Kleene's Second Recursion Theorem and Self-Knowing Programs

Last Updated April 6th, 2023

Kleene's Second Recursion Theorem

"Know Thyself"

Socrates

Consider a computable function $f(x, y)$, where x is viewed as a Gödel number of some program and y is some other input. The following are some statements that could be made in an informal program that computes f.

- Print the instructions of P_x.
- Simulate the computation of P_x on input y.
- Count the number of Jump instructions that are executed in the computation of P_x on input y.
- Send program x and y to a server on the internet.
- Return the tuple of configurations that constitutes the computation of P_x on input y.

Now suppose we take f's program statements and re-write them in a self-referencing way, to where we get statements like the following ones.

- Print *my* instructions.
- Simulate *myself* on input y.
- Count the number of Jump instructions that I execute when I'm computing input y.
- Send *myself* and y to a server on the internet.
- Return the tuple of configurations that constitutes *my* computation on input y.

A program that makes one or more references to its own Gödel number is said to be **self-knowing** or **self-referencing**.

Catch-22 for a Program P Attempting to Know Itself

1. For P to know its Gödel number, it must know each of its instructions.
2. Some instructions, such as "print myself", requires P to know its Gödel number.

Proposed Solution to Catch-22

1. Assume for the sake of argument that, after replacing statements about x with statements about itself, that there does in fact exist a program P_e with Gödel number e that computes the resulting function.
2. Then P_e is a function of the single variable y (since variable x has been assigned constant e).
3. Therefore, we have, for all y,
$$\phi_e(y) = f(e, y).$$
In other words, there is a program P_e that, on input y computes $f(e, y)$, and thus makes references (to e that has been substituted for x) to its own Gödel number.
4. Thus, we have reduced the problem to that of finding a Gödel number e that satisfies the above equation.
5. Stephen Kleene's second recursion theorem states that such an e does exist!

Kleene's Second Recursion Theorem. Let $f(x, y)$ be a computable function that takes as input a Gödel number x, and some additional input y. Then there is a Gödel number e for which $\phi_e(y) = f(e, y)$.

Example 1. Consider the URM computable function $f(x, y)$ which, on inputs x and y, simulates the computation $P_x(y)$, and returns the number of times that a jump instruction is executed during the computation $P_x(y)$. Then by the 2nd recursion theorem, there is a program P_e for which $P_e(y) = f(e, y)$, and so, for input y, P_e computes the number of times that its own self executes a jump instruction during its computation with input y. □

Proof of Kleene's Second Recursion Theorem. The idea behind the proof is to divide the construction of the desired program $P = ABC$ into three parts: A, B, and C which we now describe. Assume that y is the input to P.

Part A.
- Move y to register R_2.
- Place B's Gödel number b in R_1.

Part B.
- Use b in R_1 to compute A's Gödel number a.
- Compute C's Gödel number c.
- Compute

$$e = \gamma(\gamma^{-1}(a), \gamma^{-1}(b)\gamma^{-1}(c)) = \gamma(ABC) = \gamma(P),$$

  the Gödel number of the concatenation of A's, B's, and C's instructions.
- Place e in R_1, with y remaining in R_2.

Part C. Compute $f(e, y)$.

"Gödel Escher Bach: The Eternal Golden Braid"

Notes.

1. The most straightforward of the three is part C, since its sole purpose is to compute function f which is assumed URM computable, and so C's instructions consist of the instructions of the URM program used to compute f.

2. The clever part of the above program is understanding how A is able to compute B's Gödel number and vice versa. This is actually made possible by an elementary use of the s-m-n theorem.

3. Consider the function $g(x, y) = x$. By the s-m-n theorem, there is a total computable function $k(x)$ for which

$$\phi_{k(x)}(y) = g(x, y) = x.$$

4. Then define A's Gödel number to be equal to $k(b)$. This works because, on input y, program A outputs

$$\phi_{k(b)}(y) = b$$

   in register R_1, and has the side effect of placing y in R_2 via an initial $T(1, 2)$ statement. Therefore A works in exactly the way it was described above.

Given that $a = k(b)$ we may now describe B's program as follows.

