منطق‌پژوهی، پژوهشگاه علوم انسانی و مطالعات فرهنگی
سال اول، شمارهٔ دوم، پاییز و زمستان ۱۳۸۹، صص ۱۰۳ ـ ۱۱۸

قضیهٔ ناتمامیت گودل و فلسفهٔ ذهن

کامران قیوم‌زاده*

چکیده

یکی از کاربردهای اساسی قضایای ناتمامیت گودل در فلسفه، نقش آن‌ها در استدلال‌هایی است که بین ذهن انسان از یک طرف و یک الگوریتم (ماشین) یا نظام صوری متناهی از طرف دیگر مقایسه به عمل می‌آورند. دو استدلال متمایز در این زمینه مطرح گشته است. در هر دو استدلال درک صدق جملهٔ گودل توسط انسان، به عنوان ملاکی برای تفوق بر هر ماشینی قلمداد شده است. اما ایرادهایی چند بر هر دو استدلال وارد است. در این مقاله با شرح و بسط استدلال‌های ذکر شده، به بررسی مجادلات و مباحثات دامنه‌داری که میان دو گروه مکانیک‌گرا و ضد مکانیک‌گرا وجود داشته است پرداخته‌ایم. با تحلیل دقیق قضایای ناتمامیت و ارتباط آن‌ها با معرفت حسابی بشر می‌توان چار چوب‌های موجود در زمینهٔ بحث مورد نظر را شکل داد. با تحلیل دقیق این مسأله می‌توان گفت که هیچ نوع دلیل قاطعی، با استفاده از این قضایا، برای تفوق ذهن بشر بر ماشین وجود ندارد؛ بلکه تنها می‌توان گفت که با اعمال این قضایا و انتخاب حوزهٔ شناخت‌پذیر ریاضی بشر به عنوان یک پیش‌فرض تقابل یا تعاند انسان و ماشین به چه معناست.

کلیدواژه‌ها: قضایای ناتمامیت گودل، مکانیک‌گرایی، ضد مکانیک‌گرایی، استدلال پن‌روز، هوش مصنوعی.

۱. معرفی قضایای ناتمامیت گودل

در سال ۱۹۳۱ کورت گودل (۷۸ـ۱۹۰٦) منطق‌دان اتریشی در مقاله‌ای دوران‌ساز و تاریخی موفق به اثبات دو قضیهٔ مهم در حوزهٔ منطق ریاضی گشت، که به قضایای ناتمامیت معروف

* استادیار گروه آموزشی فلسفه و منطق دانشگاه تربیت معلم آذربایجان ghayoom_k@yahoo.com
تاریخ دریافت: ۸۹/۱۰/۲۹، تاریخ پذیرش: ۸۹/۱۱/۲۵

شده‌اند. این قضایا معمولاً به عنوان مهم‌ترین دستاورد تاریخ منطق به حساب می‌آیند. اما این قضایا از چه سخن می‌گویند؟

به طور کلی می‌توان گفت این قضایا به ذکر ویژگی‌هایی از یک نظام صوری به اندازه کافی قوی که شامل حساب مقدماتی اعداد بشود (برای مثال PA) می‌پردازد. در قضیهٔ اول به جمله‌ای در نظام‌های صوری اشاره می‌شود که هر چند صادق است، قابل اثبات نیست؛ به شرط آن که نظام صوری سازگار باشد (G1). در قضیهٔ دوم اثبات‌ناپذیری سازگاری نظام صوری مورد بحث اثبات می‌شود (G2). می‌توان به اجمال و به صورت غیر صوری به چگونگی این استدلال پرداخت:

گودل نظام صوری اصل موضوعی شده‌ای (F) را در نظر گرفت که: ۱ـ سازگار بوده و ۲ـ به اندازهٔ کافی برای بیان حساب مقدماتی اعداد قوی باشد (گودل خود، نظام صوری اصول ریاضی وایتهد و راسل را مد نظر داشت). سپس با الهام از جملات خودارجاعی همچون جملهٔ دروغگو « این جمله کاذب است» به ساخت جملهٔ «من اثبات‌ناپذیرم» (که از این پس با G(F) نمایش می‌دهیم) پرداخت. این جمله در هر زبان صوری که ویژگی دوم ذکر شده در بالا را داشته باشد قابلیت صوری شدن دارد. حال گوئیم، اگر جملهٔ G(F) در F اثبات‌پذیر باشد، از آن جا که F سازگار است، بنابراین باید G(F) صادق باشد؛ یعنی «من اثبات‌ناپذیرم» صادق است و این با اثبات‌پذیری آن در تناقض قرار می‌گیرد. اما اگر جملهٔ G(F) در F اثبات‌ناپذیر باشد، آن گاه از آن جا که خود نیز به اثبات‌ناپذیری خود اذعان می‌کند، G(F) صادق خواهد بود. از آن جا که برای صدق جملهٔ G(F) سازگاری این نظام را مفروض گرفتیم، می‌توانیم جملهٔ اثبات‌شده توسط این نظام را با $\text{Con (F)} \rightarrow \text{G(F)}$ (Consistent = Con) نمایش دهیم.

از طرف دیگر، اگر $\neg$ G(F) از این نظام اثبات‌پذیر باشد، آن گاه با قانون رفع تالی به ناسازگاری این نظام می‌رسیم $\neg$ Con (F)، که با فرض سازگاری نظام در تناقض است. پس در نظام F، هیچ یک از دو جملهٔ G(F) و $\neg$ G(F) نیز اثبات‌پذیر نیستند. این نتیجهٔ قضیهٔ اول ناتمامیت گودل است. ازاین قضیه می‌توان نتیجه گرفت که Con(F) نیز اثبات‌پذیر نیست. زیرا اگر فرض کنیم Con (F) در F اثبات‌پذیر باشد، با اعمال قاعدهٔ وضع مقدم، G(F) نیز می‌باید در F اثبات‌پذیر باشد؛ در حالی که یکی از نتایج قضیهٔ اول ناتمامیت این بود که G(F) و $\neg$ G(F) در F اثبات‌پذیر نیستند (تناقض)؛ پس Con (F) در F اثبات‌پذیر نیست (قضیهٔ دوم ناتمامیت گودل).

