

Finnur Dellsén

Frá skoðunum til trúnaðar og aftur til baka

Um bayesíska þekkingarfræði

1. Inngangur

Þessi grein fjallar um svonefnda *bayesíska þekkingarfræði* (e. *Bayesian epistemology*) og tilraunir heimspekinga til að tengja hana saman við hefðbundna þekkingarfræði. Í mjög stuttu máli gengur bayesísk þekkingarfræði út á að nota líkindafræði, sem er undirgrein stærðfræðinnar, til að segja til um hvernig við eigum að taka afstöðu til ólíkra hluta. Hefðbundin þekkingarfræði notast hins vegar sjaldan við tæki og tól stærðfræðinnar og því er aðferðafræði bayesískrar þekkingarfræði afar ólík því sem við þekkjum úr hefðbundinni þekkingarfræði. Við þetta bætist að bayesísk þekkingarfræði fjallar fyrst og fremst um annars konar afstöðu en hefðbundin þekkingarfræði – um að leggja *trúnað* á eitthvað frekar en að hafa *skoðun* – og því getur stundum virst sem tengslin milli hefðbundinnar þekkingarfræði og bayesískrar þekkingarfræði séu lítil sem engin.

Markmið þessarar greinar er þó að sýna fram á að hefðbundin þekkingarfræði geti margt lært af bayesískri þekkingarfræði sem getur varpað nýju og skýrara ljósi á ýmis álitamál í þekkingarfræði og vísindaheimspeki. Að auki má nota bayesíska þekkingarfræði til að leiðrétta villur og gagnrýna ýmsar kreddur sem gengið hefur verið út frá innan hefðbundinnar þekkingarfræði. Að mínu mati er það því ekki tilviljun að bayesískri þekkingarfræði hefur vaxið fiskur um hrygg innan heimspekinnar undanfarna áratugi og að nú sé tæpast hægt að kalla sig þekkingarfræðing án þess að kunna skil á þessum fræðum. Því miður er það engu

að síður svo að næstum ekkert hefur verið skrifað um bayesíska þekkingarfræði á íslensku og er ætlunin að bæta úr því hér með.¹

Ég mun hefja leikinn í 2. hluta á því að skoða eitt lykilhugtak hefðbundinnar þekkingarfræði – *skoðun* – og samsvarandi hugtak bayesískrar þekkingarfræði – *trúnaður*.² Í 3. hluta ræði ég svo grunnatriði bayesískrar þekkingarfræði, þar á meðal þær kröfur sem frumsendur líkindafræðinnar eru taldar gera til þess trúnaðar sem við leggjum á ólíkar fullyrðingar, auk þess sem ég mun gera grein fyrir tiltekinni reglu um það hvernig skynsamir einstaklingar eigi að bregðast við þegar nýjar upplýsingar koma til sögunnar. Í 4. hluta ræði ég svo áhrifamikil rök fyrir því að fylgja skuli bayesískri þekkingarfræði að málum, hin svonefndu *veðmálsrök*. Að því loknu reyni ég svo í 5. hluta að tengja bayesíska þekkingarfræði saman við hefðbundna þekkingarfræði með því að skoða hvaða tengsl eru milli trúnaðar og skoðana. Í ljós mun koma að þessi tengsl eru afar vandmeðfarin, og mun ég reyna að bregðast stuttlega við því í 6. hluta.

2. Skoðanir og trúnaður

Hefðbundin þekkingarfræði fæst við ýmis álitamál á borð við það hvernig skuli skilgreina þekkingu, hvort hægt sé að afla sér þekkingar á tilteknum hlutum (eins og siðferðilegum sannindum eða hinum ytri heimi), og hvernig það skuli þá best gert. Reyndar eru ekki allir sammála um að þekkingarfræði eigi að leggja sérs-taka áherslu á *þekkingu* (e. *knowledge*) og færa í staðinn rök fyrir því að áhuga-verðara sé að skoða fyrirbæri eins og *skilning* (e. *understanding*) og *rökstuðning* (e. *justification*).³ Þó má segja að meðal þekkingarfræðinga sé víðtæk samstaða um að þekkingarfræði fái við *skoðanir* (e. *beliefs*). Hefðbundin þekkingarfræði fjallar um *boðandi* eða *normatívar* spurningar sem tengjast skoðunum, svo sem hvernig fólk á að mynda sér skoðanir, hvaða skoðanir geti talist til réttnefndrar þekkingar, og hvort hægt sé að rökstyðja skoðanir um tiltekna hluti eður ei. Eins undarlega og það kann að hljóma má því segja að það séu skoðanir frekar en þekking sem eru meginviðfangsefni hefðbundinnar þekkingarfræði.

