ANαliZA

časopis za kritično misel

Letnik 3, številka 2, december 1999 · ISSN 1408-2969

RAZPRAVE

Damjan Bojadžiev

Refleksivno širjenje refleksivnega sistema

Institut “Jožef Stefan”, Ljubljana

Citiranje: Damjan Bojadžiev, »Refleksivno širjenje refleksivnega sistema«, *Analiza: časopis za kritično misel* 3, št. 2 (1999): 39–48.

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/deed.sl.

Izdajatelj: Društvo za analitično filozofijo in filozofijo znanosti, Ljubljana.

Elektronska različica je bila leta 2026 rekonstruirana iz izvirnih stavnih datotek in preverjena s tiskano izdajo. Številke strani ustrezajo tiskani izdaji; prelomi strani znotraj prispevka se lahko nekoliko razlikujejo.

Damjan Bojadžiev
Institut "Jožef Stefan", Ljubljana

REFLEKSIVNO ŠIRJENJE REFLEKSIVNEGA SISTEMA

Formalizacija zavesti in inteligence zahteva uporabo refleksivnega formalnega sistema, ki zmore vpogled v lastno delovanje. Refleksivnost se klasično povezuje z nepopolnostjo sistema, ki se odpravlja z razširjanjem, ki pa je samo precej nerefleksivno. Bolj refleksiven postopek bi lahko upošteval učinek lastnega delovanja na način, ki je podoben izvirnemu vnosu refleksije v formalni sistem (uporaba postopka na njem samem). Toda taka refleksivna razširitev, če že ni protislovna, le v sebi izvaja korake osnovnega nerefleksivnega postopka.
Ključne besede: refleksija, Gödelov stavek, nepopolnost,razširitev, samoaplikacija.

1. Uvod

Sposobnost refleksije je pomembna sestavina inteligence in osnovna lastnost zavesti in samozavedanja. Delovanje brez refleksije je "slepo", mehanično, avtomatsko, medtem ko inteligentno delovanje predpostavlja, da dejavnik ve (se zaveda), kaj počne, da lahko zavzame neko perspektivično distanco do svojega delovanja, iz katere ga lahko presoja, ocenjuje, spremeni ali opusti. Kot pravi Hofstadter, je bistvena lastnost inteligence zmožnost, da izstopi iz naloge, ki jo opravlja, in pregleda, kaj je storila (Hofstadter, 1979: 37). Prosto po Smithu, ki je podrobno analiziral vlogo samonanašanja [*self-reference*] v sistemih reprezentacije, bi lahko rekli, da refleksija lahko vpelje inteligenco v brezglavo učinkovitost (Smith, 1986). Sposobnost (človeške) refleksije je sicer videti omejena: vpogled v lastno delovanje ima svoje meje. Hofstadter domneva, da se nam vedno znova dogajajo situacije, v katerih reagiramo na isti neumen način zato, ker nimamo dovolj pregleda nad njimi, da bi opazili njihovo enakost (Hofstadter, 1979: 614). Toda podobno lahko velja tudi za refleksivno delovanje: tako kot ostaja slepo, mehanično delovanje ujeto vase, tako se tudi refleksija lahko ujame v sebi lastne zanke. Vsekakor pa je sposobnost refleksije vsaj nujni pogoj subjektivnosti, samoza-

vedanja ali zavesti (Uršič, 1985: 1569), po nekaterih pojmovanjih pa tudi zadostni pogoj. Tako Hofstadter meni, da jaz nastane, ko pridobi moč samorefleksije (Hofstadter, 1979: 709), Perlis pa pravi, da je zavest v osnovi posebna vrsta samonanašanja (Perlis, 1997: 514).

Če je sposobnost refleksije tako pomembna za človeško inteligenco in (samo)zavedanje, potem bi morala biti nekako na delu tudi v sistemih umetne inteligence. Pri teh sistemih se kakšna sposobnost običajno programira tako, da se najprej izdela bolj ali manj formalen opis te sposobnosti, tega pa se nato prevede v izbrani programski jezik. Vključitev sposobnosti refleksije v programe umetne inteligence bi tako zahtevala uporabo refleksivnega formalnega sistema, ki zmore določen vpogled v lastno delovanje. Oznaka "računalniška [*computational*] refleksija" sicer pokriva zelo raznoliko množico načinov, kako se program lahko "zaveda" sebe in mehanizma, ki ta program izvaja, ter vpliva na ta mehanizem, vendar bo tu govora predvsem o refleksiji, ki izhaja iz klasičnih sistemov formalne logike. Ta vrsta refleksije se prekriva z nekaterimi variantami računalniške refleksije na področju avtomatskega dokazovanja izrekov (Giunchiglia in Smaill, 1988) in (meta)logičnega programiranja (Bratko, 1990).

