of reading. In the following, the order given is but one sensible order. Also, some items are listed in more than one unit. They don't need to be read repeatedly of course . . . ;-) Items marked with † are optional, though all are informative and some have been influential.

Unit A

1. Detlefsen, Michael 'Gödel's theorems', *Routledge Encyclopedia of Philosophy.* Referred to as '**Detlefsen 1997**'. Surveys technical and philosophical basics concerning Gödel's theorems. Online through ND library.

2. Hilbert, David 'Knowledge of nature and logic', English translation of German audio text of 1930 radio address by Hilbert entitled 'Naturerkennen und Logik'. Historic statement of Hilbert's epistemic optimism—'Wir müssen wissen. Wir werden wissen.' Directed against the epistemic pessimism represented by the *Ignorabimus* doctrine of the du Bois-Reymonds (and others). This pessimism had made itself felt not only in foundational thinking but also in ideas for reforming the curriculum of German universities. Online at: Hilbert's 1930 Radio Address.

3. McCarty, David 'Problems and Riddles: Hilbert and the du Bois-Reymonds', *Synthese* 147 (2005): 63–79. Discusses the *Ignorabiumusstreit* of the late nineteenth and early twentieth centuries. Online through ND library.

4. Schlick, Moritz 'Unanswerable questions', *The Philosopher* 13 (1936). Raises questions concerning the cogency of the *Ignorabimus* doctrine. Available online at: Schlick 1936.

5. Curry, Haskell *Foundations of Mathematical Logic*, New York, McGraw-Hill, 1963. Ch. 1, §B. The Logical Antinomies. Brief summary of principal antinomies of importance to 20th century logic and foundations. Electronic reserve.

6. Russell, Bertrand *Principia Mathematica*, vol. I, Introduction. Ch. II, §VIII. The Contradictions. Historically important listing of main paradoxes figuring in early 20th century foundational discussions. Electronic reserve.

7. Ramsey, Frank *The foundations of mathematics*, Totowa, NJ, Littlefield, Adams & Co., 1965. Ch. I, pp. 20–21. Historically important statement of division of antinomies into *epistemological* and *logical* types. Mirrored earlier distinction by Peano (who used 'linguistic' rather than 'epistemological'). Electronic reserve.

8. Gödel, Kurt 'Review of *Carnap 1934*, CW I, 389. Describes Carnap's derivation of formally undecidable sentences from antinomies of 'second kind' (i.e. so-called 'linguistic' or 'epistemological' antinomies). Notes (i) Carnap's conclusion—'although everything mathematical is formalizable, it is . . . impossible to formalize all of mathematics in a *single* formal system'—was held by intuitionists and (ii) Tarski offered same view of antinomies of second class in 1932,1933. Referred to as '**Gödel 1935b**'.

Unit B

1. Detlefsen, Michael 'Gödel's theorems', *Routledge Encyclopedia of Philosophy*. Online through ND library.

2. Pp. 126–139 of the introductory note by Kleene to *1930b*, *1931* and *1932b*, CW I. Useful description of Gödel 1931. To be read in conjunction with Gödel 1931.

3. Gödel, Kurt 'On undecidable sentences', in CW III. Summary of basic ideas and arguments of **Gödel 1931**. Referred to as '**Gödel 1931?**'. Electronic reserve.

4. Gödel, Kurt 'Some metamathematical results on completeness and consistency', CW I, 141, 143. Announcement of Gödel's theorems.

5. Rosser, John Barkley 'An informal exposition of proofs of Gödel's theorem and Church's theorem', *Journal of Symbolic Logic* 4 (1939): 53–60. Useful informal exposition of chief ideas behind Gödel's theorems and the variety of proofs of them known in the late 1930s. Pays particular attention to 'hidden assumptions' in Gödel's 1931 concerning the conditions under which G1 applies to a 'logic' (i.e. a system in a given language). Online through ND library.

6. Gödel, Kurt 'On formally undecidable propositions of *Principia Mathematica* and related systems I', CW I, 144–195. Referred to as '**Gödel 1931**'. Classic paper in which Gödel's theorems and their proofs were first set out in detail. [Note: A more 'efficient' proof of G1, in friendlier notation and more up-to-date and exact terminology (e.g. 'primitive recursive', where Gödel uses only 'recursive') is in Smorynski 1977.]

7. Smorynski, Craig 'The incompleteness theorems', *Handbook of Mathematical Logic*, Amsterdam, North-Holland Publishers, 1977. Useful introduction to Gödel's theorems and the larger body of work they inspired. Of particular concern for this unit is the proof of G1. Electronic reserve.

Unit C

1. † Rosser, J. B. 'Extensions of some theorems of Gödel and Church', *Journal of Symbolic Logic* 1 (1936): 87–91. Important paper in which Rosser eliminates the ω-consistency condition of Gödel's original proof of G1 in favor of simple consistency. The consistency sentence for a theory T constructed in the usual way from Rosser's provability formula for T turns out to be provable in T. The notation used in the paper may be off-putting. Smorynski 1977 gives more perspicuous treatment. Online through ND library.

2. Smorynski, Craig 'The incompleteness theorems', *Handbook of Mathematical Logic*, Amsterdam, North-Holland Publishers, 1977. Useful introduction to Gödel's theorems and the larger body of work inspired by them. Of particular interest for this unit is the proof of Rosser's generalization of G1. Electronic reserve.

Unit D

1. Detlefsen, Michael 'Gödel's theorems', *Routledge Encyclopedia of Philosophy.* Online through ND library.

2. Smorynski, Craig 'The incompleteness theorems', *Handbook of Mathematical Logic*, Amsterdam, North-Holland Publishers, 1977. Of particular importance for this unit are those parts concerning the proof of G2. Electronic reserve.