Program B

Input Gödel number z.

Compute Gödel number $k(z)$.

Compute $c = \gamma(C)$.

Return

$$\gamma(\gamma^{-1}(k(z)), \gamma^{-1}(z), c).$$

Important: notice that B's program does *not* depend on knowing A's Gödel number a. If it did, then it would create a circularity error, since $a = k(b)$ already depends on B's Gödel number. However, B is able to compute a once it has its own Gödel number $z = b$ since step 2 of its algorithm yields $a = k(b)$.

Thus, we see that, after the execution of A on input y, B receives input $z = b$ which gives $a = k(z) = k(b)$, and so B outputs into R_1 the value

$$e = \gamma(\gamma^{-1}(a), \gamma^{-1}(b), \gamma^{-1}(c)) = \gamma(ABC) = \gamma(P).$$

The following diagram shows the results of all three programs combined in sequence, where $v \xrightarrow{X} w$ means that program X inputs v and outputs w. Then we have

$$y \xrightarrow{A} (b, y) \xrightarrow{B} (e = \gamma(ABC), y) \xrightarrow{C} f(e = \gamma(ABC), y).$$

Therefore, $P = ABC = P_e$ computes

$$\phi_e(y) = f(e, y),$$

and the proof is complete. $\square$

Example 2. Program P is called **totally introspective** iff, on input y, P returns a number that encodes every configuration of the computation of P on input y. Letting $\sigma(x, y, i)$ denote the encoding of the i th configuration of the computation $P_x(y)$, then we define the computable function

$$f(x, y) = \begin{cases} \prod_{i \leq t} p_i^{\sigma(x,y,i)} & \text{if } P_x(y) \downarrow \text{ in } t \text{ steps} \\ \text{undefined} & \text{otherwise} \end{cases}$$

Now, by the 2nd recursion theorem, there exists a Gödel number e for which $\phi_e(y) = f(e, y)$, meaning that P_e is totally introspective, since, on input y, it outputs an encoding of all the configurations used in the computation $P_e(y)$. $\square$

The `self` Programming Statement

The Recursion theorem gives rise to a tool that may be used when writing a program P. Namely, we may make reference to P's Gödel number, which is represented with the keyword `self`. This keyword is similar to the `this` keyword of Java, which refers to the object that a Java method is acting on. For example, the following are valid programming statements for program P:

```
int f(int y)
{
    int length = self.instructions.length;

    print("Hi! I have Godel number equal to ");
    print(self);
    print(".\nI have ");
    print(length);
    print(" instructions.\nI will return the square of the output that results ");
    print("from simulating myself on input y.\n");
    print("What are the possibilities for the output value?\n");
    length = simulate(self,y);
    return length*length;
}
```

To justify such a program, suppose $y \in \mathcal{N}$ is the input to P, and the purpose of P is to implement the unary computable function $f(y)$. Then we may do the following.

1. Transform P by adding another input x, so that we are now implementing function $f(x, y)$.
2. Replace each occurrence of `self` with x.
3. Use the method described in the proof of Kleene's 2nd Recursion Theorem to compute an e for which P_e computes
$$\phi_e(y) = f(e, y).$$
4. Therefore, P_e computes $f(y)$, with e substituted for x.
5. Therefore, P_e's references to `self` are justified, since `self` $= e$, the Gödel number of the program that computes $f(y)$.

Kleene's 2nd Recursion Theorem and Undecidability

Recall that a predicate function p is **decidable** iff there is some URM program that is capable of computing p. Otherwise, we say that p is **undecidable**.

Example 3. Let x and y be the encodings of two DFA's M_x and M_y and let predicate $\text{EQ\_DFA}(x, y) = 1$ iff $L(M_x) = L(M_y)$, i.e. M_x and M_y accept the same set of words. Then $\text{EQ\_DFA}(x, y)$ is decidable via the following algorithm that is outlined below.

I. Compute the DFA M that accepts the language

$$L(M_x) \oplus L(M_y) = (L(M_x) \cap \overline{L(M_y)}) \cup (L(M_y) \cap \overline{L(M_x)}).$$

This can be done using both the Intersection Algorithm and the algorithm for computing an NFA that accepts the union of two languages (and then converting the NFA to a DFA).

