A Solution to the Surprise Exam Paradox in Constructive Mathematics

Mohammad Ardeshir and Rasoul Ramezanian
Department of Mathematical Sciences,
Sharif University of Technology,
P. O. Box 11365-9415, Tehran, Iran
mardeshir@sharif.edu, ramezanian@sharif.edu

Abstract

We represent the well-known surprise exam paradox in constructive and computable mathematics and offer solutions. One solution is based on Brouwer's continuity principle in constructive mathematics, and the other involves type 2 Turing computability in classical mathematics. We also discuss the backward induction paradox for extensive form games in constructive logic.

1 Introduction

The teacher announces to the students:

> *you will have one and only one exam at 10am on one day next week (Monday–Friday), but you will not know in advance the day of exam.*

The students, using a *backward argument*, reason that *there can be no exam* indeed:

- Friday is not the day of the exam. Since if it were, then we will not have received the exam through Thursday, and as there is an exam during the week, then on Thursday night we will be able to know in advance the day of exam,
- Thursday is not the day of the exam. Since if it were, then we will not have received the exam through Wednesday, and as there is an exam during the week, and it is not on Friday, then on Wednesday night we will know in advance the day of exam is on on Thursday,...
- and so on,
- then none of the days of the next week is the day of exam.

The above statements constitute "The Surprise Exam Paradox" (**SEP**). There are different formulations of the paradox in logico-philosophical literature, e.g., [4, 20]. **SEP** was brought to public attention about six decades ago [11, 13, 14]. Since that time, more than 150 papers have been written on the subject to explain and (or) solve the paradox. The solutions proposed are based on different approaches, e.g., [14, 7, 3, 9]. Some authors try to solve the paradox in terms of epistemic notions [2, 13, 16, 20] and others in the context of logical notions [19], in particular, in dynamic epistemic logic [6, 10, 4]. For example, Gerbrandy sees the puzzle in the assumption that announcements are in general successful [6]. Baltag allows the students to revise their trust in the teacher once they reach the paradox in order to solve the paradox [10]. Some works relate the paradox to Gödel's self-referential statements [14, 8, 3], and claim that the teacher's announcement is self-referring in nature, and others relate the paradox to backward induction in finite dynamic game theory [12, 15].

In this paper, we reformulate the surprise exam paradox in the context of *mathematical constructivism*, and propose a solution for it. In teacher's announcement, we simply replace the word "*know*" with the word "*construct/compute*".

The teacher announces to the students:

> *you will have one and only one exam at 10am on one day next week, but you will not be able to* `construct/compute` in advance *the day of exam.*

The core of our proposed solution is the *free will* of the teacher to choose a day for the exam in *future*. In constructive mathematics, Brouwer's *continuity principle* and the notion of *choice sequences* formally declare the notion of free will of the subject [1]. In section 2, we reformulate **SEP** in Brouwerian mathematical constructivism [17] and offer a solution for it based on Brouwer's continuity principle and constructive logic [1, 5, 17]. In section 3, we discuss the paradox in computable analysis, and put forward a solution considering the classical logic. Finally, we discuss the backward induction in constructive logic in the last section of this paper.

2 A Mathematical Constructivist Solution

As is well-known, the slogan of mathematical constructivism is that a mathematical object is admitted in our domain of discourse if it can be *constructed*. Commitment to this slogan, imposes its own epistemology and logic. In epistemology, this implies that we *know* a proposition to be true if we have a *proof* for it, and a proof itself is a *construction* or *computation*. This epistemology, in turn, imposes its own logic, *intuitionistic logic*. Intuitionistic reading of the logical connectives (the so called **BHK** interpretation) results in denying some tautologies of classical logic such as $\varphi \vee \neg\varphi$ and $\neg\neg\varphi \to \varphi$. For more details, see, e.g., [1, 5, 17].