3. No general reading assignment covering the proof of G2. We leave it to the presenters to present one. Raymond Monk's *Mathematical Logic* (New York, Springer-Verlag, 1976), Chapter 17, gives fairly detailed treatment in English (and in a general arithmetic setting) of the proof of G2 (Cor. 17.3). Lemma 17.11 is one of the few proofs of the third Hilbert-Bernays-Löb Derivability Condition (condition (ii) in Monk's list on p. 299) [Note: The Δ or 'numeralization' operator of Monk's discussion is defined inductively on p. 235.]

4. †Solovay, Robert 'Provability interpretations of modal logic', *Israel Journal of Mathematics* 25 (1976), no. 3-4, 287–304. Presents many technical extensions and refinements of G1, G2, and also Gödel's completeness theorem based on the modal logical representation of formal provability. The essay has a rather high level of overhead, but we would like to present one of its several results as a possible strengthening of G2. The theorem has relevance to Units G and L, but is interesting in its own right for its insights into the mechanics of G2. (In fact, once distilled, the theorem is not too difficult and would make a good presentation topic.)

Unit E

1. Detlefsen, Michael 'Gödel's theorems', *Routledge Encyclopedia of Philosophy.* Online through ND library.

2. Feferman, Solomon 'My route to arithmetization', *Theoria* 63 (1997): 168-181. Prepublication electronic version available on Feferman's website: Feferman on Arithmetization. Useful statement of the motives and reasoning that led Feferman to his alternative proof of G2.[1]

3. Franks, Curtis 'Intensionality in Metamathematics', unpublished manuscript. Analyzes central aspects of Feferman 1960. To be circulated.

4. Friedman, Harvey. 'Fromal (sic!) statements of Goedel's second incompleteness theorem'. Argues that there (i) is a 'formal' version of G2 whose proof does not appeal to derivability conditions (or similar 'intensional' constraints on choice of provability predicates), and which nonetheless (ii) implies a suitable 'informal' version of the theorem (i.e. a statement which asserts or implies the inability of various formal theories to prove their own consistency). Online at Papers of Harvey Friedman, item #56.

[1] We are not generally assigning Feferman 1960 or Jeroslow 1973. Nonetheless, students are free, and even encouraged, to choose them as presentation topics.

Unit F

1. Detlefsen, Michael 'Gödel's theorems', *Routledge Encyclopedia of Philosophy.* Online through ND library.

2. Gödel, Kurt 'Some remarks on the undecidability results', unpublished remarks originally intended for publication in *Dialectica.* First published in CW II. Introductory note gives useful background information. Of particular interest is Gödel's suggestion that so-called 'outer' consistency is the preferred form of consistency and that the unprovability of outer consistency can be proved without appeal to the derivability conditions used in the usual proof of G2. Referred to as '**Gödel 1972a**'. Electronic reserve.

3. Kleene, Stephen C. *Introduction to Metamathematics*, Wolters–Noordhoff & North-Holland Publishing, 1971. Reprinting of 1952 original. Ch. VIII, §42 (cf. esp. p. 211). Brief argument by Kleene to the effect G1 by itself is sufficient to establish the unprovability of the consistency of T in T. Electronic reserve.

4. Wang, Hao *A survey of mathematical logic*, Science Press (Peking) & North-Holland Publishing (Amsterdam), 1964. Ch. XIII, §1 (esp. pp. 338-339). Summary of proof of G2 and a challenge to Kleene's argument that G1 by itself establishes the unprovability of the consistency of T in T. Electronic reserve.

5. Lacey, Hugh and Geoffrey Joseph 'What the Gödel Formula says', *Mind* 77 (1967): 77–83. Classical paper from the 1960s (others mentioned in its footnotes) which tries to make precise sense of the helpful intuition that the Gödel sentence expresses its own unprovability. Authors distinguish the formal language of an arithmetical system (P), the language of informal arithmetic (A), and the informal language of P's metatheory (M) and conclude that (i) no sentence in any of these languages expresses its own unprovability, and that (ii) the idea that 'the Gödel sentence' does is based on a confusion of these languages. Online through ND library.

6. Milne, Peter 'On Gödel Sentences and What They Say', *Philosophia Mathematica* 15 (June 2007): 193–226. Renews the criticism from Lacey and Joseph, *et al.*, against generally regarding Gödel sentences as statements asserting their own unprovability. Proposes a way to constrain the construction of such a sentence in order to get appropriate self-referentiality. Also touches on connections between this issue and the meaning of G2. A considerable amount of this analysis turns on the differences between two ways to prove G1: by diagonalizing 'outside the theory' and arithmetizing the diagonal function together with the proof predicate, or by diagonalizing 'inside' the theory. Online through ND library.

7. Detlefsen, Michael 'What does Gödel's second theorem say?', *Philosophia Mathematica* 9 (2001): 37–71. Electronic reserve. Discussion of the conditions under which a

formula of a formal system S may rightly be regarded as 'expressing' the consistency of a formal system T. Considers as well the question of how these conditions may compare to conditions known to guarantee a G2-type result. [Note: This issue of *Philosophia Mathematica* is available online through the ND library but the article can not be downloaded from there. The paper is thus on electronic reserve.]

8. Roeper, Peter 'Giving an account of provability within a theory', *Philosophia Mathematica* 11 (2003): 332–340. Critical discussion of Detlefsen 2001. Argues that the Hilbert-Bernays-Löb Derivability Conditions *are* necessary for the proper representation in T of the notion of *provability-in-T*. Online through ND library.

Unit G

1. Detlefsen, Michael 'Hilbert's programme and formalism', in *Routledge Encyclopedia of Philosophy.* Discussion of the central features of Hilbert's Program and the bearing of Gödel's theorems on it. Online through ND library.

2. Detlefsen, Michael 'Philosophy of Mathematics in the 20th Century', in *Philosophy of science, logic and mathematics in the 20th century*, pp. 50–121 of vol. IX of the *Routledge History of Philosophy.* London, Routledge, 1996. Read pp. 76–87, which discuss Hilbert's Program. Electronic reserve.