پس از سال ۱۹۳۱ بحث‌هایی پیرامون تبعات فلسفی این قضایا مطرح گشت؛ بخشی از این مباحثات پیرامون فلسفه‌های ریاضی مطرح در آن زمان بود؛ پس از چند دهه، این قضایا موجب مباحثی بسیار جدی در فلسفهٔ ذهن و هوش مصنوعی شد. در بخش‌های زیر تنها به مباحثی که در فلسفه‌ی ذهن مطرح گشته می‌پردازیم.

۲ .شرایط اولیهٔ بحث

برای کاربرد قضایای ناتمامیت گودل در فلسفهٔ ذهن مجبور به اتخاذ شرایط و پیش‌فرض‌های لازم و ضروری هستیم. در بحث از ماشینی بودن ذهن بشر، دو گروه در مقابل یکدیگر قرار می‌گیرند: مکانیک‌گرایان و ضد مکانیک‌گرایان. ادعای مکانیک‌گرا چنین مطرح می‌گردد: ماشینی وجود دارد که خروجی آن با خروجی انسان و یا گروهی از انسان‌ها برابر است. اما ماشین‌ها و انسان‌ها دارای چه خصوصیاتی هستند؟ خروجی ماشین از چه نوع واژه‌هایی تشکیل شده است؟ در این بحث خروجی‌ها را تنها جمله‌هایی در نظر می‌گیریم که در زبان درجهٔ اول حساب پئانو ارائه می‌شود یا حتی کمی دقیق‌تر، بنابر نظر پن‌روز خروجی‌ها را به π_1 - جمله‌ها محدود می‌کنیم (Penrose, 1995).

حال با روشن‌شدن وضعیت خروجی‌ها، اگر هر انسانی را به صورت واقعی، با پوست و گوشت خود و عمری محدود در نظر بگیریم، تنها جملات محدودی از حساب وجود دارد که هر یک از انسان‌ها می‌توانند در طول زندگی‌شان به اثبات برسانند؛ حتی اگر تعدادی محدود از انسان‌ها را در نظر بگیریم، مجموعهٔ جملات اثبات‌شدهٔ آنها نیز محدود می‌گردد. در این حالت اگر مکانیک‌گرا ادعا کند که خروجی ماشین او با یکی از مجموعه‌های محدود بالا یکسان است، قضایای گودل برای بحث نامربوط به نظر می‌رسند. بنابراین برای پیش‌برد این بحث، شرایط ایده‌آلی را در نظر می‌گیریم که ماشین‌ها در آن دارای حافظه، گنجایش، مکان و زمان بالقوه نامتناهی برای انجام اعمال خود هستند و هرگز نقص پیدا نمی‌کنند. به عبارتی دقیق‌تر از بین سخت‌افزار و نرم‌افزار، سخت‌افزار کنار گذارده می‌شود (ماشین تورینگ). مشابه این ایده‌آل‌سازی، برای انسان‌ها نیز به کار می‌رود؛ بدین صورت که عمر، انرژی، گنجایش اطلاعاتی و مواد مصرفی در اختیار او را نامحدود تصور می‌کنیم و در دیگر جنبه‌های انسانی او را شبیه انسان واقعی در نظر می‌گیریم. همچنین در نظر بگیرید که ماشین‌ها و انسان‌ها هیچ اشتباهی را مرتکب نمی‌شوند و یا به عبارتی هر دو از نرم‌افزاری

خالی از اشتباه بهره می‌گیرند (Shapiro, 1998: 275-277)، البته این تمایز بین نرم‌افزار و سخت‌افزار، مورد قبول تعدادی از فلاسفه همچون ویتگنشتاین و کرایزل نیست. کریپکی نیز در (Kripke, 1982)، از جانب ویتگنشتاین به چنین امری اشاره می‌کند. ولی ما برای پیش‌برد بحث، ناچار از این تمایز و جداسازی هستیم؛ در غیر این صورت همان طور که اشاره شد کاربردی برای قضایای ناتمامیت گودل در این مسأله پیدا نمی‌شود.

۳. استدلال لوکاس و استدلال اول پن‌روز

اولین استدلالی که با اتکا بر قضایای گودل، در زمینهٔ مقایسه بین ذهن بشر و ماشین شکل گرفته است، به جان لوکاس، فیلسوف آکسفوردی مربوط می‌شود که در مقالهٔ «اذهان، ماشین‌ها و گودل» در سال ۱۹٦۱ انتشار یافت (Lucas, 1961: 112-137). بنابر این، می‌توان از لوکاس به عنوان اولین چهره در این زمینه نام برد. استدلال لوکاس را می‌توان این گونه بیان نمود: فرض کنید که مکانیک‌گرایی ماشینی به نام M را طراحی کرده و ادعا می‌کند که این ماشین می‌تواند همهٔ احکام صادقی را که توسط بشر بیان‌پذیرند، به عنوان خروجی ارائه دهد. فرض کنیم که M سازگار باشد. از آن جا که لوکاس قضیهٔ ناتمامیت گودل را می‌داند، پس جملهٔ گودل (G) ماشین M را می‌سازد. حال از آن جایی که لوکاس می‌داند که M نمی‌تواند G را به عنوان خروجی ارائه دهد و همچنین ادعا می‌کند که خود صدق این حکم (G) را می‌داند، پس با استفاده از این مقدمات، ادعای مکانیک‌گرا را رد می‌کند.

دو ایراد اصلی بر این استدلال وارد است. اولین ایراد که حتی قبل از انتشار مقالهٔ لوکاس مطرح گشته است، توسط پاتنام ارائه شد. (Putnam, 1960: 138-164) پاتنام در این مقاله عنوان می‌کند که تنها چیزی که ما دربارهٔ جملهٔ گودل می‌توانیم بگوئیم، دیدن صدق آن است؛ زیرا صدق جملهٔ گودل در یک نظام صوری مبتنی بر سازگاری آن نظام می‌باشد و این دقیقاً مسأله‌ای است که ما قادر به اثبات آن نیستیم. ایراد دوم از طرف پاول بناسراف صورت گرفته است (Benacerraf, 1967: 9-32). او اظهار می‌کند تنها نتیجه‌ای که می‌توان از قضایای ناتمامیت گودل اخذ کرد، این ادعاست که اگر چه شاید درست باشد که بگوئیم «من یک ماشین تورینگ هستم» ولی در این صورت دیگر «نمی‌توانیم به این ماشین شناخت داشته باشیم».