In **SEP**, we expect the day of the exam to depend on the free will of the teacher. Brouwer's continuity principle states the role of free will in constructive mathematics. Imagine you have a collection of objects at your disposal, let's say the natural numbers. Pick out one of them, and record the result. Put it back into the collection, and choose again. Since you have the *ability to choose freely*, you may choose a different one, or the same again. Record the result, and put it back. You may make keep on making further choices. In this way, you are developing a *choice sequence* if you think of the sequence you are making as potentially infinite [1]. Initial segments are always finite. We cannot make an actually infinite number of choices, but we can always extend an initial segment by making a further choice. If we want to apply a function to a choice sequence, a sequence will have to act as an argument, to which then a method is applied to calculate the function's output. But we cannot construct the argument in its entirety, as a choice sequence is always an unfinished object, and therefore, one always has access only to a finite segment of the sequence. The Brouwer weak continuity principle for numbers, **WCN** says that a total function from choice sequences to natural numbers never needs more input than a *finite initial segment* to produce its output; hence all choice sequences sharing this segment will yield the same value [1, 5, 17].

$$\textbf{WCN}\text{: Assume } \Phi \in \mathbb{N}^{\mathbb{N}} \to \mathbb{N} \text{ is total, then } \forall\alpha\exists x\forall\beta(\overline{\beta}x = \overline{\alpha}x \to \Phi\beta = \Phi\alpha),$$

where $\overline{\alpha}x = \langle\alpha(0), \alpha(1), ..., \alpha(x-1)\rangle$. To state the principle more formally, we need some definitions.

Definition 2.1 *1. A* tree $\mathbf{T}$ *is an inhabited decidable set of finite sequences of natural numbers closed under predecessor, i.e.,*

- $\langle\rangle \in \mathbf{T}$,
- $\forall n(n \in \mathbf{T} \vee \neg n \in \mathbf{T})$,
- $\forall nm(n \in \mathbf{T} \wedge m \prec n \to m \in \mathbf{T})$. *($m \prec n$ means m is a prefix of n)*

2. A spread *is a tree in which each node has at least one successor,*

$$\forall n \in \mathbf{T}\exists x(n * \langle x\rangle \in \mathbf{T})$$

3. A sequence α is a branch *of the tree iff all initial segments of α belong to $\mathbf{T}$. We denote this by $\alpha \in \mathbf{T}$.*

The weak continuity principle for natural numbers can be stated as follows. Let $\mathbf{T}$ be an arbitrary spread, then

$$\textbf{WCN}: \quad \forall\alpha \in \mathbf{T}\exists y(\Phi\alpha = y) \to \forall\alpha \in \mathbf{T}\exists x\forall\beta \in \mathbf{T}(\overline{\beta}x = \overline{\alpha}x \to \Phi\beta = \Phi\alpha).$$

2.1 Reformulation of the paradox

We reformulate the surprise exam paradox in constructive analysis as follows.
The mathematician announces to the students:

> *I develop a sequence α of natural numbers such that for all n, $(1 \leq \alpha(n) \leq 5)$ and $(\alpha(n+1) = \alpha(n) \vee \alpha(n+1) = \alpha(n)+1)$. But you will not be able to construct/compute the maximum value of the sequence.*

Let T be the set of all sequences α satisfying the following conditions:

1. $\forall n(1 \leq \alpha(n) \leq 5)$,
2. $\forall n(\alpha(n+1) = \alpha(n) \vee \alpha(n+1) = \alpha(n)+1)$.

It is easy to verify that T is a spread. For each $\alpha \in$ T and $k \in \mathbb{N}$, we define

$$\max(\alpha, k) :\equiv \forall n(\alpha(n) \leq k) \wedge \exists n(\alpha(n) = k).$$

The formula $\max(\alpha, k)$ states that k is the maximum value of the sequence α.
By *Increasing Bounded Sequences* statement, **IBS** , we mean the conjunction of the following two formulas:

1. $(\forall\alpha \in \mathrm{T})\neg\neg\exists m \forall n \geq m(\alpha(n) = \alpha(m))$,
2. $\neg\exists f : \mathrm{T} \to \mathbb{N}\ \forall\alpha \in \mathrm{T}[\max(\alpha, f(\alpha))]$.