II. We have $L(M_x) = L(M_y)$ iff $L(M_x) \oplus L(M_y) = \emptyset$ (why?).

III. Thus, we need only check if the initial state of M can reach any of its accepting states along some path in the state diagram. Then $L(M_x) = L(M_y)$ iff there is no path from the initial state to an accepting state.

We state the following Theorem without proof (See Chapter 4 of Sipser).

Theorem 2. Let x and y be the encodings of two CFG's G_x and G_y and let predicate EQ_CFG$(x, y) = 1$ iff $L(G_x) = L(G_y)$, i.e. G_x and G_y derive the same set of words. Then EQ_CFG(x, y) is *undecidable.*

Now consider a predicate function $p(x)$ that outputs a 1 iff the URM program having Gödel number x has some property. Then Kleene's 2nd Recursion theorem can be used to show that p is undecidable. To accomplish this we use the following strategy.

1. Assume p decidable.
2. Define a program P which, on input y, first computes $p(\text{self})$.
3. If $p(\text{self}) = 1$, then P proceeds to act in a contradicting way so that it is necessary for $p(\gamma(P)) = 0$.
4. If $p(\text{self}) = 0$, then P proceeds to act in a contradicting way so that it is necessary for $p(\gamma(P)) = 1$.
5. From 3 and 4, we must conclude that p is undecidable.

Theorem 3. The Halting Problem is the problem of deciding whether or not the program whose Gödel number is x halts on input y. Let $H(x, y)$ be the predicate function for which $H(x, y) = 1$ iff $P_x(y) \downarrow$. Then H is undecidable. Equivalently, $H(x, y) = 1$ iff $y \in W_x$. is undecidable.

Example 4. Provide a program P_1 for which $H(\gamma(P_1), 1) = 1$, and a P_2 for which $H(\gamma(P_2), 1) = 0$.

Proof of Theorem 3. We assume the Halting Problem is decidable, i.e.

$$H(x,y) = \begin{cases} 1 & \text{if } y \in W_x \\ 0 & \text{otherwise} \end{cases}$$

is total computable. Now consider the following program P.

Input $y \in \mathcal{N}$.

If $H(\texttt{self}, y)$, loop forever.

Return 1.

Let $e = \texttt{self}$ denote the Gödel number for P. Then $P_e(e) = 1$ provided $H(e,e) = 0$ iff $P_e(e)$ does not halt, a contradiction. Similarly, $P_e(e)$ does not halt provided $H(e,e) = 1$ iff $P_e(e)$ does halt, another contradiction. Therefore, the assumption that H is decidable must be false.

Example 5. Prove that the function

$$g(x) = \begin{cases} 1 & \text{if } \phi_x \text{ is total} \\ 0 & \text{otherwise} \end{cases}$$

is undecidable. First give examples of programs P_1 and P_2 for which $g(\gamma(P_1)) = 1$ and $g(\gamma(P_2)) = 0$.

Other Applications of Kleene's 2nd Recursion Theorem

Theorem 4. Consider the set M, where $x \in M$ iff there is no $y < x$ for which $\phi_y = \phi_x$. In other words, P_x is a minimal program for function ϕ_x. Then M is not recursively enumerable.

Proof of Theorem 3. Suppose M is recursively enumerable. Then it is an exercise to show that there is a total computable unary function f whose range is equal to M. In other words $M = \{f(i) | i \in \mathcal{N}\}$. Consider the following program P.

> Input $x \in \mathcal{N}$.
>
> For each $i = 0, 1, \ldots$
>
> > If $f(i) >$ `self`, then `break`.
>
> Simulate program $P_{f(i)}$ on input x, and return y in case $P_{f(i)}(x) \downarrow y$.

Let e be the Gödel number of P. Then it follows that $\phi_e = \phi_{f(i)}$. But $f(i) > e$ which contradicts the fact that $f(i) \in M$. Therefore, the assumption that M is r.e. must be false.