پن‌روز، فیزیک‌دان و ریاضی‌دان برجسته در دو کتاب (Penrose, 1989) و (Penrose, 1994) استدلال لوکاس را به شیوه‌ای دیگر بیان می‌دارد. او با استفاده از مسألهٔ توقف در ماشین‌های

تورینگ، علیه مکانیک‌گرایی اقامه‌ی دعوی می‌کند. ما برای روشن‌تر شدن استدلال لوکاس این استدلال را نیز مطرح می‌کنیم. صورت استدلال را می‌توان به شیوهٔ زیر بیان نمود:

ماشین‌های تورینگ دارای دو نوع محاسبهٔ متوقف‌شدنی و متوقف‌نشدنی هستند. برای مثال:

(A): عددی را بیابید که حاصل جمع سه عدد مربع نیست؛ نمونه‌ای برای محاسبات متوقف‌شدنی است. در حقیقت زمانی که ماشین به عدد ۷ می‌رسد، این عدد را به عنوان خروجی ارائه می‌دهد و متوقف می‌شود. در حالی که:

(B): عددی را بیابید که حاصل جمع چهار عدد مربع نیست، و یا

(C): عددی فرد را بیابید که حاصل جمع دو عدد زوج نباشد؛ از نمونه مثال‌هایی هستند که در آن ماشین متوقف نمی‌شود. در حقیقت قضیهٔ لاگرانژ در نظریهٔ اعداد می‌گوید که هر عددی حاصل جمع چهار عدد مربع است و با اثباتی ساده می‌توان نشان داد که هیچ عدد فردی حاصل جمع دو عدد زوج نیست. اما اگر توقف‌پذیری حدس گولدباخ را در نظر بگیریم که:

(D): عدد زوجی بزرگ‌تر از ۲ را بیابید که حاصل‌جمع دو عدد اول نباشد؛

مسألهٔ توقف‌پذیری آن بر ما مشخص نیست؛ زیرا این حدس تا به حال تأیید یا رد نشده است. پس می‌توان دید که مسألهٔ توقف بعضی از مسائل بسیار ساده (C و A) بعضی مشکل (B) و بعضی نیز تا به حال حل‌نشده (D) باقی مانده‌اند.

مسأله‌ای که برای توقف ماشین‌های تورینگ وجود دارد و در تناظر با قضیهٔ گودل است، مربوط به توقف‌ناپذیری الگوریتم می‌شود. می‌توان این مسأله را چنین عنوان نمود: آیا الگوریتمی وجود دارد که مسألهٔ توقف‌ناپذیری الگوریتم‌ها را مشخص کند؟ یعنی برای مثال عنوان کند که مسائل ... ,A, B, C, D متوقف می‌شوند یا خیر. در ابتدا فرض می‌کنیم که روش محاسباتی A در دسترس ماست. زمانی که A متوقف می‌شود، نتیجه می‌گیریم که الگوریتم C(n) متوقف نخواهد شد، جایی که C(n) ما را با خانواده‌ای از الگوریتم‌ها مرتبط می‌کند که به محاسبه روی اعداد طبیعی می‌پردازند. برای مثال الگوریتم:

(E): عددی را بیابید که حاصل‌جمع n عدد مربع نباشد. به ازای ۳، ۲، ۱، ۰ = n اعداد ۷ و ۳ و ۲ و ۱ را به عنوان خروجی ارائه می‌دهد و به ازای $n \geq ٤$ بنابر قضیهٔ لاگرانژ متوقف نمی‌شود. و:

(F) : عدد فردی را بیابید که حاصل‌جمع n عدد زوج نباشد؛ به ازای هیچ مقداری از n متوقف نخواهد شد. همچنین فرض می‌کنیم A تمامی روش‌های محاسباتی ریاضیدانان را

تحت اختیار دارد که متوقف نشدن الگوریتم‌ها را حساب می‌کنند. گوییم A هرگز به ما جواب‌های کاذب ارائه نمی‌دهد؛ یعنی اگر A بگوید که $C(n)$ متوقف نمی‌شود، آن گاه $C(n)$ در حقیقت نیز متوقف نخواهد شد. اگر A از این شرایط پیروی کند، این الگوریتم را درست (Sound) می‌نامیم. اما اگر A نادرست باشد، می‌توان با یک محاسبهٔ دقیق روشن کرد که A ابطال‌پذیر نیست. به این طریق که اگر A به غلط اظهار کند $C(n)$ متوقف نمی‌شود، در حالی که $C(n)$ متوقف شود، سرانجام با توقف $C(n)$، الگوریتم A ابطال می‌شود.

اگر بخواهیم A را در محاسبات کلی خود وارد کنیم، لازم است که تمامی الگوریتم‌ها را کدبندی نماییم: $C_0 , C_1 , C_2 , C_3 ,....$ آنگاه به C_q به عنوان qمین الگوریتم اشاره می‌کنیم و زمانی که عدد n را به عنوان ورودی برای چنین الگوریتم‌هایی در نظر می‌گیریم، آن‌ها را به طریق زیر نمایش می‌دهیم $C_0(n) , C_1(n) , C_2(n), ...$. می‌دانیم که چنین لیستی از الگوریتم‌ها، محاسبه‌پذیر است؛ یعنی الگوریتمی وجود دارد که با ورودی‌های q ، n خروجی $C_q(n)$ از آن حاصل می‌شود. با این تفاصیل، اگر ورودی‌های n,q به الگوریتم A داده شود و A متوقف گردد، نتیجه می‌گیریم که $C_q(n)$ هرگز متوقف نخواهد شد. به عبارت دیگر:

(۱): اگر $A(q,n)$ متوقف شود، آن گاه $C_q(n)$ متوقف نمی‌شود.

با استفاده از روش قطری کانتور در (۱) q را مساوی n قرار می‌دهیم.

(۲): اگر $A (n,n)$ متوقف شود، آن گاه $C_n(n)$ متوقف نمی‌شود.

$A(n,n)$ تنها به یک عدد وابسته است و باید در لیست الگوریتم‌های $C_0 , C_1 , C_2 , C_3 ,....$ قرار گرفته باشد. این لیست به الگوریتم‌هایی اشاره می‌کند که روی عدد n محاسبه انجام می‌دهند. فرض می‌کنیم که $A = C_k$ باشد، پس:

(۳): $A(n,n) = C_k(n)$

حال برش قطری دوم را اجرا می‌کنیم و n را مساوی k قرار می‌دهیم و از (۳) نتیجه می‌گیریم:

(٤): $A (k , k) = C_k (k)$ و از (۲) با $n=k$

(۵): اگر $A(k , k)$ متوقف شود، آن گاه $C_k(k)$ متوقف نمی‌شود.