Frá sjónarhorni hefðbundinnar þekkingarfræði eru skoðanir tiltölulega einfalt fyrirbæri.⁴ Með tilliti til tiltekinnar fullyrðingar er annaðhvort hægt að telja⁵ fullyrðinguna sanna eða ekki. Auk þess er auðvitað hægt að telja fullyrðinguna ósanna, en það er jafngilt því að telja neitun upphaflegu fullyrðingarinnar sanna. Það breytir því ekki að annaðhvort hefur maður tiltekna skoðun eða ekki – það

1 Mér vitanlega er eina undantekningin á þessu grein mín „Tvö viðhorf til vísindalegrar þekkingar – eða eitt?“, Finnur Dellsén 2015.

2 Orðið „trúnaður“ er hér notað í merkingunni *að leggja trúnað á eitthvað* en ekki í merkingunni *að geta trúnaðar um eitthvað*. Þessu orði er ætlað að samsvara enska orðinu „credence“. Ég vil þakka Jóni Ólafssyni fyrir að stinga upp á þessari þýðingu.

3 Sjá til dæmis Kaplan 1985 og Kvanvig 2003.

4 Vissulega er mörgum spurningum um skoðanir ósvarað – svo sem um hvert sé eðli skoðana og hvert sé inntak eða viðfang þeirra. En þessar spurningar eiga heima í hugspeki og málspeki fremur en í þekkingarfræði.

5 Til einföldunar í framhaldinu mun ég nota orðasambandið „að telja P sanna“ yfir það að hafa þá skoðun að P sé sönn.

er ekkert þar á milli. Í þessum skilningi eru þær skoðanir sem hefðbundin þekkingarfræði fjallar um *tvisætar* (e. *binary*). Við nánari umhugsun má þó ljóst vera að hér er um talsverða einföldun að ræða. Staðreyndin er sú að við höfum *missterkar* skoðanir á því hvort eitthvað sé satt eða ósatt, og því er að nokkru leyti villandi að tala um skoðanir eins og um tvisætt fyrirbæri sé að ræða. Við getum til dæmis velt því fyrir okkur hversu sterkar skoðanir við höfum á eftirfarandi fullyrðingum:

- (a) $1 + 1 = 2$.
- (b) Ég er hugsandi vera.
- (c) Bjarni Benediktsson er forsætisráðherra Íslands.
- (d) Ég læsti útidyrnum í morgun.
- (e) Peningurinn í veskinu mínu dugar fyrir hádegismatnum á eftir.
- (f) Rauðvínsdrykkja dregur úr líkum á hjartaáfalli.
- (g) Bjarni Benediktsson verður forsætisráðherra eftir næstu þingkosningar.
- (h) Það hefur verið brotist inn til mín síðan ég fór að heiman.
- (i) Sturla Jónsson verður næsti forsætisráðherra Íslands.
- (j) $1 + 1 = 3$

Hverjar af þessum fullyrðingum myndum við telja sannar? Fyrir mitt leyti væri ekki út í hött að segja að ég teldi (a)-(e) sannar. En ég er þó meira viss um að $1 + 1$ sé 2 en að ég hafi læst útidyrnum í morgun. Svipaða sögu er að segja um fullyrðingarnar neðar á þessum lista. Segjum sem svo að ég telji að (h)-(j) séu ósannar. Engu að síður er ég meira viss um að $1 + 1$ sé ekki 3 en að það hafi ekki verið brotist inn til mín síðan ég fór að heiman.

Þetta bendir til þess að það sé einskonar einföldun að flokka fullyrðingar eftir því hvort maður *telji* eða *hafi þá skoðun* að þær séu sannar eða ekki. Í raun séu skoðanir okkar mun nákvæmari en við höfum gert okkur grein fyrir hingað til. Einföldunin sem um ræðir er kannski sambærileg og þegar við segjum að eitthvað sé *heitt* eða *kalt* – við gætum verið nákvæmari með því að tilgreina hitastigið en oft dugar að tala með ónákvæmum hætti um það sem í raun er afar nákvæmt fyrirbæri. Að sama skapi telja sumir þekkingarfræðingar að tal um skoðanir sem eitthvað sem við annaðhvort höfum eða höfum ekki, sé í raun ónákvæmt tal um afar nákvæmt fyrirbæri sem kallast *trúnaður* (e. *credence*).⁶ Að leggja trúnað á eitthvað er hliðstætt því að hafa tiltekna skoðun, en þó nákvæmara að því leytinu til að trúnaðurinn getur fengið mörg ólík gildi. Þannig má segja að trúnaður okkar gagnvart sumum fullyrðingum, t.d. (a), fái hámarksgildi – 100% trúnaður – en trúnaðurinn gagnvart öðrum fullyrðingum, t.d. (j), fái lágmarksgildi – 0% trúnaður. Fullyrðingar geti svo haft ýmis önnur gildi þar á milli eins og (b)-(i) benda til. Með þessu móti eigum við að geta lýst afstöðu okkar gagnvart fullyrðingum

6 Á ensku hafa ýmis orð önnur en „credence“ verið notuð yfir fyrirbærið sem hér er þýtt sem „trúnaður“, t.d. „partial belief“, „degree of belief“ og „subjective probability“. Hið síðastnefnda er afar óheppilegt í þessu samhengi, því það er kenning en ekki skilgreiningaratriði að trúnaði skuli vera hægt að lýsa sem líkum í skilningi líkindafræðinnar (sjá 3. hluta þessarar greinar).