Theorem 5. Let f be a total computable unary function. Then there is a number $n \in \mathcal{N}$ for which $\phi_n = \phi_{f(n)}$. We refer to n as a **fixed point** for f.

Proof of Theorem 4. Consider the following program P.

> Input $x \in \mathcal{N}$.
>
> Compute $y = f(\texttt{self})$.
>
> Simulate program P_y on input x, and return z in case $P_y(x) \downarrow z$.

Then

$$\phi_y = \phi_{f(\texttt{self})} = \phi_{\texttt{self}},$$

and so $n = \texttt{self}$ is a fixed point for f.

Exercises

1. With respect to Kleene's 2nd Recursion Theorem, prove that there are infinitely many values e for which $\phi_e(y) = f(e, y)$. Hint: consider program B in the proof of the theorem.

2. Recall that a function $f : \mathcal{N} \to \mathcal{N}$ is **onto** provided for every $y \in \mathcal{N}$ there is an $x \in \mathcal{N}$ for which $f(x) = y$. Consider the function

$$g(x) = \begin{cases} 1 & \text{if } \phi_x \text{ is onto} \\ 0 & \text{otherwise} \end{cases}$$

Evaluate $g(a)$, $g(b)$, and $g(c)$, where

(a) $\phi_a(y) = y^2$

(b) $\phi_b(y) = 1$

(c) $\phi_c(y) = y$.

3. Prove that the function

$$g(x) = \begin{cases} 1 & \text{if } \phi_x \text{ is onto} \\ 0 & \text{otherwise} \end{cases}$$

is not URM computable. In other words, there is no URM program that, on input x, always halts and either outputs 1 or 0 as output, depending on whether or not ϕ_x is onto. Do this by writing a program P that uses g and makes use of the `self` programming concept.

4. Recall that W_x denotes the domain of the function $\phi_x(y)$, i.e. the natural number inputs y to ϕ_x for which $\phi_x(y)$ is defined. Consider the function

$$g(x) = \begin{cases} 1 & \text{if } W_x = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Evaluate $g(a)$, $g(b)$, and $g(c)$, where

(a) $P_a = S(2), S(2), S(1), J(1, 2, 6), J(1, 1, 3)$

(b) $P_b = S(2), J(2, 3, 3), J(1, 1, 1)$

(c) $P_c = S(1), S(1), S(2), J(1, 2, 6), J(1, 1, 1)$

5. Prove that the function

$$g(x) = \begin{cases} 1 & \text{if } W_x = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

is not URM computable. In other words, there is no URM program that, on input x, always halts and either outputs 1 or 0 as output, depending on whether or not ϕ_x has an empty domain. Do this by writing a program P that uses g and makes use of the `self` programming concept. Then show how P creates a contradiction.

6. Consider the function

$$g(x) = \begin{cases} 1 & \text{if } |E_x| = \infty \\ 0 & \text{otherwise} \end{cases}$$

In other words $g(x) = 1$ iff function $\phi_x(y)$ has an infinite range, meaning that it outputs an infinite number of different values. Evaluate $g(a)$, $g(b)$, and $g(c)$, where

(a) $\phi_a(y) = y^2$

(b) $\phi_b(y) = y$

(c) $\phi_c(y) = \text{sgn}(y)$.

7. Prove that the function

$$g(x) = \begin{cases} 1 & \text{if } |E_x| = \infty \\ 0 & \text{otherwise} \end{cases}$$

is not URM computable. In other words, there is no URM program that, on input x, always halts and either outputs 1 or 0 as output, depending on whether or not ϕ_x has an infinite range. Do this by writing a program P that uses g and makes use of the `self` programming concept. Then show how P creates a contradiction.

8. Rice's theorem states that if $\mathcal{C}_1$ denotes the set of unary computable functions, and $\mathcal{B}$ is a nonempty proper subset of $\mathcal{C}_1$, then the predicate function

$$B(x) = \begin{cases} 1 & \text{if } \phi_x \in \mathcal{B} \\ 0 & \text{otherwise} \end{cases}$$

is undecidable. Prove Rice's theorem by writing an informal program P that uses $B(x)$ and makes use of the `self` programming concept. Then show how P creates a contradiction. Hint: assume $B(x)$ is decidable, and take advantage of the fact that the set of functions $\mathcal{B}$ is both nonempty and not all of $\mathcal{C}_1$.