We justify that **IBS** is a reformulation of **SEP** as follows.

The teacher announces to the students:

> *you will have one and only one exam at 10am on one day next week (Monday–Friday), but you will not know in advance the day of exam.*

The mathematician also announces to the same students:

> *I develop a sequence α of natural numbers such that for all n, $(1 \leq \alpha(n) \leq 5)$ and $(\alpha(n+1) = \alpha(n) \vee \alpha(n+1) = \alpha(n)+1)$. But you will not be able to construct/compute the maximum value of the sequence.*

We claim that if the students cannot compute the maximum value of the sequence developed by the mathematician, then they would not know in advance the day of the exam taken by the teacher.

The mathematician develops a sequence α based on the *behavior* of the teacher as follows.

He first chooses an arbitrary natural number $n_1 \geq 1$ and lets $\alpha(1) = \alpha(2) = \cdots = \alpha(n_1) = 1$. Then, he observes teacher's behavior on Monday, 10am. If the exam is given, then the mathematician defines $\alpha(n) = 1$, for all $n > n_1$. Otherwise, he chooses another arbitrary natural number $n_2 \geq 1$, and lets $\alpha(n_1+1) = \alpha(n_1+2) = \cdots = \alpha(n_1+n_2) = 2$.
Again, the mathematician observes teacher's behavior on Tuesday, 10am. If the exam is given, he defines $\alpha(n) = 2$, for all $n > n_1 + n_2$. Otherwise, he chooses an arbitrary natural number $n_3 \geq 1$, and lets $\alpha(n_1+n_2+1) = \alpha(n_1+n_2+2) = \cdots = \alpha(n_1+n_2+n_3) = 3$.
In a similar way, the mathematician observes the teacher's behavior on Wednesday, 10 am. If the exam is given, he defines $\alpha(n) = 3$, for all $n > n_1+n_2+n_3$. Otherwise, he chooses an arbitrary natural number $n_4 \geq 1$, and lets $\alpha(n_1+n_2+n_3+1) = \alpha(n_1+n_2+n_3+2) = \cdots = \alpha(n_1+n_2+n_3+n_4) = 4$.
As for Thursday, the mathematician observes the teacher's behavior on Thursday, 10 am. If the exam is given, he defines $\alpha(n) = 4$, for all $n > n_1+n_2+n_3+n_4$, and otherwise, he defines $\alpha(n) = 5$, for all $n > n_1+n_2+n_3+n_4$.

It is obvious that if the students have a procedure to predict the day of the exam *in advance*, then they would be able to compute the maximum value of the sequence α developed by the mathematician as well. In Theorem 2.4 and Proposition 3.4, we show that there is no constructive/computable function that computes the maximum value of sequences in the spread T.

It is interesting to note that the first step of student's argument, i.e., Friday is not the day of the exam day, is plausible. In our reformulation **IBS** , the first step of the student's argument is as follows. If $\alpha \in \mathrm{T}$ is a sequence developed by the mathematician, and in a finite initial segment of α developed until now, say $\langle\alpha(1), \alpha(2), ..., \alpha(k)\rangle$, there exists $1 \leq i \leq k$ such that $\alpha(i) = 5$, then we can compute the maximum value of the sequence α, based on this finite initial segment, and the maximum value would be five.

2.2 The Solution

The student, using a backward argument, proves the following Proposition and its Corollary 2.3.

Proposition 2.2 $(\forall \alpha \in \mathrm{T})\neg\neg\exists m \forall n \geq m(\alpha(n) = \alpha(m))$.