Ko se refleksivni sistem loti refleksije svoje dejavnosti, postane ta refleksija del njegove dejavnosti, in ta (refleksivna) dejavnost lahko postane predmet nadaljnje refleksije. Takšni refleksiji, ki poleg svojega siceršnjega predmeta zajema tudi samo sebe, bi lahko rekli dvojna refleksija. Enostaven psihološki primer bi bila, na primer, ugotovitev, da mi nagnjenost k refleksiji jemlje čas in energijo, ki bi ju lahko posvetil bolj objektnim dejavnostim. Ta vrsta "celovite" refleksije se sicer omenja v literaturi o osnovah umetne inteligence – Perlis ji pravi krepko [*strong*] samonanašanje (Perlis 1997) –, vendar tam ni ravno izrazito navzoča. Podobno velja za literaturo o sistemih formalne refleksije, v katerih se refleksija obravnava v zvezi z (ne)popolnostjo sistema. V nadaljevanju bo izvajanje omejeno na to – v primerjavi z računalniško – klasično formalno obravnavo refleksije. Sistemi, ki so jih raziskovali Gödel, Turing in kasneje Feferman, so orisani v naslednjem razdelku, nato pa je v tretjem razdelku nakazana formalna opredelitev ideje o možnosti bolj celovite refleksije. Osnovna ideja je ta, da bi moralo biti razširjanje nepopolnega refleksivnega sistema tudi samo (čim bolj) refleksivno. Splošna logična forma te ideje je morda nekoliko dvomljiva – spominja, na primer, na Johnsonov izrek o postavi goniča debelih volov (kdor debele vole goni, naj bo tudi sam debel), ali na zahtevo, naj bo kritika poezije tudi sama poetična (hvala, Meži) – toda ideja sama je povsem smiselna in ustreza naravi konstrukcij, ki vzpostavijo izhodiščno nepopolnost refleksivnega sistema.

2. Refleksija in nepopolnost

V tridesetih letih se je izkazalo, da so nekateri precej elementarni formalni sistemi sposobni določenega vpogleda vase. Tako je Gödel dokazal, da formalna aritmeti-

ka ne vsebuje le običajnih resnic o naravnih številih, temveč vsebuje tudi določene resnice o njej sami. Nekatere trditve o odnosih med števili – to so običajno precej zapleteni odnosi med precej velikimi števili – namreč hkrati šifrirajo določene trditve o sistemu teh trditev samem: na primer to, da kakšna trditev je ali ni dokazljiva iz njegovih aksiomov. Sistem torej lahko "gleda" samega sebe, svojo strukturo, vendar je pri tem omejen z zahtevo po neprotislovnosti. Če je namreč v sistemu hkrati dokazljiva kakšna trditev in njena negacija, potem po klasičnih zakonih logike iz tega sledi karkoli, in sistem je neuporaben ("vse je hkrati res in ni res"). Neprotisloven sistem pa ne more, kot je dokazal Gödel, zajeti določenih resnic o sebi in svojem siceršnjem predmetu (številih). Primer takšne resnice je ravno trditev neprotislovnosti sistema: če je sistem neprotisloven, potem svoje neprotislovnosti ne more dokazati (drugi Gödelov izrek). Ker je v aritmetičnem sistemu trditev njegove lastne neprotislovnosti šifrirana v neki (precej obskurni) trditvi o številih, ta rezultat pomeni, da tudi ta trditev o številih ni dokazljiva, če je sistem neprotisloven.