3. Detlefsen, Michael 'Ideals of completeness', unpublished manuscript. Historico-philosophical survey of various ideals of completeness that have featured significantly in the history of mathematics. To be circulated.

4. Hilbert, David 'Neubegründung der Mathematik. Erste Mitteilung', *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1 (1922): 157–177. English trans. in P. Mancosu (ed.), *From Brouwer to Hilbert*, Oxford, Oxford U. Press, 1998. First systematic statement of Hilbert's Program. Among the clearest technical statements of Hilbert's Program. Also discusses how it differs from other foundational schemes. Gives (p. 202) summary characterization of finitary reasoning that was repeated by Hilbert in other essays. Electronic reserve.

5. Hilbert, David 'Über das Unendliche', *Mathematische Annalen* 95 (1926): 161–190. English trans. in van Heijenoort (ed.) *From Frege to Gödel*, Boston, Harvard U. Press, 1967. Perhaps the most comprehensive philosophical statement of Hilbert's mature foundational viewpoint. Electronic reserve.

6. Hilbert, David 'Die Grundlagen der Mathematik', *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 6 (1928): 65–85. English trans. in van Heijenoort (ed.) *From Frege to Gödel.* Mature philosophical statement of Hilbert's Program written as sequel to the 1926 paper. Replies to intuitionist criticisms of Hilbert's ideas. Electronic reserve.

7. Gödel, Kurt 'Discussion on providing a foundation for mathematics', CW I, 201, 203, 205. Brief statement of difference between consistency and soundness proofs. Postscript emphasizes that Gödel's theorems leave the viability of Hilbert's Program open since it was (is?) not certain that finitary reasoning is formalizable in the system(s) covered by the proof.

8. Gödel, Kurt 'Lecture at Zilsel's', CW III. Retrospective statement of the significance of G1 and G2. Discusses epistemological aims of foundational investigation and comments on attempts to salvage Hilbert's program. Referred to as '**Gödel 1938a**'. Electronic reserve.

9. †Mancosu, Paolo 'Hilbert and Bernays on Metamathematics', pp. 149–188 in Mancosu (ed. & trans.) *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s.* Useful exposition of Hilbert's foundational ideas, some of which were perfected through his collaboration with Bernays. Electronic reserve.

10. Franks, Curtis 'David Hilbert's naturalism', excerpt from *Mathematics Speaks for Itself* dissertation 2006. We will circulate chapter 1 'David Hilbert's naturalism' as well as a short portion of chapter 2.

11. Raatikainen, Panu 'Hilbert's program revisited', *Synthese* 137 (2003): 157–177. Discussion of recent work on Hilbert's Program and the effects of Gödel's theorems (and related results) on it. Online through ND library.

Unit H

1. Myhill, John 'Some remarks on the notion of proof', *Journal of Philosophy* 62 (1960): 461–471. Discussion of the significance of Gödel's theorems for the formalization of informal notions of provability. Suggests that proof-concepts may inevitably induce some type of hierarchical ordering on proofs. Online through ND library.

Unit I

1. Gödel, Kurt 'Some basic theorems on the foundations of mathematics and their implications', CW III. Gödel's famous Gibbs lecture in which he indicates a possible link between his incompleteness and the age-old philosophical question concerning the mechanizability of human reason. Referred to as '**Gödel 1951**'. Electronic reserve.

2. Lucas, J. R. 'Minds, machines and Gödel', *Philosophy* 36 (1961): 112–127. Widely discussed argument that Gödel's theorems refute mechanist views of human reasoning. Electronic reserve.

3. Benacerraf, Paul 'God, the devil, and Gödel', *The Monist* 51 (1967): 9–32. Discussion, mainly critical, of Lucas' argument. Electronic reserve.

4. Feferman, Solomon ‘Penrose’s Gödelian argument’, *PSYCHE* 2 (1996): 21–32. Critical discussion of Penrose’s recent attempts to show that G1 (and/or G2) have anti-mechanist bearing. Online through ND library.

5. Penrose, Roger ‘Beyond the Doubting of a Shadow: A Reply to Commentaries on Shadows of the Mind’, *PSYCHE* 2 (1996). Reply to Feferman. Online through ND library.

6. Detlefsen, Michael, Critical essay on Roger Penrose’s *The Emperor’s New Mind*, *Shadows of the Mind* and *The Large, the Small and the Human Mind*, *Studies in the History and Philosophy of Modern Physics* 29 (1998): 123–36. Critical discussion of Penrose’s attempts to revive Lucas-style anti-mechanist arguments. Online through ND library.

Unit J

1. Chaitin, Gregory ‘Gödel’s theorem and information’, *International Journal of Theoretical Physics* 22 (1982): 941–954. Argues that G1 limits the extent to which (or respects in which) information can be ‘compressed’ *via* formal axiomatization. Online on Chaitin’s website at: Gödel’s theorem & Information.

2. Chaitin, Gregory ‘A random walk in arithmetic’, *New Scientist* 125 (no. 1709) (Mar 24, 1990): 44–46. Argues that Gödel’s theorems, and certain (similar?) results of the author, show that, in a serious sense, there is an element of randomness in the realm of arithmetical truth. Online on Chaitin’s website at: Random Walk.

3. Raatikainen, Panu Review of *Exploring randomness* and *The Unknowable*, *Notices of the AMS* 48 (9) (2001): 992–996. Critical discussion of Chaitin’s claims concerning the relation of his work to the older work of Gödel, Church and Turing and of its larger significance. Online through ND library.

Unit K

1. Gödel, Kurt ‘On the lengths of proofs’, CWI, 397–9. Claims (without proof) that it is possible to dramatically reduce the lengths of proofs by moving from a system of lower type (e.g. quantifiers ranging over individuals) to a system of higher type (e.g. quantifiers ranging over sets and/or functions of individuals).