Proof. To show $\forall \alpha \in \mathrm{T}\neg\neg\exists m \forall n \geq m[\alpha(n) = \alpha(m)]$, it is enough to show $\neg\exists \alpha \in \mathrm{T}\neg\exists m \forall n \geq m[\alpha(n) = \alpha(m)]$, by the constructive valid statement $\neg\exists x A(x) \leftrightarrow \forall x \neg A(x)$. So assume, for some $\alpha_0 \in \mathrm{T}$, $\neg\exists m \forall n \geq m[\alpha_0(n) = \alpha_0(m)]$. Then $\forall m \neg\forall n \geq m[\alpha_0(m) = \alpha_0(n)]$. Since $\alpha_0 \in \mathrm{T}$, we have $\alpha_0(n) \leq \alpha_0(n+1) \leq 5$, for all $n \in \mathbb{N}$. We prove that for every $k \in \mathbb{N}$,

$$(\forall n(\alpha_0(n) \leq k)) \rightarrow (\forall n(\alpha_0(n) < k)) \ (1).$$

Assume there exists $t \in \mathbb{N}$ such that $\alpha_0(t) = k$. Since α_0 is non-decreasing and $\forall n(\alpha_0(n) \leq k)$, we have $\forall n > t[\alpha_0(n) = k]$. This contradicts with the assumption $\forall m \neg\forall n \geq m[\alpha_0(m) = \alpha_0(n)]$. Hence $\forall n(\alpha_0(n) < k)$.
Now let $k = 5$. By (1), we derive $\forall n(\alpha_0(n) \leq 4)$. Repeating this argument by using (1), we will have $\forall n(\alpha_0(n) = 0)$. This contradicts with $\forall m \neg\forall n \geq m[\alpha(m) = \alpha(n)]$. $\dashv$

From Proposition 2.2, we can simply derive the following Corollary.

Corollary 2.3 *For each* $\alpha \in \mathrm{T}$, $\neg\neg\exists k(\max(\alpha, k))$.

Using Brouwer's continuity principle, we show that it is not possible to construct for each $\alpha \in \mathrm{T}$, a natural number k such that $\max(\alpha, k)$ holds.

Theorem 2.4 (**WCN**). *There is no function* $f : \mathrm{T} \rightarrow \mathbb{N}$, *such that* $\forall \alpha \in \mathrm{T}[\max(\alpha, f(\alpha))]$.

Proof. Suppose there exists a function $f : \mathrm{T} \rightarrow \mathbb{N}$ such that $\forall \alpha \in \mathrm{T}[\max(\alpha, f(\alpha))]$. Let $\beta = \overline{\langle 1 \rangle}$ be the sequence of ones. Then $f(\beta) = 1$. By **WCN** , there exists $j \in \mathbb{N}$ such that for all $\alpha \in \mathrm{T}$, if $\bar{\alpha}(j) = \bar{\beta}(j)$ then $f(\alpha) = f(\beta) = 1$. Let $\alpha_0 \in \mathrm{T}$ be a sequence that its first j-th elements are one and all other elements are two, i.e.,

$$\alpha_0(n) = \begin{cases} 1 & n \leq j, \\ 2 & n > j. \end{cases}$$

For the sequence α_0, $\max(\alpha_0, 2)$ contradicts with $f(\alpha_0) = 1$. $\dashv$

Theorem 2.4 and Corollary 2.3 show that **IBS** holds, and thus there is no paradox in teacher's announcement.

3 A Computable Analysis Solution

In the previous section, we discussed the paradox in constructive logic and by Brouwer's continuity principle, we proved **IBS** . In this section, we consider the classical logic and type-2 Turing Computability [18].

Type 2 machines are proposed in computable analysis to generalize computability from finite to infinite sequences of symbols [18]. A Type 2 Turing machine is a Turing machine for which not only finite but also infinite sequences of symbols may be considered as inputs or outputs. Let Σ be a finite set of alphabets, and let Σ^ω denotes the set of all sequences $\alpha : \mathbb{N} \rightarrow \Sigma$.