á borð við (a)-(j) með afar nákvæmum hætti þar sem gerður er greinarmunur á afstöðu okkar innan þeirra flokka fullyrðinga sem við sögðumst áður telja sannar eða ósannar.

Í bayesískri þekkingarfræði er sumsé gengið út frá því að afstöðu fólks til ólíkra fullyrðinga megi lýsa – að minnsta kosti að hluta – sem trúnaði sem hafi gildi á bilinu 0 (eða 0%) upp í 1 (eða 100%), að báðum meðtöldum.⁷ Eins og við munum víkja að síðar (í 5. hluta) telja sumir bayesískir þekkingarfræðingar að hægt sé að gera grein fyrir hefðbundnum (tvísætum) skoðunum sem ákveðinni tegund trúnaðar. Aðrir bayesískir þekkingarfræðingar telja hins vegar að skoðanahugtakið sé ekki lengur gagnlegt til að lýsa afstöðu fólks eftir að hið mun nákvæmara trúnaðarhugtak hefur verið skilgreint.⁸ Hvað sem þessum innbyrðis ágreiningi líður, eru allir bayesískir þekkingarfræðingar á því að þekkingarfræði eigi fyrst og fremst að snúast um að gera grein fyrir því hversu mikinn trúnað við eigum að leggja á ólíkar fullyrðingar.

3. Grunnatriði bayesískrar þekkingarfræði

Líkt og önnur þekkingarfræði er bayesísk þekkingarfræði *boðandi* eða *normatívu* grein. Hún fjallar ekki um það hvernig hlutirnir *eru í raun og veru* heldur um það hvernig þeir *ættu að vera* – ekki um það hvernig fólk í reynd tekur afstöðu til fullyrðinga, heldur um það hvernig það ætti að taka slíka afstöðu. Þetta gerir bayesísk þekkingarfræði með því að lýsa því hvernig *fullkomlega skynsamur* einstaklingur tekur afstöðu til ólíkra fullyrðinga.⁹ Meginkenning bayesískrar þekkingarfræði er að slíkur einstaklingur leggi trúnað á ólíkar fullyrðingar í samræmi við frumsendur líkindafræðinnar. Með þessum hætti tengir bayesísk þekkingarfræði *trúnað* skynsamra einstaklinga við þá undirgrein stærðfræðinnar sem nefnist *líkindafræði* (e. *probability theory*).

Áður en lengra er haldið skulum við skoða stuttlega þessar frumsendur líkindafræðinnar. Venjan er að setja fram þrjár frumsendur fyrir líkindafræði og eru þær kenndar við rússneska stærðfræðinginn Alexander Kolmogorov. Ef við látum „ $p(A)$ “ tákna líkurnar á að ótilgreind fullyrðing A sé sönn, má setja frumsendur Kolmogorovs fram svona:¹⁰

- 7 Reyndar telja margir bayesískir þekkingarfræðingar núorðið að best sé að lýsa afstöðu fólks sem *trúnaðarbili*, þ.e.a.s. einhverju bili á rauntölubilinu milli 0 og 1. Samkvæmt þessu er trúnaður fólks dreifður á milli tiltekinna talna fremur en að vera einhver slík tiltekin tala. Sjá til dæmis James M. Joyce 2005.
- 8 Sjá til dæmis Jeffrey 1970.
- 9 Það skal tekið fram að bayesísk þekkingarfræði gerir að mörgu leyti algjörlega óraunhæfar kröfur til skynsamra einstaklinga. Þetta á hún reyndar sameiginlegt með flestum kenningum í þekkingarfræði þótt kröfurnar sem bayesísk þekkingarfræði geri séu ef til vill enn óraunhæfari vegna þess að hún fjallar um mun nákvæmara fyrirbæri (þ.e.a.s. trúnað fremur en skoðanir), auk þess sem hún gerir ráð fyrir því að einstaklingar geti myndað sér afstöðu í samræmi við lögmál líkindafræðinnar.
- 10 Þessar frumsendur hafa verið settar fram á ýmsu formi. Í ljósi þess að við höfum áhuga á beitingu þeirra á þekkingarfræði, eru þær hér settar fram fyrir *fullyrðingar* (frekar en til dæmis fyrir *atburði* eins og oft er gert). Fjallað hefur verið um frumsendur Kolmogorovs í heimspæklegu samhengi í *Hvað eru vísindi?*, Erlendur Jónsson 2008.