با جایگزینی $C_k(k)$ که مساوی $A(k , k)$ است، در (۵) داریم:

(٦): اگر $C_k(k)$ متوقف شود، آن گاه $C_k(k)$ متوقف نمی‌شود.

از (٦) نتیجه می‌گیریم که $C_k(k)$ در حقیقت متوقف نمی‌شود و از آن جا که $A(k , k) = C_k(k)$ پس $A(k , k)$ نیز متوقف نخواهد شد. پس روش A قادر نیست مشخص کند که

$C_k(k)$ متوقف نمی‌شود، حتی اگر در حقیقت $C_k(k)$ متوقف نشود. به علاوه، اگر می‌دانیم که A درست است، پس می‌دانیم که $C_k(k)$ متوقف نمی‌شود. بنابر این، واقعیتی را می‌دانیم که A قادر به اظهار آن نیست؛ پس می‌توان نتیجه گرفت که A قادر نیست فهم ما را تحت اختیار بگیرد. حال می‌توان گفت هیچ مجموعه‌ای از قواعد محاسباتی که به طریقی درست شناخت‌پذیر باشد، نمی‌تواند برای تصمیم‌گیری درباره الگوریتم‌هایی که متوقف نمی‌شوند کامل باشد. زیرا الگوریتم‌هایی هم چون $C_k(k)$ وجود دارند که از این مجموعه قواعد تخطی می‌کنند. به علاوه از شناخت A و درستی آن می‌توانیم الگوریتم $C_k(k)$ را بسازیم که پی به عدم توقف آن ببریم. پس می‌توان حکم زیر را نتیجه گرفت.

T: ریاضی‌دانان برای تعیین صدق ریاضی، از الگوریتم درستی که شناخت‌پذیر باشد استفاده نمی‌کنند.

T نتیجهٔ استدلال اول پن‌روز است که ذهن بشر را با الگوریتم درستی که شناخت‌پذیر است مقایسه می‌کند و به محال بودن وجود چنین الگوریتمی برای انسان حکم می‌دهد.

پن‌روز معتقد است که پدیدارهای محاسبه‌ناپذیر، تنها می‌توانند در ذهن موجودات زنده شکل گیرند. پدیدارهایی همچون دیدن رنگ‌ها، احساس درد و یا درک یک نوای موسیقی که به شعور (Awareness) موجودات وابسته است. شعور جنبهٔ منفعل آگاهی (Consciousness) در نظر گرفته می‌شود. در حالی که پدیدارهایی هم چون برخاستن از خواب یا تصمیم برای انتخاب امری، که به ارادهٔ آزاد (will Free) وابسته‌اند، جنبهٔ فعال آگاهی در نظر گرفته می‌شوند. در واقع آگاهی دارای هسته‌ای مرکزی است که تمامی جنبه‌های مختلفش (شعور و ارادهٔ آزاد) به آن هسته وابسته‌اند و این اعمال فعالانه و منفعلانه تنها اعمال عالم فیزیکی‌اند که محاسبه‌ناپذیرند. درک صدق جملهٔ گودل نیز به این هستهٔ مرکزی وابسته است که خاص انسان‌ها است و ویژگی آن، درک متقنی (Unassailable) است که از طریق براهین ریاضی به دست می‌آید؛ در حالی که چنین امری برای سایر پدیدارهای ذهنی میسر نیست. پس با استفاده از قضیهٔ گودل می‌توان پدیدارهای محاسبه‌ناپذیر ذهن را ثابت کرد. پن‌روز با استدلال بر وجود چنین پدیدارهایی، فیزیک جدیدی را پیشنهاد می‌کند که علاوه بر پیش‌بینی پدیدارهای محاسباتی که در فیزیک کوانتوم و نسبیت و همین طور فیزیک کلاسیک قابل عرضه است، به پیش‌بینی پدیدارهای محاسبه‌ناپذیر نیز بپردازد و به عقیده‌ی او این فیزیک، تنها با تلفیق فیزیک کوانتوم و فیزیک نسبیت میسر است. (Penrose, 1994)

۴. استدلال دوم پن‌روز

با توجه به ایرادهایی که بر استدلال لوکاس گرفته شده است و از آن جا که استدلال اول پن‌روز، تنها حکم به عدم امکان الگوریتم درستی می‌دهد که برای انسان‌ها شناخت‌پذیر باشد، بنابراین می‌توان الگوریتم‌هایی را در نظر گرفت که نادرست (نظر پاتنام و در پاره‌ای از موارد تورینگ) و یا شناخت‌ناپذیر (نظر بناسراف) باشند. از این رو پن‌روز به ارائهٔ استدلال دیگری می‌پردازد که سازگاری نظام را از پیش، فرض نمی‌کند. استدلال دوم پن‌روز به‌گونهٔ زیر است:

ریاضیدانی را در نظر بگیرید که تحت نظام F قرار گرفته است و زمانی که می‌گوید «من F هستم» به این معناست که «F تمامی روش‌های دسترس‌پذیر برهانی ریاضی بشر را دربر می‌گیرد». ریاضیدان چنین می‌گوید:

A: اگر چه من نمی‌دانم که ضرورتاً F هستم، اما می‌توانم بگویم که اگر F بودم، دستگاه F باید درست (Sound) می‌بود. و حتی می‌توان بیشتر از این گفت که F' نیز باید درست می‌بود، جایی که F' از جمع F و «من F هستم» به دست می‌آید. از این که من F هستم می‌توان گفت که جملهٔ گودل G(F') باید صادق باشد و به علاوه این جمله نتیجه‌ای از دستگاه F' نیست. بنابر این من درست فهمیده‌ام که «اگر من F باشم آن گاه G(F') باید صادق باشد» و این دقیقاً همان چیزی است که F انجام می‌دهد. بنابراین من قادرم چیزی را درک کنم که فرای توانایی F است و از این جا نتیجه می‌گیرم که من به هیچ وجه نمی‌توانم F باشم. (Penrose, 1995) در این استدلال، پن‌روز از سازگاری F استفاده نمی‌کند و در واقع این استدلال گودلی از نوع جدیدی به حساب می‌آید.