Pojem nepopolnosti, ki je značilna za refleksivne sisteme, se običajno bolj povezuje s prvim Gödelovim izrekom. Ta izrek se ukvarja s pogoji dokazljivosti stavka, ki trdi ravno svojo lastno nedokazljivost (in hkrati izraža določeno obskurno aritmetično trditev, vendar te dvojnosti interpretacije v nadaljevanju ne bom več izpostavljal). Gödel je ugotovil, da stavek, ki zanika lastno dokazljivost, ni dokazljiv, če je sistem konsistenten (neprotisloven). Če pa sistem ustreza še nekoliko strožji obliki zahteve konsistence, potem Gödelov stavek ni niti ovrgljiv, *tj*., ni dokazljiva niti njegova negacija. Oboje skupaj pomeni, da dovolj konsistenten refleksiven sistem vsebuje stavek, ki ni niti dokazljiv niti ovrgljiv. Tak sistem se imenuje nepopoln [*unvollständig*, *incomplete*]. Več o tem v dodatku, tu pa bi kazalo poudariti, da nepopolnost ni preprosto posledica dejstva, da je sistem refleksiven. Refleksivnost sama po sebi namreč ni odločilna: ključno je dejstvo, da sistem operira z negacijo. Gödelov stavek pač združuje tri komponente, od katerih sta refleksivni le prvi dve: formalizacija relacije dokazljivosti v sistemu s sredstvi sistema samega, konstrukcija stavka, ki govori o sebi, in uporaba negacije. Produkt teh komponent je stavek, ki o sebi trdi, da ni dokazljiv, in *ta* stavek je razlog za nepopolnost (stavek, ki bi o sebi trdil kaj pozitivnega – na primer, da je dokazljiv – ne bi povzročil nepopolnosti).

Za nadaljnje namene tega teksta pa je pomembno predvsem to, da je v svetu naravnih števil, ki smo se ga s formalno aritmetiko namenili formalizirati, Gödelov stavek dejansko resničen (utemeljitev, zakaj je tako, bi peljala v prevelik ovinek). Če naj torej sistem formalne aritmetike zajame vse resnice o naravnih številih, potem bomo sistem, ki je bližji temu idealu, dobili tako, da izhodiščnemu sistemu, recimo mu S, dodamo njegov Gödelov stavek, recimo mu G(S). Toda če je razširjeni sistem S+G(S) konsistenten, bo tudi ta sistem nepopoln, po istem kopitu kot prej sistem S sam. V njem bo namreč mogoče konstruirati nov Gödelov stavek, ki o sebi pravi, da ni dokazljiv v razširjenem sistemu, in v njem res ni dokazljiv, če

je ta sistem konsistenten.

Zgodba o nepopolnosti pa se tu ne konča oziroma se komaj začne: tudi sistem S+G(S)+G(S+G(S)) ima svoj Gödelov stavek, in tako naprej. Pri tem seveda ne gre za kakšen logični regres, temveč za povsem dobrodušno *pro*gresijo (regres bi nastal, če bi za to, da pridemo v nek sistem, že morali biti v naslednjem). Dodajanje Gödelovih stavkov se lahko nadaljuje poljubno daleč, saj vsakič le pripravi teren za konstrukcijo novega Gödelovega stavka. Obzorje, proti kateremu beži ta postopek, pa tudi ni navadna neskončnost: mogoče je povzeti sistem, ki bi nastal po neskončno mnogo korakih postopka, v nov sistem, ki bo spet imel svojo Gödelovo peto, in tako naprej. To je neskončni "Gödelodrom", po katerem dirkata duh in stroj v znanem argumentu o kognitivni prednosti človeškega duha pred formalnim sistemom, ki ga je nazadnje popularizíral Penrose (1989, 1994).

3. Refleksivne razširitve

Razširjanje sistema z dodajanjem njegovega Gödelovega stavka je v določenem smislu že refleksivno, je nekakšna refleksija sistema. Gödelov stavek je namreč ekvivalenten trditvi, da je sistem konsistenten (Feferman, 1995), tako da se z njegovim dodajanjem sistem razširi z določeno ugotovitvijo o sebi: z ugotovitvijo lastne konsistence, ki mu sicer ni dostopna, v kolikor je konsistenten (po drugem Gödelovem izreku). Sistem si seveda dejansko ne "prilasti" ugotovitve lastne konsistence, ker ta ugotovitev ne postane del *tega* sistema, saj njeno dodajanje proizvede nov sistem ("v hipu, ko bi si sistem prilastil trditev lastne konsistence, bi se, ravno skozi to prilastitev, že spremenil v drug sistem"). Vsekakor pa je razširjanje sistema z njegovim Gödelovim stavkom refleksivno v tem smislu, da se sistemu doda stavek, ki govori o tem sistemu samem. Dosti bolj refleksiven pa bi bil postopek, ki ne bi razširil sistema s stavkom, ki govori o tem sistemu, temveč s stavkom, ki "že" govori o sistemu, ki ga to dodajanje proizvede. Če stopicanje osnovnega postopka zapišemo z enačbo