2. Dawson, John ‘The Gödel incompleteness theorem from a length-of-proof perspective’, *American Mathematical Monthly* 86 (1979): 740–747. Brief but useful discussion of a cryptic statement of Gödel’s in his 1936 paper ‘On the length of proofs’. Traces effects of Gödel’s paper in work of Kreisel’s and Parikh’s in the 1960s and 1970s. Online through ND library.

3. Parikh, Rohit ‘Existence and feasibility in arithemtic’, *JSL* 36, 1971: 494–508. Takes as a starting point that successor, addition and multiplication are ‘feasibly’ computable functions, but that exponentiation (i.e. x^y) is not. Shows that in a certain sense of ‘concrete’, concrete subsystems of PA cannot prove the existence of exponentiation functions. Also constructs some inconsistent theories whose only problematic proofs are ‘unfeasibly’ long and demonstrates certain advantages of working with such theories. Discusses consequences of adopting a thoroughly ‘concrete’ viewpoint in arithmetic. Online through ND library.

4. Parikh, Rohit ‘Some results on the lengths of proofs’, *Transactions of the American Mathematical Society* 177, 29–36. Presents analog of Gödel’s theorem on the length of proofs for a system of arithmetic (PA^*) that includes full induction and treats addition and multiplication as ternary relations. Shows that a natural coding of the relation(s) ‘. . . is a proof in T of at most n lines’ is decidable in PA^* for a variety of systems T. Online through ND library.

5. Pudlák, Pavel ‘On the lengths of proofs of consistency: a survey of results’, *Annals of the Kurt Gödel Society*, vol. 2, 1996, pp. 65-86. A survey of progress on questions about proof length from 1979 onward. Of particular interest is the treatment of the question, first posed by Gödel, first systematically investigated by Friedman (in unpublished work from 1979), and here nicely sharpened by Pudlák, ‘What is the minimal length of a proof of $Con_T(n)$?’ The last formula says that there is no proof in T of a contradiction whose length is shorter than n. Thus, if the answer is some $m < n$, then one has a sort of feasible consistency result: One can prove in a feasible way that no feasible proofs in T are of contraditions. On the other hand, if the answer is n or greater, then this can be seen as a type of strengthening of G2. Available online in the *Surveys and general exposition* section of Bibliography of Pavel Pudlak.

V. Course Requirements and Grading

Homework is a problem in a course that is not a mathematics course but requires knowledge of a certain body of mathematical ideas, methods and techniques. Our decision this time around is to dispense with the usual type of homework problems. We’d rather have you invest your mathematical time and effort (and practice in teaching) in presentations that bear directly on the content of the course. Proper preparation for and execution of these presentations will require learning the technical material that’s essential to the course. Thorough preparation and active participation are thus key. As concerns particulars . . .

1. Each student will be asked to give two or more presentations to the seminar on topics given below. Volunteers for topics are preferred, but starred topics will be assigned if there aren’t volunteers. Sizes and make-ups of presentation teams will be worked out case-by-case.

2. Presentation projects will typically involve both an in-class presentation and a written report. Reports should not be transcripts of the oral presentations, nor should they be mere sketches of them. They should be efficiently written but should also give enough detail to fully convey the thinking behind a proof or argument.

3. Once we've received the written report/summary for a presentation, we'll identify any changes that need to be made and inform the presenter(s) of these. We'll also give a date by which the changes are to be made. The final report, to be sent to us electronically by that date, should incorporate these changes. When that has been satisfactorily done, the written presentation will be circulated to the class.

4. Coverage of non-starred topics for which no one volunteers will be minimal.

5. A student may volunteer to do more than two presentations. Time permitting, these will be honored. The quality of these presentations will be reflected in the course grade.

6. ***Anyone wanting to volunteer for Unit B*** should let us know by **August 30th**. All other choices for presentations should be given to us by **September 4th**.

Each student will also write a paper on one of the paper topics listed under heading VI below or on a topic of the student's choice approved in advance by the instructors. This paper will be due **Wednesday, December 12th**.

The in-class presentations and write-ups will count for $30-40\%$ of a student's grade. Class participation will count for another 10%. The course paper will count for $50-60\%$.

VI. Presentation and Paper Topics

I. Presentation Topics

Unit B*: The basic proof of the first incompleteness theorem. This will require giving a fairly thorough treatment of each of the following:

1. The basics of Gödel numbering.

2. The representation of metamathematical properties and relations by arithmetical properties and relations.

   (a) Representing metamathematical properties and notions by arithmetical formulas.

   (b) Proof that all primitive recursive relations (sets) are strongly representable in basic arithmetic.

   (c) Proof that all recursively enumerable sets are weakly representable in basic arithmetics

(d) The Diagonal Lemma.

Unit C*: Rosser's 1936 refinement of G1. Rosser's proof. Proof of the provabilitiy of the Rosser consistency formula (i.e., the formula that is defined from Rosser's provability formula in the same way that Gödel's consistency formula is defined from his provability formula). How did Rosser's proof change the usual proof of G1, and what's the significance of these changes? Of what, if any, significance is the provability of the Rosser consistency formula?

Unit D*: The proof of the second incompleteness theorem. This should focus particularly on giving the proofs of the HBL Conditions More specifically, it should focus on the proofs of the second and third conditions. The proof (to be covered in presentations pertaining to B) that all recursively enumerable sets are weakly representable implies the first condition.

Unit E: Jeroslow's 1973 proof of G2 from a reduced set of derivability conditions. Special attention should be given to Jeroslow's particular fixed point construction (viz. his so-called *literal* fixed point) and how it enables one to avoid appeal to the second HBL Condition. Consideration should also be given to the possible significance of Jeroslow's 'reduction' of the HBL Conditions.