این استدلال دارای چند نقص اساسی است که می‌توان به مواردی از آن اشاره کرد:

I- بنابر اظهارات خود پن‌روز، اگر جملات ریاضی را به π_1– جمله ها محدود کنیم، آن گاه مشخص نیست که چرا «من F هستم» باید یک π_1– جمله باشد.

II- در این استدلال از محمول «دانستن» استفاده شده است که واژه‌ای مبهم و مفهومی (Intentional) است و نمی‌توان آن را وارد نظام‌های صوری نمود.

III- ایراد سوم به «درستی» (Sound) نظام‌ها بر می‌گردد که هر گونه تعریفی از این واژه به شرط برخورداری از دو شرط ۱. از راست (True) بودن نظام، درستی آن نتیجه می‌شود و ۲. درستی در زبان نظام تعریف‌پذیر باشد، به عدم اعتبار هر صورت اصلاح شده‌ای از

استدلال پن‌روز می‌انجامد. این ایـراد توسـط پرلیندسـتروم (Lindstrom, 2001: 241-250) اظهار شده است که صورت آن در ذیل می‌آید:

فـرض کنیـد $Sd(F)$ فرمـول مناسـبی بـرای «F درسـت اسـت» باشـد. همچنـین $HC(F)$ (Humanly complete) به معنای «F تمامی روش‌های دسترس‌پذیر براهین ریاضی را در برمی‌گیرد» به کار رود. زمانی که پن‌روز اعلام می‌کنـد کـه «مـن F هسـتم» ایـن جملـه مستلزم $HC(F)$ و $Sd(F)$ است. هم چنـین $F \vdash S \rightarrow S'$ ا ت ا $F + S \vdash S'$ و تعریـف آخـر عبارت است از $F^{+} = F + Sd(F)$. می‌توان استدلال پن‌روز را به صورت زیر نمایش داد:

:(B)

۱	$Sd(F) \Rightarrow Sd(F^{+})$	فرض
۲	$Sd(F^{+}) \Rightarrow G(F^{+})$	برهان گودل
۳	$Sd(F^{+}) \Rightarrow F^{+} \nvdash G(F^{+})$	برهان گودل
۴	$Sd(F) \Rightarrow G(F^{+})$	(۱)، (۲)، قانون تعدی
۵	$Sd(F) \Rightarrow F^{+} \nvdash G(F^{+})$	(۱) ، (۳)، قانون تعدی
۶	$HC(F) \Rightarrow F \vdash Sd(F) \longrightarrow G(F^{+})$	فرض $HC(F)$ و برهان ریاضی بودن (۴)
۷	$\neg (Sd(F) \,\&\, HC(F))$	(۵) و (۶)
۸	¬ "من F هستم"	(۷) و تعریف من F هستم.

البته استدلال دوم پن‌روز کاملاً با استدلال (B) منطبق نیست، بلکـه بایـد (۱) را بـا (۹) جایگزین کرد.

۹. «من F هستم» $\Rightarrow$ Sd (F + «من F هستم»)

مهم‌ترین نکته در این استدلال، تعریف «درستی» است. اگر (۱) راست باشد، آنگاه برای هر توسیع E ای از F که در L (زبان F) صورت‌بندی شود، می‌توان حکم زیر را وارد کرد:

$$Sd(E) \Rightarrow Sd(E^{+})$$

می‌توان به جای این حکم ازحکم ضعیف‌تر (۱۰) نیز استفاده کرد:

۱۰. $Sd(E) \Rightarrow Con(E^{+})$

به‌خصوص زمانی که $E = F$ می‌توان (۱۰) را جایگزین (۱) کرد. حال اگر بتوان B را با این جایگزینی بی‌اعتبار کرد، استدلال اصلی B نیز بی‌اعتبار می‌شود.

هر تعریفی از «درستی» که شامل شروط زیر باشد، برای اعتبار بخشیدن به استدلال B ممکن نیست.

این شرایط عبارتند از :

۱. $F \Rightarrow Sd(F)$ راست است .۱۱

۲. «درستی» در L تعریف‌پذیر است (اگر چنین نباشد برای مثال (٦) نمی‌تواند صادق باشد).

برای آن که صدق (۱۰) از عمومیت خارج شود، از روش خودارجاعی برای ساخت جملهٔ S استفاده می‌کنیم (با در نظر گرفتن این حقیقت که $Sd(F + x)$ فرمولی از L با متغیر عددی آزاد x است).

۱۲. $F \vdash S \Leftrightarrow \neg Sd\,(F + S)$

فرض می‌کنیم که F راست است. با فرض $\neg\ Sd\,(F + S)$ و (۱۲) به راست بودن S می‌رسیم و چون F و S هر دو راست هستند، $F + S$ نیز راست است و با استفاده از (۱۱) نتیجه می‌گیریم که $Sd\,(F + S)$ و تناقض حاصل می‌شود. بنابر این باید نتیجه بگیریم کـه $Sd\,(F + S)$. از آن جا که با عبارت (۱۲) به نادرست بودن S می‌رسیم، پس $F + S$ نیز نادرست است و این نتیجه حاصل می‌شود که درستی (Sound) مستلزم راستی (True) نیست. با استفاده از (۱۲) می‌بینیم که : $E = F + S$ و $\neg Con\,(F + S + Sd\,(F + S))$ مثال نقضی برای (۱۰) می‌شود. این استدلال بدون در نظر گرفتن چگونگی تعریف «درستی »B را نامعتبر می‌کند.

حال می‌توان از B به A رجوع کرد. این امر با جانشینی (۹) به جای (۱) به دست می‌آید. می‌توان اصل ضعیف‌تر (۱۳) را جانشین (۹) کرد.

۱۳. «من E هستم» $\Rightarrow$ Con (E + «من E هستم»)

همچون (۱۰) نمی‌توان مثال نقضی برای (۱۳) به دست آورد؛ حتی اگر A نامعتبر باشد. اما ممکن است هیچ دستگاه E ای نتواند «من E هستم» را تصدیق کند. بنابراین اگر چه ممکن است (۱۳) صادق باشد، اما متقن بودن آن واضح نیست. در ضمن در استدلال (A)، «من F هستم» یک جملهٔ ریاضی در L فرض شده است. در حالی که اگر این فرض کاذب باشد، نمی‌توان گفت که چگونه باید («من E هستم» + Sd (E یا («من E هستم» + Con (E و یا («من E هستم» + G (E را فهمید و یا حتی این حکم که مطابق با (٦) است : («من F هستم» + G (F

$\Rightarrow$ «من F هستم» $\vdash F \Rightarrow HC(F)$ نمی‌تواند راست باشد و یا حتی معنی بدهد. بنابراین A از اعتبار آشکاری برخوردار نیست؛ در حالی که در استدلال B احتیاجی به فرض «من E هستم» به عنوان جمله‌ای ریاضی نیست.