$$S' = S + G(S),$$

potem bi refleksivno varianto zapisali v obliki

$$S' = S + G(S').$$

Po tej varianti se sistemu doda stavek, ki "vračuna" učinek lastnega dodajanja k sistemu: ni le Gödelov stavek sistema S, kateremu se doda, temveč sistema S+G', ki ga proizvede njegovo lastno dodajanje sistemu S. Stavek G(S') je sicer mogoče konstruirati, kar je posledica Kleenejevega izreka o rekurziji (Rogers, 1987: 187), toda tak sistem S' bo dejansko protisloven. Stavek G(S'), ki je eden od aksiomov sistema S', je namreč ekvivalenten trditvi neprotislovnosti tega sistema, drugi Gödelov izrek pa pravi, da trditev o lastni neprotislovnosti lahko vsebuje le protisloven sistem (neprotisloven sistem ne more dokazati lastne konsistence). Tako kot ima "statična" refleksija sistema določene meje (Gödelove in druge izreke), če

naj bo sistem konsistenten, tako ima tudi "dinamična" refleksija pri razširjanju sistema določene meje, če naj ohrani konsistenco sistema.

Prevelika doza refleksije torej pelje v protislovje, vendar to še ne pomeni, da ideja refleksivnega razširjanja ni smiselna in da krepka doza refleksije ni mogoča ali zaželena. Nasprotno, osnovni postopek razširitve sistema z dodajanjem njegovega Gödelovega stavka je globoko nerefleksiven v nekem temeljnem pomenu, namreč v tem, da ostaja "slep" za dejstvo, da vsak posamezni korak njegovega izvajanja zgolj reproducira okoliščine, ki omogočajo njegovo (ponovno) izvajanje. Bolj refleksiven postopek bi se "zavedel" tega, da vsakič proizvede nov nepopoln sistem, in bi nekako "vračunal" svoj lastni učinek: ne bi deloval le nad vsakokratnim sistemom, temveč bi se na nek način v njem reflektiral, tako da bi v osnovi deloval tudi nad samim seboj. Nekoliko bolj koherentno rečeno, refleksivni postopek bi znal dodati svoj opis oziroma opis svojega učinka k sistemu, nad katerim sicer deluje, in bi nato deloval nad tem refleksivno dopolnjenim sistemom. V fazi statične refleksije sistema je ravno ta prijem ključen pri konstrukciji stavka, ki govori o samem sebi: določen postopek (diagonalizacija) se uporabi na nečem, kar vsebuje opis učinka tega postopka (Hofstadter, 1979: 446-7). Podobno deluje Webbov refleksivni Turingov stroj, ki najprej požene podprogram, v katerem simulira samega sebe (Webb, 1980: 235). Obstoj takega stroja je, tako kot obstoj aksiomov, ki trdijo lastno konsistenco, posledica Kleenejevega izreka, ki tudi temelji na večkratni uporabi omenjenega postopka (diagonalizacije). Ideja, da naj bo razširjanje refleksivnega sistema tudi samo refleksivno, je tako povsem smiselna in popolnoma ustreza naravi Gödelovih konstrukcij.

Idejo, da naj bo razširjanje refleksivnega sistema tudi samo refleksivno, je mogoče razviti v vmesnem prostoru, ki ostaja odprt med minimalno refleksivnostjo osnovnega postopka in pretirano refleksivnostjo protislovne variante. Ta prostor pokriva shema

$$S' = S + G(S^*),$$

kjer je sistem S* nekje med sistemoma S in S'. Možna vmesna varianta je, na primer,

$$S' = S + G(S + G(S)),$$

pri kateri sistem S – namesto s stavkom G(S) – razširimo s stavkom G(S+G(S)). To lahko storimo, ker tudi stavek G(S+G(S)) ni dokazljiv v S, če G(S) ni, saj G(S+G(S)) implicira G(S). Če torej dokažemo G(S+G(S)), potem sledi tudi G(S), in obratno: če G(S) ni dokazljiv, tudi G(S+G(S)) ne more biti. Stavek G(S+G(S)) pa implicira G(S) zato, ker G(S+G(S)) o sebi pravi, da ni dokazljiv v S+G(S), da torej ni dokazljiv niti v S – če nekaj ni dokazljivo v določenem sistemu, ne more biti dokazljivo niti v manjšem sistemu –, to pa o sebi pravi G(S). Sistem S je torej mogoče razširiti ne le s stavkom G(S), temveč tudi s stavkom G(S+G(S)), rezultat pa je sistem S+G(S)+G(S+G(S)), ker dodajanje G(S+G(S)) implicitno doda tudi

njegovo posledico G(S).