Unit E: Feferman's alternative proof of G2 (cf. Feferman *1960*, Feferman *My route to arithmetization*, pdf available on request). Particular attention should be paid to the central notions of PR- and RE-formulas and their roles in representing metamathematical notions. What was the motivation for Feferman's alternative? Does it provide a philosophically significant alternative to proofs that invoke the HBL Conditions?

Unit F: Gödel's remarks on so-called 'outer' consistency. Does focusing on 'outer' consistency rather than ordinary consistency justify one in ignoring the defense of the Derivability Conditions when considering the effects of Gödel's theorems on Hilbert's Program (and related philosophical questions)?

Unit F: Milne's discussion of the semantic significance of the difference(s) between diagonalizing 'inside' and 'outside' a theory. Sketch the salient features of each way of proving G1, and explain the restricted type II Gödel sentence Milne produces and how it is supposed to express its own unprovability.

Unit G: Myhill's argument concerning the effects of G1 on the formalization of proof. Do Gödel's theorems imply that there is an ineliminable need for some type of hierarchical structuring of mathematical proofs?

Unit I: The Lucas-Penrose argument that Gödel's theorems imply the non-mechanizability of human reasoning.

Unit J: Chaitin's approach to incompleteness. Special attention should be paid to the statement and explanation of Chaitin's 'information-theoretic incompleteness theorem'—the theorem that for every formal system F, there is a finite constant c such that F cannot prove any true statement of the form '$K(w) > c$', even though '$K(w) > c$' is true for infinitely many w. (Here '$K(w)$' stands for the Kolmogorov complexity of the number w.) In addition to stating and explaining the above theorem, the chief task is to identify those things (e.g. those features of the system F) that determine the value of c. In this connection, pay particular attention to the question of what if any role the complexity of F (conceived in an interesting, relevant way, which you should state) plays in determining the value of c.

Unit K: Dawson's unpacking of Gödel's claim about 'speed-up' and Parikh's remarks [1971] connecting speed-up with 'feasibility'.

II. Paper Topics

The following topics are possible topics for course papers. You are also free to choose other topics provided that your topic is cleared by the instructors.

1. In the text of his 1951 Gibbs Lecture, Gödel stated his first theorem informally as follows:

> ... *whatever well-defined system of axioms and rules of inference may be chosen, there always exist diophantine problems of the type described which are undecidable by these axioms and rules, provided only that no false propositions of this type are derivable.*
>
> CW II, 308

He then went on to elaborate his notion of a 'well-defined' system of axioms and rules of inference as follows:

> If I speak of a well-defined system of axioms and rules here, this only means that it must be possible actually to write the axioms down in some precise formalism or, if their number is infinite, a finite procedure for writing them down one after the other must be given. Likewise the rules of inference are to be such that, given any premises, either the conclusion (by any one of the rules of inference) can be written down, or it can be ascertained that there exists no immediate conclusion by the rule of inference under consideration. This requirement for the rules and axioms is equivalent to the requirement that it should be possible to build a finite machine, in the precise sense of a 'Turing machine', which will write down all the consequences of the axioms one after the other.
>
> CW II, 308

Discuss and evaluate the notion of 'well-definedness' that Gödel appeals to in the above. Consider, in particular, what its epistemological end(s) might be. More particularly still, consider whether (and, if so, how) it might plausibly be argued that one who rightly claims to know that all propositions belonging to a given set of propositions are true must be able 'actually to write the [propositions] down in some precise formalism or, if their number is infinite, [to state] a finite procedure for writing them down'. If you think Gödel is wrong in this claim, describe and defend what you regard as a more satisfactory alternative standard of definition for totalities (including infinite totalities) of which one may justifedly assert the truth of all members.

2. In his paper 'Some remarks on the notion of proof', John Myhill raises the possibility of interpreting Gödel's incompleteness theorem as having the following implication (cf. p. 461):

> ...that for any [correct, MD†] formal system S, adequate for the arithmetic of the natural numbers, there are *correct inferences* which cannot be formalized in S. This is problematic because no one has ever given an explanation of what 'correct inference' means, and it is suggestive because it challenges us to attempt such an explanation.

[Note †: The insertion of the word 'correct' at this juncture is justified by the reiteration Myhill gives of his claim on p. 462.]

Later (cf. p. 463), Myhill elaborates his earlier contention by writing:

> I am asserting that there is an *absolute* sense of 'provable', neither syntactical nor semantical nor psychological, and that in this sense of 'provable', Gödel (sic!) undecidable statements are provable. The proof which I assert not to be formalizable in elementary arithmetic is as follows: The axioms of elementary arithmetic are true, and the rules of inference are truth-preserving. Therefore every theorem of elementary arithmetic is true. Therefore '0=1' is not a theorem of elementary arithmetic. Therefore a certain statement p (the arithmetization of the statement that '0=1' is not a theorem) is true.

Give a clear statement of Myhill's arguments for these claims. Then analyze and evaluate them. If you think his arguments fail but are correctable, present and defend what you take the best such correction to be. If you think there is no correct argument from Gödel's theorems to Myhill's conclusions, defend that view.

3. In his 1951 Gibbs Lecture, Gödel argued as follows (cf. CW III, 310):

> ...if the human mind were equivalent to a finite machine, then objective mathematics not only would be incompletable in the sense of not being contained

> in any well-defined axiomatic system, but moreover there would exist *absolutely* unsolvable diophantine problems of the type described above, where the epithet 'absolutely' means that they would be undecidable, not just within some particular axiomatic system, but by *any* mathematical proof the human mind can conceive. So the following disjunctive conclusion is inevitable: *Either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems of the type specified* (where the case that both terms of the disjunction are true is not excluded, so that there are, strictly speaking, three alternatives). It is this mathematically established fact which seems to me of great philosophical interest.