IV- در سال ۲۰۰۳ شاپیرو در مقاله‌ای (Shapiro, 2003: 19-42) ایراد چهارمی را بر استدلال دوم پن‌روز مطرح کرد که می‌توان به طور خلاصه آن را چنین بیان نمود:

پن‌روز در استدلال دوم خود، در مقدمات به کار گرفته شده، از یک محمول صدق نامحدود (محمول شناخت‌پذیری نامحدود) استفاده می‌کند و بنابراین، چنین مقدماتی ناسازگار هستند و هر شیوه‌ای برای محدود کردن این محمول‌ها، یا به عدم اعتبار استدلال پن‌روز منجر می‌شود و یا باید پیش‌فرض‌هایی را قبول کند که مکانیک‌گرا می‌تواند به راحتی آن‌ها را رد کند.

همان‌طور که مشاهده شد، استدلال دوم پن‌روز نیز مانند استدلال اول وی، و همچنین استدلال لوکاس از اعتبار کافی برخوردار نیست. بنابراین بهتر است که این بحث را از زاویه‌ای دیگر مورد بررسی قرار دهیم.

۵. استلزامات پذیرش نگرهٔ مکانیک‌گرایی

فرض کنید K مجموعهٔ تمامی قضایا و احکام حساب درجهٔ اولی است که توسط انسان اثبات و شناخته می‌شود (ریاضیات ذهنی). برای سادگی بحث، فرض می‌کنیم تنها اعداد گودل این قضایا و احکام، در K وجود دارند. مکانیک‌گرایان ادعا می‌کنند که همهٔ شیوه‌های حسابی بشر، الگوریتم‌هایی بیش نیستند و با استفاده از نگرهٔ چرچ اظهار می‌کنند ماشین تورینگی وجود دارد که K را می‌شمارد و یا به عبارتی K شمارش‌پذیر بازگشتانه است. بر خلاف این گرایش، ضدمکانیک‌گرایان به وجود شیوه‌ها و قاعده‌هایی در کشف ریاضی و حساب توسط انسان اشاره می‌کنند که توسط هیچ ماشین تورینگی شبیه‌سازی نمی‌شود.

فرض کنید T (ریاضیات عینی) مجموعهٔ تمامی حقایق در حساب درجه اول باشد. فرض کنیم $K \subseteq T$. در این جا می‌توان دو شرط را مورد بررسی قرار داد: ۱- $K = T$ و ۲- $K \subset T$ (Shapiro, 1998, 273-302).

در ابتدا فرض می‌کنیم $K = T$ و هم چنین S جمله‌ای از حساب باشد. با استفاده از منطق دو ارزشی $S \in K$ و یا $\neg S \in K$. در مورد اول خواهیم داشت $S \in K$ و بنابر این S

«اصولاً» شناخت‌پذیر خواهد بود. در حالت دوم $\neg S \in K$ و کاذب بودن S «اصولاً» شناخت‌پذیر خواهد بود. بنابر این اگر $T = K$، آن گاه ریاضی‌دانان، می‌توانند صدق و کذب هر جمله‌ای از حساب مانند S را مشخص کنند. به عبارتی هر جمله‌ای از حساب، توسط بشر تصمیم‌پذیر است. از طرف دیگر با استفاده از قضیهٔ تارسکی می‌دانیم که T در زبان حساب تعریف‌پذیر نیست و یا به عبارتی دیگر T شمارش‌پذیر بازگشتانه نمی‌باشد. در حقیقت هیچ دستگاه صوری استنتاجی و یا ماشین تورینگی وجود ندارد که تمامی حقایق ریاضی موجود را به عنوان قضیه اعلام کند. پس اگر $K = T$ قضیهٔ تارسکی بر K نیز اعمال می‌گردد و بنابر این در می‌یابیم که K شمارش‌پذیر بازگشتانه نیست و در نتیجه مکانیک‌گرا بر خطا است.

برای جدی‌گرفتن مکانیک‌گرا باید فرض دوم ($K \subset T$) را در نظر گرفت. با فرض این که $S \in T$ و $S \notin K$ جمله‌ای از حساب را یافته‌ایم که مطلقاً تصمیم‌ناپذیر است و البته نقیض آن نیز $\neg S$ به همین صورت. پس اعتبار ادعای مکانیک‌گرا در وجود جملاتی از حساب است که هیچ فاعل ایده‌آلی قادر به اثبات آن نباشد و این جملات، مطلقاً تصمیم‌ناپذیر باشند.

گودل خود در سخنرانی گیبس اظهار کرده بوده است که اگر K شمارش‌پذیر بازگشتانه باشد، آن‌گاه جمله‌ای مطلقاً تصمیم‌ناپذیر به‌صورت $\forall x_1 \ldots \forall x_n \exists y_1 \ldots \exists y_m (Px_1 \ldots x_n y_1 \ldots y_m = ۰)$ وجود دارد که از درجهٔ ٤ یا پایین‌تر است. او می‌گوید یا ریاضیات پایان‌ناپذیر است به این معنی که اصول موضوعهٔ بدیهی آن هیچ‌گاه محدود به مجموعه‌ای از قواعد متناهی نمی‌شود و ذهن انسان به طور مداوم و همیشگی بر هر ماشین متناهی، تسلط می‌یابد و یا مسائل دیوفانتی مطلقاً حل‌ناپذیری از نوع معین وجود دارد.