Na prvi pogled se celo zdi, da bi pri razširitvi s stavkom G(S+G(S)) preskočili en korak osnovnega postopka in iz sistema S neposredno prestopili v S+G(S) +G(S+G(S)). Toda ta preskok je le navidezen, saj je pri formulaciji stavka G(S+G(S)) najprej treba izračunati sistem S+G(S), z njim pa stavek G(S). Ta varianta je torej le navidezno hitrejša, ker se prehod iz sistema S v sistem S+G(S) +G(S+G(S)) izvede tako, da postopek najprej znotraj sebe simulira prehod na vmesni sistem S+G(S). Podobno bi se dalo reči za postopke, ki izvajajo večje končne ali celo neskončne skoke, vendar je za to potreben bolj izdelan pojmovni aparat (kodiranje in preslikave (neskončnih) vrstnih števil).

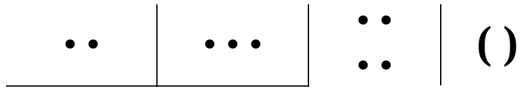
Pri nakazanih variantah refleksivne razširitve se nadalje utegne zdeti, da ne dosegajo povsem izhodiščne ideje refleksivnega razširjanja, ker ne omogočajo dejanskega preskakovanja vmesnih sistemov. Lahko bi rekli, da so ti postopki nekoliko preveč "ekstenzionalni", da bi bili resnično refleksivni: znotraj sebe izvajajo ustrezno zaporedje osnovnih korakov postopka, namesto da bi le, na osnovi opisa postopka, predvideli rezultat njegovega izvajanja – ne da bi ga dejansko izvedli – in nato presegli ta rezultat. Toda predvidevanje učinka nekega programa zahteva vsaj simulacijo njegovega izvajanja – če ne kar njegovo izvajanje samo –, tako da kakšna bolj idealna refleksija dejansko ni mogoča.

4. Sklep

Ideja, da naj bo razširjanje refleksivnega sistema tudi samo (čim bolj) refleksivno, je smiselna in izpeljiva, vendar dejansko ne prekaša enostavne ideje standardnega postopka razširitve. Zahteva neprotislovnosti omejuje refleksivnost razširitve na oblike, ki ne omogočajo hitrejšega razširjanja sistema.

Dodatek: formalni sistemi (aritmetike)

Gödel je konstruiral svoj stavek kot zadnjo formulo v zaporedju precej zapletenih aritmetičnih formul (njih 46). Beseda "formula" predvidoma že prikliče del tega, za kar pri formalnih sistemih gre, ostalo pa je, na primer, mogoče potegniti iz smisla, v katerem je šah bolj formalna igra od nogometa (pri šahu je povsem vseeno, kakšne objekte se premika, pri nogometu pa ni povsem vseeno, kakšen objekt se brca). Pri formalnem sistemu je osnovno to, da se pravila, po katerih se iz nekih izhodiščnih trditev izpeljujejo njihove posledice, lahko sklicujejo le na obliko zapisa trditev, ne pa na njihov pomen. Drugače rečeno, ves pomen, ki ga formalni sistem zajame, mora biti ekspliciran v sami obliki zapisa. Ker bi bilo pojasnilo kakšnega dejanskega sistema nekoliko preveč obsežno, vzemimo enostaven filmski zgled – formulo iz Zemeckisovega filma *Contact* (1997):



Že sama oblika zapisa – grupiranje in število pik – navaja na domnevo, da gre za neko trditev o številih 2, 3 in 4. Dejanski sistemi formalne aritmetike uporabljajo podobno sistematičen, minimalističen zapis številnikov, ki je le nekoliko manj slikoven – namesto niza

0, 1, 2, 3, 4, …

se, na primer, uporablja niz

0, 0’, 0’’, 0’’’, 0’’’’, …

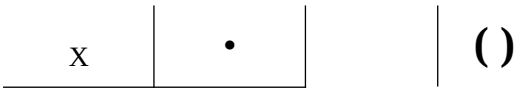
Ker so tudi ostali deli zgornje formule precej sugestivni, lahko nadalje domnevamo, da gre za zapis neenačbe

$2 + 3 \neq 4$,

točneje rečeno, za zapis metaaritmetične trditve (trditve o aritmetični trditvi)

“2 + 3 = 4” ni res.