- I. $p(A) \geq 0$, fyrir allar fullyrðingar A . Með öðrum orðum er engin fullyrðing með minna en 0% líkur.
- II. $p(A) = 1$ ef A er klifun (e. tautology). Með öðrum orðum eru líkurnar á klifunum alltaf 100%.
- III. $p(A_1 \vee \dots \vee A_n) = p(A_1) + \dots + p(A_n)$ þar sem A_1, \dots, A_n eru ósamrýmanlegar fullyrðingar.¹¹ Með öðrum orðum má leggja saman líkurnar á ósamrýmanlegum fullyrðingum og fá þannig líkurnar á að í það minnsta ein þeirra sé sönn.

Talað er um að fall $p(-)$ sem uppfyllir þessi skilyrði sé *líkindafall* (e. *probability function*). Eins og þessi framsetning gefur til kynna eru líkindaföllin sem við höfum áhuga á þess konar að þau tengja saman annars vegar fullyrðingar og hins vegar tiltekna tölu á bilinu 0 til 1 (eða 0 til 100%). Hugmyndin á bak við bayesíska þekkingarfræði er því sú að trúnaðurinn sem við leggjum á ólíkar fullyrðingar eigi að samrýmast þessum frumsendum. Með öðrum orðum á lágmarkstrúnaður ekki að vera minni en 0 samkvæmt frumsendu I, trúnaður gagnvart klifunum á að vera 1 samkvæmt frumsendu II, og samkvæmt frumsendu III á trúnaðurinn sem við leggjum á *sundurgreiningu* (e. *disjunction*) ólíkra fullyrðinga, sem útiloka hver aðra, að vera jafn þeim samanlagða trúnaði sem við leggjum á hverja fullyrðingu um sig.

Þetta skiptir máli vegna þess að líkindafræðin mun þá geta sagt okkur ýmislegt um það hversu mikinn trúnað við eigum að leggja á ólíkar fullyrðingar. Nánar tiltekið munum við geta *reiknað út* með hjálp líkindafræðinnar hvaða trúnað við eigum að leggja á ólíkar fullyrðingar ef við ætlum okkur að vera fullkomlega skynsöm. Það er best að skoða dæmi til að átta sig í því hvernig þetta virkar. Í frægri sálfræðirannsókn Amos Tversky og Daniels Kahneman (sem síðar hlaut Nóbelsverðlaunin í hagfræði) er konu lýst svo: „Linda er 31 árs gömul, einstæð, ákveðin og mjög klár. Hún lærði heimspeki í háskóla og beitti sér fyrir jafnrétti og félagslegu réttlæti, auk þess sem hún tók þátt í mótmælum gegn kjarnorkuverum.“¹² Þátttakendur í rannsókninni voru svo meðal annars spurðir hvort þeir teldu líklegra að Linda væri (a) bankastarfsmaður, eða (b) bankastarfsmaður sem væri virk í femínistahreyfingunni. Í ljós kom að yfirgnæfandi meirihluti svarenda sagði að (b) væri líklegra en (a), en það má túlka sem svo að svarendur hafi lagt meiri trúnað á (b) en (a).

Samkvæmt bayesískri þekkingarfræði voru þessir svarendur ekki fullkomlega skynsamir, því það að leggja meiri trúnað á (b) en (a) stangast á við frumsendur líkindafræðinnar. Ástæðan er sú að sanna má eftirfarandi reglu út frá þessum frumsendum:

Rökleiðslureglan. Ef A_1 leiðir af sér A_2 , þá gildir: $p(A_1) \leq p(A_2)$.

¹¹ Þessa þriðju frumsendu mætti kalla *endanlegu samlagningarfrumsenduna* (e. *axiom of finite additivity*). Stundum er þessari frumsendu breytt til að ná utan um það þegar við erum að eiga við óendanlega en teljanlega margar fullyrðingar – sú frumsenda nefnist á ensku „axiom of countable additivity“.

¹² Tversky og Kahneman 1983: 297.

Takið nú eftir því að í rannsókninni sem lýst var hér að ofan leiðir (b) af sér (a), enda er ekki hægt að vera bankastarfsmaður sem er virkur í feministahreyfingunni án þess að vera bankastarfsmaður. Bayesísk þekkingarfræði kveður því á um að ekki geti verið skynsamlegt að leggja meiri trúnað á (b) en (a). Við nánari umhugsun kemur þetta ágætlega heim og saman við það sem flestir, ef ekki allir, telja að sé rétt svar við spurningunni sem Tversky og Kahneman lögðu fyrir þátttakendurna í rannsókninni: Þótt margir svari því til að (b) sé líklegra en (a) þegar spurningin er fyrst lögð fyrir þá, átta flestir, ef ekki allir, sig á því við nánari umhugsun að þetta getur ekki verið rétt. Tversky og Kahneman nýttu sér svo þessa staðreynd til að færa rök fyrir því að rökhusun mætti skipta í tvö ólík „kerfi“ sem væru misvel til þess fallin að taka skynsamlega afstöðu til hluta.¹³