9. For each constant $n \geq 1$, show that $\lfloor x^{1/n} \rfloor$ is a primitive-recursive function of x.

10. Prove that there exists an n for which $\phi_n(x) = \lfloor x^{1/n} \rfloor$. Hint: use the s-m-n theorem and Theorem 4.

11. Recall that program P_x has the self-output property iff $x \in E_x$. By writing an informal program that makes use of the programming construct `self`, prove that the self-output property is undecidable.

12. Show that there is a number e for which $\phi_e(x) = e^{10}$, for all $x \in \mathcal{N}$.

13. Consider the following description of a function $f(n)$. On input n, return the Gödel number of the program P' that is the result of appending program P_n with a minimum number of successor instructions $S(1), \ldots, S(1)$ so that it is always guaranteed that, should P_n halt on an input, then the final instruction of P' will be one of these successor instructions. Then by the Church-Turing thesis, f is total computable. Moreover, prove that, if n is a fixed point for $f(n)$, i.e. $\phi_n = \phi_{f(n)}$, then necessarily $\phi_n(x)$ is undefined for all x.

Exercise Solutions

1. Since the proof of Kleene's 2nd Recursion Theorem constructs e as $e = \gamma(ABC)$, by changing the instructions of B, we get a new value for e, since B has changed. We only have to make sure that B's instructions are changed in a trivial way that does not affect its functionality as described in the proof.

2. A function $\phi_x(y)$ is onto iff $E_x = \mathcal{N}$, where E_x denotes the range of ϕ_x. Thus,

   (a) $g(a) = 0$ since $\phi_a(y) = y^2$ is not onto since $E_a = \{1, 4, 9, 25, \ldots\} \neq \mathcal{N}$,

   (b) $g(b) = 0$ since $\phi_b(y) = 1$ is not onto since $E_b = \{1\} \neq \mathcal{N}$, and

   (c) $g(c) = 1$ since $\phi_c(y) = y$ is onto since $E_c = \mathcal{N}$.

3. We have the following program P.

   Input $y \in \mathcal{N}$.

   If $g(\texttt{self}) = 1$, loop forever.

   Return y;

   If $g(\texttt{self}) = 1$, then P has a range equal to $\mathcal{N}$ which is impossible since it does not terminate on any input (loops forever). If $g(\texttt{self}) = 0$, then P does not have a range equal to $\mathcal{N}$, which is contradicted by the fact that P returns y on input y, and so has the set of return values $\{0, 1, \ldots\} = \mathcal{N}$.

4. We have the following answers.

   (a) $g(a) = 0$ since P_a terminates on input 1 (verify!) and thus $W_a = \{1\} \neq \emptyset$.

   (b) $g(b) = 1$ since P_b does not terminate on any input (why?) and thus $W_b = \emptyset$.

   (c) $g(c) = 1$ since P_c does not terminate on any input (why?) and thus $W_c = \emptyset$.

5. We have the following program P.

   Input $y \in \mathcal{N}$.

   If $g(\texttt{self}) = 1$, Return 0.

   Loop Forever.

   If $g(\texttt{self}) = 1$, then it means $W_{\text{self}} = \emptyset$, but P returns 0 for each input y, which implies $W_{\text{self}} = \mathcal{N}$, a contradiction.

   If $g(\texttt{self}) = 0$, then it means $W_{\text{self}} \neq \emptyset$, but P loops forever on each input y, which implies $W_{\text{self}} = \emptyset$, a contradiction.

6. We have the following answers.

   (a) $g(a) = 1$ since $\phi_a(y) = y^2$ has an infinite range: $E_a = \{1, 4, 9, 25, \ldots\}$,

   (b) $g(b) = 1$ since $\phi_b(y) = y$ has an infinite range $E_b = \mathcal{N}$, and

   (c) $g(c) = 0$ since $\phi_c(y) = \text{sgn}(y)$ has finite range equal to $\{0, 1\}$.