Zgled pa je mogoče tudi razširiti z enostavnim zapisom osnovne aritmetične trditve (aksioma), ki ponazarja vlogo spremenljivk:



(nobeno število ne predhodi ničli, *tj*., katerokoli – X – število vzamemo, ni res, da bi njegov seštevek z 1 bil 0). Če uspemo na podobno elementaren način zapisati tudi ostale osnovne značilnosti naravnih števil (Peanove aksiome) in dodamo pravila sklepanja, ki se sklicujejo le na to obliko zapisa, dobimo sistem formalne aritmetike.

Prednost formalnega zapisa je ta, da se pri sklepanju, ki izhaja le iz oblike zapisa, lahko izključi možnost zmote – sklepanje postane mehansko in se načelno lahko prepusti stroju (računalniku). Da ne bo pomote: računalniki seveda že čisto dobro obvladajo računanje s konkretnimi števili, tu pa gre za dokazovanje splošnih trditev o številih – na primer izreka, da obstaja neskončno praštevil – in metateoretičnih izrekov, kakršen je Gödlov izrek (ki ga je tudi že dokazalo nekaj računalnikov oziroma programov). Nakazani postopek formalizacije pa tudi ni omejen na aritmetiko: na podoben način se sistematično opisuje tudi druge matematične ali nematematične teorije.

Ko je formalni sistem enkrat vzpostavljen, ga je mogoče naprej opazovati s povsem formalnega stališča, neodvisno od (številskega) pomena, ki smo ga imeli v mislih v fazi formalizacije. Lahko se, na primer, vprašamo, ali je sistem popoln, *tj*., ali je vsaka trditev bodisi dokazljiva bodisi ovrgljiva? Popolnost v tem smislu se včasih bolj eksplicitno imenuje *deduktivna* popolnost, ker gre za domet sklepanja (dedukcije) znotraj sistema: ali je s sklepanjem po pravilih sistema mogoče utemeljiti ali ovreči vsako trditev, formulirano v terminih sistema. Drugo vpraša-

nje pa je, kakšna je zveza med dokazljivostjo v sistemu in resničnostjo. Če je ta zveza v določenem smislu optimalna, se takemu sistemu tudi reče popoln. V primeru izhodiščne, številske interpretacije sistema je zveza med dokazljivostjo in resničnostjo optimalna, če je vse, kar je res v tej interpretaciji sistema, tudi formalno dokazljivo v njem. Popolnost te vrste – vse resnice so dokazljive – se včasih bolj eksplicitno imenuje *semantična* popolnost, ker gre za dokazljivost trditev, ki imajo določen pomen. Bolj pogosto pa se tako imenuje bolj abstraktna oblika zveze med dokazljivostjo in resničnostjo oziroma optimalne zveze med njima. Izhodiščni pomen sistema lahko namreč tudi popolnoma zanemarimo in se vprašamo, kako sama forma zapisa omejuje *možne* interpretacije sistema t.j. kako vse je mogoče interpretirati simbole sistema tako, da se ohrani resničnost osnovnih trditev (aksiomov). To je nadaljnji smisel formalnosti formalnega sistema, ko ga opazujemo le kot igro simbolov, v katero šele naknadna interpretacija vnese kakšen pomen. Na tej stopnji abstrakcije bo (semantično) popoln tak sistem, v katerem je dokazljivo vse, kar je resnično v vseh možnih interpretacijah sistema.