Rétt er að vekja athygli á því að sú niðurstaða að ekki geti verið skynsamlegt að leggja meiri trúnað á (b) en (a) fékkst í bayesískri þekkingarfræði með því að sanna tiltekna líkindafræðilega reglu og beita henni svo á dæmið sem um ræðir. Þessari aðferð er hægt að beita almennt í bayesískri þekkingarfræði þegar við stöndum frammi fyrir þekkingarfræðilegu álitamáli. Með þessum hætti er líkindafræði óspart notuð innan bayesískrar þekkingarfræði til að svara heimspökilegum spurningum með því að sanna viðeigandi líkindafræðireglur. Sem dæmi um álitamál sem reynt hefur verið að leysa með þessum hætti má nefna hvort og að hvaða marki treysta skuli sérfræðingum,¹⁴ hvort staðfestar forspár vísindalegra kenninga séu betri rök fyrir þeim en annars konar athuganir sem styðja kenningarnar,¹⁵ og hvort sú staðreynd að hlutur af tilteknu tagi hefur ekki fundist, leiði líkur að því að engir slíkir hlutir séu til.¹⁶

Sú líkindafræðilega regla sem oftast er beitt í heimspökilegum umfjöllunum af þessu tagi er sú sem kenningin sjálf er nefnd eftir, *regla Bayes* (e. *Bayes' theorem*).¹⁷ Til að átta okkur á þessari reglu þarf fyrst að skilja hugtakið *skilyrtar líkur* (e. *conditional probability*) og setja fram örlítið táknmál. Skilyrtar líkur eru líkurnar á því að eitthvað sé satt að því gefnu að eitthvað annað sé satt. Þetta er skrifað „ $p(A|B)$ “, sem stendur þá fyrir líkurnar á því að A sé sönn að því gefnu að B sé sönn.¹⁸ Rétt eins og röklegu tengslin milli A_1 og A_2 tengja saman líkurnar á A_1 og A_2 samkvæmt rökleiðslureglunni sem við skoðuðum hér að ofan, þá mynda skilyrtu líkurnar á A að gefnu B tengsl milli þess hvaða líkur eru á A annars vegar og B hins vegar. Um þetta fjallar regla Bayes, sem setja má fram á eftirfarandi formi:

$$\text{Regla Bayes. } p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

13 Ekki verður farið nánar út í þessa sálma hér, en áhugasömum er bent á nýlega metsölubók Daniels Kahneman, *Thinking, Fast and Slow*, Kahneman 2011.

14 Sjá Goldman 2001.

15 Sjá til dæmis Horwich 1982 og Maher 1988.

16 Sjá Sober 2009, Strevens 2009 og Finnur Dellsén 2016.

17 Reglan er kennd við 18. aldar prestinn Thomas Bayes (1701–1761). Bayes setti regluna þó fram á dálítið öðru formi og hafði engin bein áhrif á þróun bayesískrar þekkingarfræði að öðru leyti.

18 Í líkindafræði eru skilyrtu líkurnar $p(A|B)$ yfirleitt skilgreindar sem $p(A \& B)/p(B)$. Með þessa skilgreiningu að vopni er hægt leikur að sanna reglu Bayes á því formi sem hún er sett fram í textanum.

Með öðrum orðum eru skilyrtu líkurnar á A að gefnu B í réttu hlutfalli bæði við líkurnar á A og við skilyrtu líkurnar á B að gefnu A, en í öfugu hlutfalli við líkurnar á B. Oft er betra að notast við aðra útgáfu reglunnar sem er jafngild þessari:

$$\text{Regla Bayes á öðru formi. } p(A|B) = \frac{p(A)p(B|A)}{p(A)p(B|A) + p(\neg A)p(B|\neg A)}$$

Til eru fleiri útgáfur reglunnar en við látum þetta duga hér. Vert er að athuga að regla Bayes er í sjálfu sér ekki heimspekileg kenning heldur setning í líkindafræði, en þekkingarfræðilegt mikilvægi hennar kemur fram ef við gerum ráð fyrir því að skynsamir einstaklingar eigi að dreifa trúnaði sínum í samræmi við frumsendur líkindafræðinnar. Ástæðan er sú að regluna *leiðir af* þessum frumsendum (rétt eins og rökleiðsluregluna) og því þarf trúnaður skynsamra einstaklinga að virða regluna samkvæmt bayesískri þekkingarfræði.

Við þetta má bæta að í bayesískri þekkingarfræði er einnig fjallað um hvernig við eigum að breyta trúnaði okkar eða uppfæra hann í ljósi nýrra gagna sem við öflum okkur. Um þetta fjallar setning sem að sumra mati er kjarninn í bayesískri þekkingarfræði, svonefnd *bayesísk skilyrðing* (e. *Bayesian conditionalization*).¹⁹ Nánar tiltekið kveður bayesísk skilyrðing á um að trúnaður manns gagnvart tiltekinni fullyrðingu A, eftir að gagnanna G hefur verið aflað, skuli vera jafn mikill og sá trúnaður sem maður hafði áður á A að gefnu G. Í bayesískri þekkingarfræði er þetta jafngilt því að setja