State, discuss and evaluate Gödel's argument. If you think it's basically successful, give a clear and careful statement of its premises and conclusion and a defense of its premises and reasoning. If you think it's unsuccessful, state why and give a reasoned view of what the prospects for its successful modification are.

4. In his 1994 paper 'On the incompleteness theorems' (*JSL* 59 (1994): 1414–1419), Henriyk Kotlarski describes the diagonalization lemma as 'intuitively unclear' (1414). In a later paper ('The incompleteness theorems after 70 years', a paper given May 28th–June 1st, 2001 at the Tarski Centenary Conference), he wrote that the diagonalization lemma, though 'very intuitive in the natural language, is highly unintuitive in formal theories like Peano arithmetic'. (cf. Introduction).

In 1989, George Boolos published what he described as a 'new' proof of Gödel's incompleteness theorem, one which, he said, avoided appeal to the diagonalization lemma. Specifically, he wrote (cf. p. 387, in the reprinting in *LL&L*):

> Both our proof and the standard one make use of Gödel numbering. Moreover, the unprovable truths in our proof and in the standard one can both be seen as obtained by the substitution of a name for a number in a certain crucial formula. There is, however, an important distinction between the two proofs. In the usual proof, the number whose name is substituted is the code for the formula into which it is substituted; in ours, it is the unique number of which the formula is *true*. **In view of this distinction, it seems justified to say that our proof, unlike the usual one, does not involve *diagonalization*.**

Clarify and evaluate both Kotlarski's and Boolos' claims. Say specifically what it is or arguably might be about the diagonalization lemma's 'unintuitive' character that would make is something we'd want to avoid in proving Gödel's theorem(s). In addition, give a clear statement of Boolos' argument for the claim that his proof *does* avoid diagonalization, evaluate that argument and say whether it avoids the aspects of diagonalizaton that Kotlarski and others have thought make it undesirable.

VII. Supplementary Reading

1. Bauer-Mengelberg, Review: *The Undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions*, *Journal of Symbolic Logic* 31 (1966): 484–494. Corrects many misprints and other infelicities of production in **U**. Online through ND library.

2. Beth, Evert W. *The foundations of mathematics*, Amsterdam, North-Holland, 2nd. rev. ed. 1968. Ch. 17, The Paradoxes. §§154–171. Useful summary discussion of the paradoxes with more historical background than is given in either Whitehead and Russell, Ramsey or Curry. Electronic reserve.

3. Boolos, George Review of 'Minds, machines and Gödel' (J. R. Lucas); 'God, the devil, and Gödel' (P. Benacerraf), *Journal of Symbolic Logic* 33 (1968): 613–615. Online through ND library.

4. Boolos, George 'A new proof of the Gödel incompleteness theorem', *Notices of the American Mathematical Society* 36 (1989): 388–390. Reprinted in *Logic, Logic and Logic*, Harvard U. Press, 1998. Referred to as '**Boolos 1989**'. Proof of G1 based on the Berry paradox rather than the Richard or Epimenides paradoxes. Online through ND library.

5. Boolos, George, Letter from George Boolos, *Notices of the American Mathematical Society* 36 (1989): 676. Reprinted in *Logic, Logic and Logic*, Harvard U. Press, 1998 (as appendix to the previous paper). Online through ND library. Referred to as '**Boolos 1989a**'.

6. Buss, Samuel, Review: John Dawson's *Logical Dilemmas: The life and work of Kurt Gödel*, *SIAM Review* 40 (1998): 397–400. In addition to reviewing Dawson's book, it gives a brief but useful overview of the larger body of Gödel's work. Online through ND library.

7. Calude, Cristian 'Incompleteness, Complexity, Randomness and Beyond', *Minds and Machines* 12 (2002): 503–517. Largely sympathetic discussion of Chaitin's work by one of his collaborators. Has a somewhat higher ratio of speculation to argument than is perhaps ideal. Online through ND library.

8. Chaitin, Gregory 'Information-theoretic limitations of formal systems', *Journal of the ACM* 21 (1974): 403–424. Early paper in which the Kolmogorov-Chaitin definition of complexity was related to the general phenomenon of incompleteness of formal systems. Online at: Informational limits of formal systems.

9. Chaitin, Gregory 'Randomness and mathematical proof', *Scientific American* 232 (5) (1975): 47–52. Popular discussion of the significance of the work of Chaitin and others in algorithmic information theory for the formalizability of mathematical proof. Online at: Randomness & Proof.

10. Chaitin, Gregory 'Randomness and Gödel's theorem', *Mondes en developpement* 54–55 (1986): 125–128. Popular discussion of the connection between algorithmic information theory and Gödelian incompleteness. Online at: Randomness & Gödel's Theorem.

11. Chaitin, Gregory 'Randomness in arithmetic', *Scientific American* 259 (1) (1988): 80–85. Popular discussion of the connections Chaitin sees between algorithmic information theory and questions concerning the solvability of arithmetical problems. Online at: Randomness in Arithmetic.

12. Church, Alonzo 'Comparison of Russell's resolution of the semantical antinomies with that of Tarski,' *Journal of Symbolic Logic* 41 (1976): 747–760. Interesting discussion of differences separating Russell's and Tarski's approaches to the resolution of the so-called 'semantical antinomies'. Online through ND library.

13. Cleslinski, Cezary 'Heterologicality and Incompleteness', *Mathematical Logic Quarterly* 48 (2002): 105–110. Semantical proof of G2 based on the Grelling-Nelson paradox. Online through ND library.

14. Davis, Martin 'Hilbert's Tenth Problem is Unsolvable', *American Mathematical Monthly* 80 (1973): 233–269. Excellent survey of the work leading up to the solution of Hilbert's Tenth Problem concerning solutions to Diophantine equations. Online through ND library.