To idejo semantične popolnosti je mogoče razložiti še drugače. Ko namreč enkrat zanemarimo izhodiščni pomen sistema in se ukvarjamo le z *možnimi* pomeni, lahko še na drugačen način opredelimo relacijo logičnega izhajanja: A je posledica B, če je A res v vseh tistih (možnih) interpretacijah, v katerih je res tudi B. To relacijo naj bi na ravni čiste oblike ponavljala pravila sklepanja v formalnem sistemu: če je A posledica B v zgornjem smislu, potem naj bo A tudi izpeljivo iz B po formalnih pravilih sklepanja. Tako pridemo na idejo popolnega sistema v drugem, splošnem semantičnem smislu: kar je resnično v vseh interpretacijah sistema, naj bo tudi formalno dokazljivo. Drugače rečeno, kar je resnično v vseh interpretacijah, v katerih so resnični aksiomi sistema, naj bo tudi formalno izpeljivo iz njih.

Različni smisli popolnosti seveda ponujajo precej priložnosti za zmote in nesporazume, tako da je McCawley v svojem uvodu v logiko za jezikoslovce – zato da bi "zmanjšal verjetnost, da bo bralec predpostavil, da ima Gödelov izrek o nepopolnosti kaj opraviti s semantično popolnostjo" (McCawley, 1981: 74) – napisal poseben razdelek o vrstah popolnosti. Gödel je pač dokazal zgolj in samo *deduktivno* nepopolnost aritmetike (obstoj trditev, ki niso niti dokazljive niti ovrgljive). Potencialna zmeda pa je dejansko še hujša, ker se Gödel ni ustavil na tem, da *ni* dokazal semantične nepopolnosti aritmetike, temveč je – popolnoma nasprotno – dokazal njeno semantično *popolnost* (dokazljivost vsega, kar je res v vseh interpretacijah), in sicer, še preden je dokazal njeno (deduktivno) nepopolnost. Tako je nastala samo logikom razumljiva domislica, da je Gödel najprej dokazal popolnost, nato pa še nepopolnost aritmetike. Logična posledica tega srečanja (semantične) popolnosti in (deduktivne) nepopolnosti pa je naslednja: če (Gödlov) stavek ni dokazljiv, dokazljivo pa je vse, kar je res v vseh interpetacijah, potem obstaja interpretacija, v kateri stavek ni resničen (taki interpretaciji se včasih reče nestandarden model aritmetike).

damjan.bojadziev@ijs.si
http://nl.ijs.si/~damjan/me.html

Reflexive Extension of Reflexive Systems

Formalizing consciousness and intelligence requires a reflexive formal system, capable of insight into its own workings. Reflexivity is classicaly connected with incompleteness, and usually overcome in ways which are themselves not very reflexive. A more reflexive procedure might take into account its own effect, in a way similar to the initial introduction of reflexivity into the system (self-application of diagonalization). But such reflexive extension, if consistent, only performs within itself the steps of the basic non-reflexive procedure.
Keywords: reflexivity, Gödel sentence, incompleteness, extension, self-application.

Literatura

Bratko, I. (1990). *Prolog programming for Artificial Intelligence*. **KRAJ**: Addison-Wesley. Feferman, S. (1995). "Penrose's Gödelian argument". *Psyche,* 2 (7), HTML verzija na psyche.cs.monash.edu.au.

Giunchiglia, F. & Smaill, A. (1988). "Reflection in Constructive and Non-constructive Automated Reasoning". *International Workshop on Meta-Programming in Logic META*, str. 123-140.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach – An Eternal Golden Braid*. **KRAJ**: Basic Books.

McCawley, J. D. (1981). *Everything that Linguists have Always Wanted to Know about Logic (but were ashamed to ask)*. **KRAJ**: University of Chicago Press.

Penrose, R. (1989). *The Emperor's New Mind*. **KRAJ**: Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind*. KRAJ: Oxford University Press.

Perlis, D. (1997). "Consciousness as Self-Function". *Consciousness Studies,* 4 (5-6). str. 509-525.

Rogers, H. jr. (1967, 1987). *Theory of recursive functions and effective computability*. **KRAJ**: MIT Press.

Smith, B. (1986). "Varieties of Self-reference". V Halpern, J. Y. (ur.), *Theoretical Aspects of Reasoning about Knowledge*, **KRAJ**: Morgan Kaufmann.

Uršič, M. (1985). “Dialektični um ni ’naš rešitelj’“, *Nova Revija*, 43-4, str. 1569-1570.

Webb, J. C. (1980). *Mechanism, Mentalism and Metamathematics - An Essay on Finitism*. **KRAJ**: D. Reidel Publ. Co.