$$p_e(A) = p_f(A|G)$$

þar sem $p_f(-)$ lýsir líkunum fyrir öflun gagnanna G og $p_e(-)$ lýsir líkunum eftir öflun gagnanna G. Bayesísk skilyrðing kveður í raun á um tiltekna tegund af samkvæmni yfir tíma, þ.e.a.s. að allar breytingar á skoðunum okkar eigi að taka mið af skilyrtum líkum fyrri tíma. Rétt er að vekja athygli á því að samkvæmt reglu Bayes er þetta jafngilt því að setja:

$$p_e(A) = \frac{p_f(A)p_f(A|G)}{p_f(G)}$$

Þetta er sú útgáfa af bayesískri skilyrðingu sem oftast er sett fram, enda segir hún okkur í vissum skilningi mun meira en sú fyrri þótt auðvitað séu þessar setningar jafngildar í líkindafræði.

Best er að átta sig á hvað þetta þýðir allt saman með því að skoða dæmi. Kári, sem er sjö ára, bauð mér nýlega að spila við sig spil sem ég kannaðist ekki við. Kári sagðist kunna reglurnar og ég treysti honum þess vegna fyrir því að segja til um hver hefði unnið hverja umferð í spilin. Eftir 10 umferðir hafði Kári hins vegar unnið hverja einustu umferð, og ég fór að gruna Kára um græsku. Ég vildi

19 Rétt er að taka fram að margir bayesískir þekkingarfræðingar hafa sett fram efasemdir um bayesíska skilyrðingu og vilja annaðhvort hafna henni með öllu eða setja fram hófsamari útgáfu af henni. Sjá til dæmis Christensen 1991 og 1996.

samt ekki saka Kára um að hafa svindlað nema ég hefði fyrir því haldgóð rök og því spurði ég sjálfan mig hversu mikinn trúnað ég ætti að leggja á að Kári hefði svindlað þegar fyrir lá að hann hafði unnið 10 sinnum í röð. Látum „S“ tákna þá fullyrðingu að Kári hafi svindlað og „U“ tákna þá fullyrðingu að Kári hafi unnið 10 sinnum í röð. Samkvæmt bayesískri skilyrðingu og reglu Bayes gildir þá að trúnaðurinn sem ég átti að leggja á S eftir að Kári hafði unnið 10 sinnum í röð er:

$$p_e(S) = \frac{p_f(S)p_f(U|S)}{p_f(S)p_f(U|S) + p_f(-S)p_f(U|-S)}$$

Við getum nú fyllt inn í það sem er á hægri hlið jöfnunnar með því að gefa okkur tölur fyrir þessar líkur. Líkurnar á að Kári svindli tel ég almennt mjög litlar (enda treysti ég Kára) og því set ég $p_e(S) = 1\%$. (Um leið set ég $p_e(-S) = 99\%$, því ég veit að það eru samanlagt 100% líkur á að S sé sönn eða að S sé ósönn.) Líkurnar á að Kári vinni 10 sinnum röð, að því gefnu að hann hafi svindlað, tel ég afar miklar og set því $p_f(U|S) = 90\%$. Á hinn bóginn tel ég afar ólíklegt að vinna 10 sinnum í röð ef maður er ekki að svindla, enda met ég það svo að líkurnar á að vinna hverja umferð sé 50% og set því $p_f(U|-S) = 0,5^{10} \approx 0,1\%$. Ef við stingum þessum tölum svo inn í jöfnuna hér að ofan fáum við $p_e(S) = 90,3\%$. Samkvæmt þessu ætti ég því að leggja mjög mikinn trúnað á að Kári litli hafi svindlað.

4. Hvað er svo gott við að vera bayesískur?

Við höfum nú séð hvað felst í bayesískri þekkingarfræði og hvernig kenningin myndar nokkurs konar reiknistokk fyrir það hvaða trúnað við eigum að leggja á ólíkar fullyrðingar á ólíkum tímum. En hver eru rökinn fyrir þessari kenningu? Ein rök, sem sjaldan eru útlustuð frekar, eru þau að kenningin virðist samrýmast afar vel einstökum dæmum (eins og dæminu hér að ofan) um hvaða trúnað við eigum að leggja á ólíkar fullyrðingar. Það virðist einfaldlega vera rétt að ég hafi sterk rök fyrir því að saka Kára litla um að svindla, og bayesísk þekkingarfræði gerir vel grein fyrir þessu. Önnur tegund af rökum, sem reyndar eru nokkuð umdeild, eru kennd við hollensk veðmál en ég mun vísa til þeirra sem *veðmálsrakanna* (e. *Dutch Book Arguments*). Þessi rök byggjast á ákveðnum tengslum sem virðast vera til staðar á milli þess að leggja tiltekinn trúnað á fullyrðingu annars vegar og svo þess að vera tilbúinn að veðja einhverju á að fullyrðingin sé sönn. Við skulum líta aðeins nánar á þessi rök.