پس کسانی همچون گودل، پن‌روز و لوکاس معتقدند که $K = T$ و حاضر به قبول هیچ جملهٔ مطلقاً تصمیم‌ناپذیری نمی‌باشند. در حقیقت آن‌ها به وجود شیوه‌هایی غیر صوری و غیر الگوریتمی در ذهن انسان اذعان دارند که موفق به کشف جملات صادقی در حساب می‌گردد که توسط هیچ ماشین و یا الگوریتمی اثبات‌پذیر نیست. از طرف دیگر، اگر K شمارش‌پذیر بازگشتانه باشد، آن گاه بعضی از جملات حساب برای انسان تصمیم‌ناپذیر است. مکانیک‌گرایان از وضع مقدم استفاده می‌کنند؛ یعنی می‌پذیرند که K شمارش‌پذیر بازگشتانه است و نتیجه می‌گیرند که $K \neq T$. در حالی که ضدمکانیک‌گرایان از رفع تالی استفاده می‌کنند، یعنی می‌پذیرند که $K = T$ و نتیجه می‌گیرند K شمارش‌پذیر بازگشتانه نیست.

اما اگر مکانیک‌گرا بر حق باشد و K شمارش‌پذیر بازگشتانه باشد؛ چه مستلزماتی حاصل

می‌آید؟ e را به عنوان عدد گودل ماشین تورینگی در نظر می‌گیریم که K را می‌شمارد. پس K = W_e. فرض می‌کنیم که شرایط استنتاج‌پذیری هیلبرت - برنیز برقرار است؛ یعنی:

۱. برای هر S در زبان حساب اگر $S \in W_e$ آن گاه جمله‌ای حسابی که معین می‌کند $S \in W_e$ در W_e است.

۲. برای هر جملهٔ P و S در زبان حساب، جمله‌ای از حساب که معین می‌کند:

اگر $(S \to P) \in W_e$ آن گاه اگر $S \in W_e$ آن گاه $P \in W_e$ در W_e است.

۳. برای هر جملهٔ S در زبان حساب، جمله‌ای از حساب که معین می‌کند [اگر $S \in W_e$ آن گاه این حکم که $S \in W_e$ در W_e است] خود در W_e است. در این نمادشناسی، محمول اثبات‌پذیری با فرمول $x \in W_e$ بیان شده است. شرایط ۱ و ۳ بیان کنندهٔ این امرند که اگر S اثبات‌پذیر است، آن گاه این جمله نیز اثبات‌پذیر است که S اثبات‌پذیر است. و جملهٔ ۲ بیان می‌کند که جملات اثبات‌پذیر تحت قاعدهٔ وضع مقدم بسته هستند.

Con_e جمله‌ای از حساب است که بیان می‌کند W_e سازگار است. با فرض این که Con_e صادق باشد، قضیهٔ دوم ناتمامیت گودل مستلزم این امر است که Con_e در W_e و از آن جا در K نباشد. بنابر این Con_e طبق فرض مکانیک‌گرا، مطلقاً تصمیم‌ناپذیر است؛ یعنی اگر K = W_e و شرایط استنتاج برقرار باشد، آن گاه هیچ کس نمی‌تواند سازگاری W_e را بداند. بنابر این می‌توان گفت که هیچ‌کس نمی‌تواند بداند که $W_e \subseteq T$ و همین طور عبارت $K \subseteq T$ تنها یک فرض است که هیچ‌کس آن را نمی‌داند و اگر کسی به صدق $K \subseteq T$ مطمئن نباشد، آن گاه نمی‌تواند صدق جمله‌ی $K \subseteq W_e$ را بداند. تنها راهی که کسی بتواند صدق $K \subseteq W_e$ را بداند، آن است که درستی W_e و یا K را بشناسد و این امر به صورتی متقن امکان‌پذیر نیست. پس اگر مکانیک‌گرا صادق باشد آن گاه ماشین تورینگی وجود ندارد که ما بتوانیم بدانیم تمامی جملات حساب را که شناخت‌پذیر هستند، می‌شمارد. این همان نتیجه‌ای است که بناسراف در پاسخ لوکاس گفته بود: حتی اگر انسان ایده‌آل یک ماشین تورینگ باشد، نمی‌تواند این حقیقت را بداند. اگر ما درست بودن W_e را ندانیم، آن گاه نمی‌توانیم W_e را به عنوان موکل خود برای جستجوی حقایق ریاضی انتخاب کنیم.

لوکاس در جواب خود به ایرادهایی که عنوان شده است اظهار می‌کند اگر F نظامی صوری باشد که مکانیک‌گرا به عنوان مدلی برای K در نظر گرفته است، دارای دو جملهٔ G_F به عنوان جملهٔ گودل و همچنین Con_F به عنوان حکم سازگاری F است. او ادعا می‌کند که

تحت این شرایط می‌داند که G_F صادق است. پاتنام این ایراد را وارد می‌کند که لوکاس صدق G_F را نمی‌داند، بلکه تنها چیزی که لوکاس می‌تواند بداند، صدق این جمله است که: اگر F سازگار است، آن گاه G_F صادق است؛ و این جمله‌ای است که هر ماشین و یا هر نظام صوری می‌داند. لوکاس از آن جا که نمی‌تواند Con_F را ثابت کند، تعیین سازگاری نظام را به مکانیک‌گرا واگذار می‌کند: اگر مکانیک‌گرا ادعا کند که نظام صوری و یا ماشین او ناسازگار است، این مسأله بی‌ارزش می‌شود؛ اما اگر به سازگاری نظام صوری اذعان کند، با وضع مقدم می‌توان به صدق G_F رسید و این جمله‌ای است که لوکاس می‌داند و ماشین از اثبات آن عاجز است. در این جا نیز می‌توان شاهد بود که قول مکانیک‌گرا بر سازگاری نظام صوری با یقین ریاضی برابر نیست و بنابر این برای اعمال وضع مقدم نمی‌توان به قول مکانیک‌گرا تکیه کرد.

همچنین مباحثی با ابتناء بر اعداد ترتیبی و به پشتیانی از نگرهٔ ضدمکانیک‌گرا توسط لوکاس در سال‌های اخیر (Lucas, 1996) صورت گرفته است؛ ولی می‌توان با کمی دقت مشاهده نمود که این گونه براهین و استدلال‌ها، در حقیقت صورت دیگری از استدلال‌های پیشین هستند که بدون آن که به حل مسأله‌ای منجر شده باشند، تنها جای مسأله را از جملات ریاضی به اعداد ترتیبی مربوطه انتقال داده است.