15. Davis, Martin Review of Gödel's 'Discussion on Providing a Foundation for Mathematics' (1931a); 'Postscript", *Journal of Symbolic Logic* 55 (1990): 343. Online through ND library.

16. Dawson, John 'The published work of Kurt Gdel: an annotated bibliography', *Notre Dame Journal of Formal Logic* 24 (1983): 255–284. List of Gödel's publications. Online through ND library.

17. Dawson, John 'Addenda and corrigenda', *Notre Dame Journal of Formal Logic* 25(1984): 283–287. Online through ND library.

18. Dawson, John W. *Logical Dilemmas: The life and work of Kurt Gödel*, A. K. Peters, Wellesley, MA, 1997.

19. Dawson, John W. and Cheryl A. 'Future Tasks for Gödel Scholars', *Bulletin of Symbolic Logic* 11 (2005): 150–171. Online through ND library.

20. Dawson, John ‘The Reception of Gödel’s Incompleteness Theorems’, *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. II, 1984, 253–271. Discussion of the reception of Gödel’s announcement and publication of his incompleteness proofs. Argues that the reaction was complex, ranging from doubt, criticism and rival claims to priority to informed admiration and acceptance. Online through ND library.

21. Detlefsen, Michael *Hilbert’s Program: An essay on mathematical instrumentalism*, D. Reidel, Dordrecht & Boston, 1986. Book-length critical investigation of the arguments used to support claims to the effect that Gödel’s theorems refute Hilbert’s Program. The focus is particularly on arguments from G2, though arguments from G1 by Smorynski and Prawitz are briefly discussed in an appendix. The general thesis is that there are deep gaps in the anti-Hilbertian arguments Gödel’s theorems, gaps whose elimination seem to require fundamental advances in our understanding of such things as (i) the conditions necessary for the proper representation of metamathematical notions by expressions in formal languages of arithmetic, (ii) what types of formalisms can best be seen as representing complexes of rational human beliefs, and (iii) how best to think of the notion of complexity of proof (what might be called ‘discovermental complexity’) that figures centrally in Hilbert’s conception of ideal methods in mathematics.

22. Detlefsen, Michael ‘On an Alleged Refutation of Hilbert’s Program Using Gödel’s First Incompleteness Theorem’, *Journal of Philosophical Logic* 19 (1990): 343–77. Critical discussion of an argument by Craig Smorynski that G1 refutes Hilbert’s Program. Online through ND library (Periodical Archive Online option).

23. Detlefsen, Michael ‘Löb’s Theorem as a Limitation on Mechanism’, *Minds and Machines* 12 (2002): 353–381. Discussion of the bearing of G¨odel’s theorems (viz. G2) and a related theorem of Löb’s on the question of the mechanizability of human reasoning. Argues that Löb’s theorem implies a certain type of limitation on mechanizability. Online through ND library. (Errata: p. 370, line 14, ‘unjustified’ should be ‘justified’; p. 375, line -7, ‘is are’ should be ‘are’.)

24. Feferman, Solomon ‘Arithmetization of metamathematics in a general setting’, *Fundamenta Mathematicae* 49 (1960): 3–92. First full published version of Feferman’s alternative proof of G2. In this proof, Feferman formulates a set of conditions on provability (and their related consistency) formulas that are intended (i) to ensure that the formula ‘expresses’ (in the relevant sense) the associated notion of provability (resp. consistency), (ii) that guarantee a G2-type unprovability result and for which it is (iii) effectively decidable whether or not a given candidate formula satisfies them. Online through ND library.

25. Feferman, Solomon 'Gödel's life and work', CW I, 1–36. Worthwhile biographical essay on Gödel.

26. Gödel, Kurt 'On undecidable propositions of formal mathematical systems', in CW I. Note: Davis' introductory note to the reprinting of this lecture in U is useful and should be read in conjunction with it. For certain differences between the version in U and the version in CW I, see the introductory note to the latter.

27. Gomez-Torrente, Mario Review of Dawson's *Logical Dilemmas: The life and work of Kurt Gödel*, *Isis* 89 (1998): 356–357. Online through ND library.

28. Hellman, Geoffrey 'How to Gödel a Frege-Russell: Gödel's Incompleteness Theorems and Logicism', *Noûs* 15 (1981): 451–468. Argument that G2 has serious negative implications for logicist philosophies of mathematics. Online through ND library.

29. Hilbert, David 'Probleme der Grundlegung der Mathematik', *Atti del Congresso internazionale dei matematici, Bologna 3–10 settembre 1928* (Bologna, 1929). Reprinted with emendations and additions in *Mathematische Annalen* 102 (1930): 1–9. English trans. of the latter in P. Mancosu (ed.), *From Brouwer to Hilbert*, Oxford, Oxford U. Press, 1998. Referred to as '**Hilbert 1930**'.

30. Hilbert, David & Paul Bernays, *Grundlagen der Mathematik*, Erster Band. Berlin, Verlag Julius Springer, 1934.

31. Hilbert, David & Paul Bernays, *Grundlagen der Mathematik*, Zweiter Band. Berlin, Verlag Julius Springer, 1939.

32. Jech, Thomas 'On Gödel's second incompleteness theorem', *Proc. Amer. Math. Soc.* 121 (1994): 311–313. Gives model theoretic proof of a set-theoretic version of G2—a theorem which says that if a set theory is consistent, it can't prove that there is a model of set theory. The proof makes use of a construction like that in Jeroslow 1973—that is, a 'diagonal' sentence that says that there's a model of its negation. Online through ND library.

33. Jeroslow, Robert 'Consistency Statements in Formal Theories', *Fundamenta Mathematicae* 72 (1971): 17–40. Using Feferman's approach to arithmetization, the author examines, among other things, various weak theories and finds some that can prove RE consistency formulae for themselves. Online through ND library.