Til eru ýmsar útgáfur af veðmálsrökunum en þau hefjast jafnan á því að gert er ráð fyrir eftirfarandi forsendu um tengsl trúnaðar og sanngjarnra veðmála:

T-V: Ef einstaklingur hefur tiltekinn trúnað d á fullyrðingu A þá er sanngjarnt frá hans bæjardyrum séð að borga d kr. fyrir veðmál sem gæfi 1 kr. ef A væri sönn, en 0 kr. ef A væri ósönn.

Gerum til dæmis ráð fyrir því að ég sé algjörlega óákveðinn um hvort það fari að

rigna á eftir – það er að segja, ég legg jafn mikinn trúnað á að svo verði eins og að svo verði ekki. Samkvæmt T-V væri sanngjarnt frá mínum bæjardryrum séð að borga hálfá krónu fyrir veðmál sem gæfi eina krónu ef það færi að rigna. Ef ég væri hins vegar 90% viss um að það myndi rigna, ætti ég að vera tilbúinn að borga 0,9 kr. Og svo framvegis.²⁰

Veðmálsrökin notfæra sér þessi tengsl milli trúnaðar og veðmála til að rökstyðja að trúnaður verði að samrýmast frumsendum líkindafræðinnar hjá skynsömum einstaklingum. Við getum sýnt fram á þetta fyrir hverja frumsendu um sig en við skulum nú skoða frumsendu III sérstaklega. Ímyndum okkur til dæmis að ég leggi 30% trúnað á að það muni rigna í Reykjavík klukkan átta í fyrramálið (R) og 20% trúnað á að það muni snjóa á sama stað á sama tíma (S).²¹ Gerum jafnframt ráð fyrir því að ég brjóti gegn frumsendu III með því að leggja 40% en ekki 50% trúnað á að það muni annaðhvort rigna eða snjóa (R ∨ S). Hollensku veðmálsrökin byggjast á því að benda á að nú sé hægt að bjóða mér eftirfarandi veðmál sem ættu öll að teljast sanngjörn frá mínum bæjardryrum séð:

Veðmál 1: Borgaðu 3 kr. og fáðu 10 kr. til baka ef það rignir en 0 kr. ef það rignir ekki.

Veðmál 2: Borgaðu 2 kr. og fáðu 10 kr. til baka ef það snjóar en 0 kr. ef það snjóar ekki.

Veðmál 3: Fáðu 4 kr. með þeim skilyrðum að þú þarft að borga 10 kr. til baka ef það rignir eða snjóar en 0 kr. ef hvorugt gerist.²²

Vandinn er sá að ef ég tek öllum þessum veðmálum, sem öll eru sanngjörn frá mínum bæjardryrum séð samkvæmt T-V, þá mun ég tapa, sama hvað gerist. Ástæðan er í raun einföld. Eftir að hafa tekið veðmálunum (en áður en það liggur fyrir hver niðurstaða þeirra verði) hef ég borgað samtals 5 kr. en fengið 4 kr. til baka. Ég er því komin í 1 kr. tap strax í upphafi. Nú eru þrír möguleikar til staðar eins og eftirfarandi tafla sýnir:

Rignir?	Snjóar?	Hvað fæ ég til baka úr veðmálunum?
Já	Nei	Græði 10 kr. á veðmáli 1; tapa 10 kr. á veðmáli 3.
Nei	Já	Græði 10 kr. á veðmáli 2; tapa 10 kr. á veðmáli 3.
Nei	Nei	Fæ ekkert út úr neinu veðmáli.

Tafla 1: Gróði og tap af ofangreindum veðmálum eftir því hvort það rignir eða snjóar á morgun.

20 Í árdaga bayesiskrar þekkingarfræði var því stundum haldið fram að unnt væri að smætta trúnað niður í það að vera tilbúinn að taka veðmálum af ákveðnu tagi. T-V væri þá ekki einungis sönn, heldur beinlínis skilgreiningin á því hvað trúnaður er. Sjá til dæmis de Finetti 1974-5. Fáir bayesískir þekkingarfræðingar eru þessarar skoðunar nú á dögum.

21 Hér er gert ráð fyrir því að „rigning“ og „snjókoma“ séu skilgreindar þannig að það geti ekki rignt og snjóað samtímis.

22 Athugið að í veðmálum 1 og 2 er fórnarlambið í hlutverki þess sem kaupir veðmálið en í veðmáli 3 er fórnarlambið að selja veðmál.

Af þessu sést að óháð því hvað gerist mun ég fá nákvæmlega 0 kr. samtals til baka úr þessum þremur veðmálum. Þannig að þegar á heildina er litið er ljóst að *óháð því hvað gerist mun ég tapa einni krónu*.²³ Hér er að sjálfsögðu bara um eitt dæmi að ræða en rökin má setja fram með fullkomlega almennum hætti: Öll frávík frá frumsendu III leiða til sams konar niðurstöðu. Og raunar gildir þetta um hinar frumsendurnar tvær líka. Af þessum sökum virðist óskynsamlegt að leggja trúnað á staðhæfingar með öðrum hætti en þeim sem frumsendur líkindafræðinnar segja til um.