نتیجه‌گیری

با توجه به مباحث گذشته قضایای ناتمامیت گودل، اگر چه نمی‌توانند پاسخی قاطع به مسألهٔ ما بدهند، اما باعث به وجود آمدن محدودیت‌هایی خاص برای هواداران مکانیک‌گرایی و ضدمکانیک‌گرایی شده است. در ابتدا برای کاربرد این قضایا، مجبور به پذیرش شرایطی ایده‌آل برای انسان و ماشین می‌شویم (انسان‌ها فاقد محدودیت حافظه‌ای، زمانی و مکانی می‌باشند و ماشین را نیز ماشین تورینگ در نظر می‌گیریم). در غیر این صورت این قضایا در این مبحث دارای هیچ گونه کاربردی نمی‌باشند.

با فرض K برای ریاضیات ذهنی و T برای ریاضیات عینی، اگر *بدانیم* که $K = T$ آن گاه با توجه به قضیهٔ تارسکی می‌توان ادعا کرد که ضد مکانیک‌گرا بر حق است. اما این مسأله‌ای است که تاکنون با هیچ روش معرفت‌شناسانه‌ای به اثبات نرسیده است. از طرف دیگر اگر *بدانیم* که $K \subset T$، آن گاه انسان سازگاری خود را خواهد دانست و بنابر

این نگرهٔ مکانیک‌گرا غیر ممکن می‌شود.[۱] اما از آن جا که انسان نمی‌تواند حکم $K \subset T$ و سازگاری ماشین M را بداند، بنابر این می‌تواند ادعا کند که یا بر هر ماشینی برتری دارد و یا ماشین ناسازگار است.

حال فرض کنیم که $K = T$. اگر انسان صدق حکم $K = T$ را *نداند*؛ آن گاه بدون این که انسان بداند، بر هر ماشینی برتری دارد. از طرف دیگر اگر مکانیک‌گرا بر حق باشد و $K \subset T$ و انسان صدق نگرهٔ مکانیک‌گرا و صدق حکم $K \subset T$ را نداند، آن گاه انسان هیچ گاه نمی‌تواند بداند که الگوریتمی خاص بر ذهن او حاکم است و حتی نمی‌تواند بداند که چنین الگوریتمی سازگار یا ناسازگار است. اگر انسان *نداند* که $K \subset T$ و نگرهٔ ضدمکانیک‌گرا صادق باشد، آن گاه سازگاری خود را نخواهد دانست و بنابر این برتری خود را بر هر ماشینی نخواهد شناخت.

حالت بعدی این مسأله عبارت از ناسازگاری انسان ایده‌آل می‌باشد، یعنی $K \cap T \not\subseteq T$. در این مرحله تنها می‌توان گفت که انسان نمی‌تواند بر ناسازگاری خود شناخت داشته باشد.[۲] در این وضعیت، اگر ضدمکانیک‌گرا صادق باشد، آن گاه به دلیل ناسازگاری انسان ایده‌آل و برتری او بر ماشین، سازگاری ماشین ضروری است. اما اگر نظریهٔ مکانیک‌گرا صادق باشد، آن گاه ماشین نیز بالطبع ناسازگار است و او نمی‌تواند بداند که ماشین و یا الگوریتمی ناسازگار بر ذهن او حاکم است.

مسألهٔ دیگر آن که اگر جهان بر طبق الگوریتم سازگار خاصی حرکت کند، آن گاه انسان قادر به شناسایی چنین الگوریتمی نمی‌باشد؛ زیرا از آنجا که چنین الگوریتمی بر ذهن انسان نیز حاکم است و ذهن انسان در چنین حالتی سازگار می‌شود، مطابق آن چه در بالا گفته شد (عدم شناسایی بر $K \subset T$ و صدق نگرهٔ مکانیک‌گرا) انسان نمی‌تواند وجود چنین الگوریتمی را بشناسد و لذا وجود این الگوریتم را بر کل جهان درک نمی‌کند. پس نظر پن‌روز دربارهٔ فیزیک حاکم بر پدیده‌های محاسبه‌ناپذیر، تنها یکی از امکان‌های منطقی بحث است.

در پایان باید گفت که راه کارهایی که می‌توانند به حل این مسأله کمک کنند، علاوه بر بحث‌های فلسفی و تحلیلی فوق، به چند دسته تقسیم می‌شوند:

۱. اثبات و یا رد سازگاری انسان و ماشین به روش‌های تجربی و استقرایی، از طریق پیشرفت‌هایی که ممکن است در آینده در حوزه‌های فیزیولوژی، پزشکی و یا سایر علوم تجربی روی دهد؛

۲. راه‌حل‌هایی که برای جملات خودارجاع پیشنهاد می‌شود.

۳. پیشنهاد روش‌های جدید برای الصاق و یک‌پارچه کردن اثبات‌های صوری، شهودی و سایر برهان‌های غیرصوری تحت یک مقوله.

پی‌نوشت

۱. زیرا در این صورت K تنها شامل صدق های ریاضی و یا حساب خواهد بود.

۲. تنها در صورتی می‌توان ناسازگاری انسان را پذیرفت که از منطق فراسازگاری (Paraconsistent) استفاده کند.

منابع

Benacerraf, P. (1967), "God, the devil, and Gödel", *The Monist* 51 9-32.

Kripke, S. (1982), *Wittgenstein on rules and private language*, Harvard University Press.

Lindstrom, P. (2001), "Penrose's new argument", *Journal of Philosophical Logic* 30 241 – 250.

Lucas, J. R. (1961), "Minds, machines and Gödel", *Philosophy* 36 112 – 37.

Lucas, J. R. (1996), *Minds, machines and Gödel : A retrospect , Machines and thought : The legacy of Alan Turing*, Vol. 1 edited by Milican, P. and Clark, A., Oxford University Press.

Penrose, R.(1989). *The emperor's new mind: concerning computers, minds, and the laws of physics*, Oxford University Press,

Penrose, R.(1994). *Shadows of the mind: A search for the missing science of consciousness*, Oxford University Press,

Penrose, R. (1995), "Beyond the doubting of shadow: A reply to commentaries on "Shadows of the mind" ", *Psyche* 2 No 23.

Putnam. H. (1960), *Minds and machines, Dimensions of mind: A symposium* (Sidney Hood, editor), New York University Press 138 – 64.

Shapiro, S. (1998), "Incompleteness, mechanism, and optimism", *Bull Symbolic logic* 4 273 – 302.

Shapiro, S. (2003), "Mechanism, truth, and Penrose's new argument", *Journal of Philosophical Logic* 32 19- 42.