7. Consider the following program P.

   Input $y \in \mathcal{N}$.

   If $g(\texttt{self}) = 1$, Return 0.

   Return y.

If $g(\texttt{self}) = 1$, then it means $|E_{\text{self}}| = \infty$, but the program returns 0 for each input y, which implies $E_{\text{self}} = \{0\}$ which is finite, a contradiction.

If $g(\texttt{self}) = 0$, then it means $|E_{\text{self}}|$ is finite, but the program returns y on each input y, which implies $E_{\text{self}} = \mathcal{N}$, a contradiction.

8. Assume $B(x)$ is decidable. Since $\mathcal{B}$ is nonempty there exists a unary computable function $f \in \mathcal{B}$. Similarly, since $\mathcal{B}$ is not all of $\mathcal{C}_1$, there is a unary computable function $g \notin \mathcal{B}$. Now consider the following program P.

   Input $x \in \mathcal{N}$.

   If $B(\texttt{self}) = 1$,

   Simulate g on input x.

   Return $g(x)$ if it is defined.

   Simulate f on input x.

   Return $f(x)$ if it is defined.

   Since f and g are computable, so is P. Let e denote the Gödel number of P. Assume $B(e) = 1$. By definition, this means that $\phi_e \in \mathcal{B}$. But in examining P we see that P simulates g so that $\phi_e = g \notin \mathcal{B}$, a contradiction. Similarly, if $B(e) = 0$, then $\phi_e \notin \mathcal{B}$. But in this case P simulates f so that $\phi_e = f \in \mathcal{B}$, a contradiction. Therefore, B cannot be decidable.

9. The function $\lfloor x^{1/n} \rfloor$ may be computed as

$$\mu(z \leq x)(z^n > x) - 1.$$

10. Function $f(n, x) = \lfloor x^{1/n} \rfloor$ is computable by the previous exercise. Therefore, by the s-m-n theorem, there exists a total computable function $k(n)$ for which $\phi_{k(n)}(x) = \lfloor x^{1/n} \rfloor$. Finally, by Theorem 4, there is an integer n for which

$$\phi_n(x) = \phi_{k(n)}(x) = \lfloor x^{1/n} \rfloor.$$

11. Assume $E(x)$ is decidable, where $E(x) = 1$ iff $x \in E_x$. Now consider the following program P.

   Input $x \in \mathcal{N}$.

   If $E(\texttt{self}) = 1$,

   Loop forever.

   Return `self`.

   Since $E(x)$ is decidable, P is computable. Let e denote the Gödel number of P. Assume $E(e) = 1$. By definition, this means that $e \in E_e$, meaning that P returns e on some input x. However, since $E(e) = 1$, P does not terminate on any input, meaning that $E_e = \emptyset$, a contradiction.

   Similarly, if $E(e) = 0$, then $e \notin E_e$. But in this case P returns e, meaning that $e \in E_e$, a contradiction. Therefore, $E(x)$, i.e. the Self-Output property, is not decidable.

12. Function $f(y, x) = y^{10}$ is primitive recursive, and hence computable. Therefore, by the s-m-n theorem, there exists a total computable function $k(y)$ for which $\phi_{k(y)}(x) = y^{10}$. Finally, by Theorem 4, there is an integer e for which

$$\phi_e(x) = \phi_{k(e)}(x) = e^{10}$$

for all $x \in \mathcal{N}$.

13. Since $f(n)$ is total computable, by Theorem 4 there is an integer n for which $\phi_n(x) = \phi_{f(n)}(x)$ for all $x \in \mathcal{N}$. But the way in which Gödel number $f(n)$ is constructed is such that, whenever $\phi_n(x) = y$ is true, then P_n halts, which in turn implies that $P_{f(n)}$ halts with $\phi_{f(n)}(x) = y + 1$, since $P_{f(n)}$ is the same as P_n, except that in its final instruction it adds 1 to register R_1. Thus, if $\phi_n(x)$ is defined, then we have $\phi_n(x) = y \neq \phi_{f(n)}(x) = y + 1$. Therefore, we must conclude that $\phi_n(x)$ must always be undefined, meaning that $W_n = \emptyset$.