34. Jeroslow, Robert 'Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel's Second Incompleteness Theorem', *Journal of Symbolic Logic* 38 (1973): 359–367. Jeroslow shows how to obtain a G2-type result for consistency formulas defined in the usual way from provability formulas that do *not* satisfy the second HBL Derivability Condition. The third HBL Condition is, however, still necessary. To achieve

this result Jeroslow adopts a different type of fixed-point construction (called a *literal* fixed point) in which the undecidable diagonal formula $\mathcal{J}$ is not just provably equivalent to a certain antinomical sentence, but provably *identical* to it. Jeroslow also uses a new form of antinomical sentence, one which says not that it's unprovable, but that it's negation *is* provable. Online through ND library.

35. Kikuchi, Makoto 'A note on Boolos' proof of the incompleteness theorem', *Math. Logic Quarterly* 40 (1994): 528–532.

36. Kikuchi, Makoto & K. Tanaka 'On formalization of model-theoretic proofs of Gödel's theorems', *Notre Dame Journal of Formal Logic* 35 (1994): 403–412. Formalization of a proof of G2 and Boolos' proof of G1 in a fragment of WKL_0 that is Π^0_2-conservative over PRA. Online through ND library.

37. Kikuchi, Makoto 'Kolmogorov complexity and the second incompleteness theorem', *Arch. Math. Logic* 36 (1997): 437–443. Extends Boolos' idea of deriving Gödel's theorem from Berry's paradox. Proof that Kolmogorov complexity can be used not only to formulate and prove a version of the first incompleteness theorem (as both Kolmogorov and Chaitin showed), but also a version of the second incompleteness theorem. The proof is not formalizable in PRA, though it is formalizable in WKL_0. Online through ND library.

38. Kotlarski, Henryk 'The incompleteness theorems after 70 years'. Abstract of Kotlarski's talk at the 1997 Tarski centenary conference. Just what the title suggests. Online at: Gödel's Theorems after 70 Years.

39. Kotlarski, Henryk 'On the Incompleteness Theorems', *The Journal of Symbolic Logic* 59 (1994): 1414–1419. Model-theoretic proofs of the incompleteness theorems. Online through ND library.

40. Kotlarski, Henryk 'New proofs of old results', *Mathematical Logic Quarterly* 44 (1998): 474–480.

41. Löb, Martin 'Solution of a problem of Leon Henkin', *Journal of Symbolic Logic* 20 (1955): 115–118. Answers a question of Henkin's asking for a characterization of those sentences H of an arithmetical theory T that are provably equivalent to sentences asserting their provability. Löb proves that $\vdash_T Prov_T(\ulcorner A \urcorner) \leftrightarrow A$ for all and only those A such that $\vdash_T A$. In the process of doing this, he offers a streamlined version of the derivability conditions used by Hilbert and Bernays in their 1939 *Grundlagen der Mathematik* II to give a generalized proof of G2. Online through ND library.

42. Mancosu, Paolo, 'Between Vienna and Berlin: The Immediate Reception of Gödel's Incompleteness Theorems', *History and Philosophy of Logic* 20 (1999): 33–45. This is a more extensive and detailed discussion of the reaction to Gödel's theorems. It

elaborates at greater length on the reactions of mathematicians and philosophers in Berlin and Vienna and supplies a few details missing from Dawson's discussion. Online through ND library.

43. Pour-El, Marian B. 'Effectively extensible theories', *Journal of Symbolic Logic* 33 (1968): 56–68. Useful discussion of how to make the notion of *theory extension* precise for purposes of proving a strengthened generalized form of Gödel's theorems. Introduces the basic notion of a *presentation of* a theory for these purposes (cf. p. 57 and Def. 5). Online through ND library.

44. Pudlak, Pavel 'A note on applicability of the incompleteness theorem to human mind', *Annals of Pure and Applied Logic* 96 (1999): 335–342. Argues that a proper understanding of the relations between consistency and reflection principles explains why it's wrong to use Gödel's incompleteness theorem (he seems to mean the 2nd incompleteness theorem) to argue that human thinking is essentially different from what can be simulated by a machine. Online through ND library.

45. Raatikainen, Panu 'On interpreting Chaitin's incompleteness theorem', *Journal of Philosophical Logic* 27 (1998): 569–86. Critical discussion of various interpretations of Chaitin's theorem concerning the limits on the complexity of formal proofs. Of particular concern is the interpretation of Chaitin's result as implying that for any formal theory of arithmetic T there is a finite constant c such that for no number n can T prove n to have (Kolmogorov) complexity greater than c. Online through ND library.

46. Raatikainen, Panu 'On the philosophical relevance of Gödel's incompleteness theorems', *Revue internationale de philosophie* 59 (2005): 513–534. Electronic version available on the author's website. Useful critical survey of attempts to bring Gödel's theorems to bear on a variety of philosophical issues.

47. Smorynski, Craig 'The development of self-reference: Löb's Theorem', in T. Drucker (ed.), *Perspectives on the history of mathematical logic*, Boston, Birkhäuser, 1991.

48. Smorynski, Craig 'Fifty years of self-reference in arithmetic', *Notre Dame Journal of Formal Logic* 22 (1981): 357–374. Online through ND library.

49. Tieszen, Richard 'Gödel's path from the incompleteness theorems (1931) to phenomenology (1961), *Bulletin of Symbolic Logic* 4 (1998): 181–203. Account of a possible line of thought leading Gödel from his incompleteness theorems to his interest in Husserl's phenomenology. Online through ND library.

50. van Plato, Jan Review: *Kurt Gödel: Collected Works, IV-V: Correspondence*, *Bulletin of Symbolic Logic* 10 (2004): 558–563. Online through ND library.

51. Vopenka, Petr ‘A new proof of Gödel’s result on non-provability of consistency’, *Bulletin de l’Académie polonaise des sciences. Série des sciences mathématiques, astronomiques, et physiques* 14 (1966): 111–115. First semantical proof of G2. ND library.