Definition 3.1 *A Type 2 machine M is a Turing machine with a one-way input tape to input sequences from Σ^ω, an output tape to output strings in $\{1\}^*$, and finitely many work tapes. A function $f : \Sigma^\omega \to \mathbb{N} = \{1\}^*$ is type-2 computable, if and only if there is a type-2 machine M that for input instance α provide the output $f(\alpha)$.*

Theorem 3.2 *Let $f : \Sigma^\omega \to \{1\}^*$ be a total type-2 computable function. Then for all $\alpha \in \Sigma^\omega$, there exists m such that for all β, if $\overline{\beta}(m) = \overline{\alpha}(m)$, then $f(\beta) = f(\alpha)$.*

Proof. See [18], Theorem 2.2.6. $\dashv$

We define the *Classic Increasing Bounded Sequences* statement, **CIBS** , as conjunction of the following two statements:

1. for all $\alpha \in \mathrm{T}$, $\exists m \forall n \geq m(\alpha(n) = \alpha(m))$,
2. there is no type-2 computable function $f : \mathrm{T} \to \mathbb{N}$ such that $\forall \alpha \in \mathrm{T}[\max(\alpha, f(\alpha))]$.

The next two propositions show that the **CIBS** holds.

Proposition 3.3 *For each $\alpha \in \mathrm{T}$, $\exists k(\max(\alpha, k))$.*

Proof. Since $\neg\neg p \to p$ holds in classical logic, by Corollary 2.3, we are done. $\dashv$

Proposition 3.4 *There is no type-2 computable function $f : \mathrm{T} \to \mathbb{N}$, such that $\forall \alpha \in \mathrm{T}[\max(\alpha, f(\alpha))]$.*

Proof. The proof is similar to the proof of Theorem 2.4. Note that the weak continuity principle is valid for type-2 computable functions, by Theorem 3.2. $\dashv$

4 The Backward Induction

The *backward induction paradox* and its relation to the surprise exam paradox has been discussed in literature [12, 3, 15]. In this section, we consider backward induction on the centipede game in constructive logic. The centipede game is an extensive form game. There are 100 gold coins on a table. Two agents play a game on these coins as follows. Each agent, in its turn, either takes two coins and the game is over, or takes one coin and the other agent gets its turn.

The solution given by the backward induction is that the first player takes two coins and the game finishes. But it may seem paradoxical, since the experimental subjects regularly cooperate in the centipede game.

We claim that the way of use of backward induction in the centipede game is not acceptable in constructive logic. The *backward induction principle*, **BWI** , says:

$$\{\exists n \psi(n), \forall n(\psi(n+1) \to \psi(n))\} \vdash \psi(1).$$

Let $\psi(n)$ mean that "*the game is over in round n*" for some n, $1 \leq n \leq 100$. In fact, in argument for the centipede game, it is *not* proved that $\exists n\psi(n)$, but the weaker version, i.e., $\neg\neg\exists n\psi(n)$ is proved.

The backward argument on the centipede game usually goes as follows:

A1. Since the game may last at most 100 rounds, then *there exists* a round n which is the *last round*, i.e., $\exists n\psi(n)$.

A2. On the other hand, if one agent ends the game at round $n+1$, due to *rationality* of the other agent, the other one would have ended the game at the previous round n, i.e., $\forall n(\psi(n+1) \rightarrow \psi(n))$.

The argument $A1$ is not constructively valid. Since in constructive logic, to show "there exists n" means to "construct or present n". Argument $A1$, in constructive logic, just says that *it is not false* that there exists a round n that is the *last round.* Hence $\neg\neg\exists n\psi(n)$, and one of the premises of the **BWI** principle is not fulfilled. It is an easy exercise in constructive logic to show that the following deduction

$$\{\neg\neg\exists n\psi(n), \forall n(\psi(n+1) \rightarrow \psi(n))\} \vdash \psi(1).$$

in not valid. So we cannot deduce its conclusion, i.e., $\psi(1)$. However, it is routine to verify the following deduction in constructive logic.

$$\{\neg\neg\exists n\psi(n), \forall n(\psi(n+1) \rightarrow \psi(n))\} \vdash \neg\neg\psi(1).$$

We claim that we cannot deduce $\psi(1)$ from $\neg\neg\psi(1)$ in constructive logic, i.e.,

$$\nvdash \neg\neg\psi(1) \rightarrow \psi(1).$$

In constructive logic, $A \vee B$ is true whenever either we have an evidence (proof) for A or an evidence (proof) for B. Note that before the game starts, for each n, the round n is *in future* and has not occurred yet. Thus it is *unknown* whether round n is the *last round* or not, while backward induction argument is done *before* the game starts. Hence, constructively, for each n, $\nvdash \psi(n) \vee \neg\psi(n)$.

It is easy to show that for any formula A, if we have $\nvdash A \vee \neg A$ and $\vdash \neg\neg A$, then $\nvdash \neg\neg A \rightarrow A$. Since if we have $\vdash \neg\neg A \rightarrow A$, then by $\vdash \neg\neg A$, we can deduce $\vdash A$, and consequently $\vdash A \vee \neg A$.

5 Concluding Remarks

We introduced **IBS** as a reformulation of **SEP**. Almost all of the solutions of the surprise exam paradox in the literature are based on the critique of *teacher's announcement*, not *student's argument.* Some works present formal versions of the paradox such that teacher's announcement is a self-referential statement [14, 8, 3]. In some other works, teacher's announcement is an unsuccessful update [6]. In our reformulation of **SEP**, there is nothing wrong with teacher's announcement, but student's argument is *not* constructively valid.

The surprise exam paradox can be expressed for less than five days as well. For example, if the teacher says that *I will take an exam on Monday or Tuesday*, it is still a legal version of the paradox. However if we express the paradox for just one day, then the teacher's announcement is *false*. Our reformulation, i.e., **IBS** , works for any spread T_k, $k > 1$, consisting of all sequences α satisfying the conditions:

- $\forall n(1 \leq \alpha(n) \leq k)$,
- $\forall n(\alpha(n+1) = \alpha(n) \vee \alpha(n+1) = \alpha(n) + 1)$.

If $k = 1$, then the increasing bounded sequence statement, **IBS**, is false.

References

[1] M. van Atten, **On Brouwer**, Wadsworth Philosophers Series, 2004.

[2] R. Binkley, *The Surprise Examination in Modal Logic*, J. Phil, 1968.

[3] T. Y. Chow, *The Surprise Examination or Unexpected Hanging Paradox*, Amer. Math. Monthly, 1998.

[4] H. van Ditmarsch, B. Kooi, *The secret of my success*, Synthese 151, 2006.

[5] M. Dummett, **Elements of Intuitionism**, Oxford University Press, second edition, 2000.

[6] J. Gerbrandy, *The Surprise Examination in Dynamic Epsitemic Logic*, Synthese, 2007.

[7] W. Kaplan, R. Montague, *A paradox Regained*, Notre Dame J. Formal Logic 1, 1960.

[8] S. Kritchman, R. Raz, *The Surprise Examination Paradox and the Second Incompleteness Theorem*, Notices of the AMS, vol 57, 2010.

[9] K. Levy, *The solution to the Surprise Exam Paradox*, The Southern Journal of Philosophy, 2009.

[10] A. Marcoci, *The Surprise Examination Paradox, A review of two so called solutions in dynamic epistemic logic*, 2011.

[11] D. O'Conner, *Pragmatic Paradoxes*, Mind 57, 1948.

[12] P. Pettit, R. Sugden *The Backward Induction Paradoxes*, The Journal of Philosophy, Vol 86, 1989.

[13] W. Quine, *On a so-called Paradox*, Mind 62, 1953.

[14] R. Shaw, *The Paradox of the Unexpected Examination*, Mind 67, 1958.

[15] J. H. Sobel, *Backward Induction Arguments: A Paradox Regained*, Philosophy of Scince, 1993.

[16] R. A. Sorensen, *Conditional Blindspots and the Knowledge Squeeze: a solution to the prediction paradox*, Australian Journal. Phil, 1984.

[17] A. S. Troelstra, D. van Dalen, **Constructivism in Mathematics, An introduction**, Vols. 1, 2, North-Holland, 1988.

[18] K. Weihrauch, **Computable Analysis, An introduction**, Springer, 2000.

[19] P. Weiss, *The prediction paradox*, Mind 61 1952.

[20] T. Williamson, **Knowledge and its limits**, Oxford University Press, 2002.