Takið eftir því að rökin sem hér hafa verið rakin eiga að sýna fram á að trúnaður skynsamra einstaklinga lúti á *hverjum tíma* lögmálum líkindafræðinnar. Rétt er að nefna að sams konar rök hafa verið sett fram til að rökstyðja að skynsamir einstaklingar eigi að *breyta* trúnaði sínum í samræmi við bayesíska skilyrðingu. Hugmyndin á bak við þessi rök eru þau að þeim sem fylgir einhverri annarri reglu við að breyta trúnaði frá einum tíma til annars sé hægt að bjóða veðmál á ólíkum tímum sem öll eru sanngjörn frá hans bæjardyrum séð en sem tryggja að viðkomandi mun tapa peningum þegar á heildina er litið, sama hvað gerist í millitíðinni. Rétt er að taka fram að þessi rök eiga ekki að sýna að skynsamir einstaklingar þurfi að nota bayesíska skilyrðingu, heldur að bayesísk skilyrðing sé eina reglan sem skynsamir einstaklingar geti notað til að uppfæra trúnað. Með öðrum orðum þá útiloka rökin ekki að skynsamir einstaklingar geti skipt um skoðanir án þess að fylgja nokkurri einustu reglu.²⁴

5. Happdrættis- og formálaþversagnirnar

Eins og áður segir gengur bayesísk þekkingarfræði út á að segja til um hvenær skynsamlegt er að hafa tiltekinn trúnað. Hefðbundin þekkingarfræði fjallar hins vegar um hvaða *skoðanir* sé skynsamlegt eða réttlætanlegt að hafa. Tal um trúnað annars vegar og skoðanir hins vegar virðist samt að mörgu leyti varða svipaða hluti, það er að segja einhvers konar *afstöðu* til fullyrðinga. Þetta bendir til þess að skoðanir og trúnaður séu alls ekki óskyldir hlutir og að segja megi eitthvað vitrænt um þekkingarfræðilegu tengslin á milli skoðana og trúnaðar. Ein hugmynd um þessi tengsl sem notið hefur talsverðs fylgis á rætur að rekja til enska nýaldarheimspekingsins Johns Locke:

Regla Lockes: Skynsamlegt er að telja tiltekna fullyrðingu sanna (þ.e. hafa þá skoðun að hún sé sönn) þá og því aðeins að skynsamlegt sé að leggja nægilega mikinn trúnað á fullyrðinguna.²⁵

Til dæmis væri væntanlega skynsamlegt að hafa þá skoðun að Kári hafi svindlað í

23 Spurning: Af hverju er engin lína í töflunni sem sýnir möguleikann á rigningu og snjókomu samtímis? Svar: Af því að við gerum ráð fyrir að R og S útiloki hvort annað (sjá síðustu neðanmálgrein), enda fjallar frumsetning III aðeins um slík tilvik.

24 Sjá einkum van Fraassen 1989.

25 Sjá Foley 1992.

spilinu sem ég nefndi hér að ofan í ljósi þess að það virðist skynsamlegt að leggja yfir 90% trúnað á það. En þótt regla Lockes virðist afskaplega trúleg við fyrstu sýn, eru veruleg vandamál tengd þessari hugmynd.

Eitt álitamál við reglu Lockes felst í að segja til um hvað sé átt við með að skynsamlegt sé að leggja „nægilega“ mikinn trúnað á einhverja fullyrðingu. Er nóg að skynsamlegt sé að hafa 70% trúnað á að það rigni úti til að sú skoðun að það sé rigning sé einnig skynsamleg? En 80%? Hvað með 90%? Hér er í raun tvenns konar vandi á ferð. Annars vegar virðast mörkin vera *ónákvæm*, svipað og mörkin milli þess að einhver fjöldi sandkorna myndi hrúgu eða ekki. Í besta falli virðumst við geta sagt að mörkin liggja á einhverju tilteknu bili, t.d. milli 85% og 90%. Hinn vandinn snýst um að það virðist ekki vera nein leið til að segja til um hver mörkin eiga að vera: Eiga þau að dreifast á milli 85% og 90%, eða milli 86% og 91% (eða eitthvað allt annað)? Vandinn hér snýst um að rökstyðja að ein mörk séu „réttari“ en einhver önnur.

Við skulum hins vegar ekki eyða of miklu púðri í þessar vangaveltur því að meginrökin gegn reglu Lockes eru af allt öðrum toga. Þessum rökum má skipta í tvennt og tengja hvorn hluta um sig við *happdrættisþverstæðuna* (e. *lottery paradox*) annars vegar og *formálaþverstæðuna* (e. *preface paradox*) hins vegar.²⁶ Áður en við skoðum þessar þverstæður skulum við taka eftir því að þótt það sé ekki sagt berum orðum í reglu Lockes, er nánast alltaf gert ráð fyrir því að sá trúnaður sem skynsamlegt er að leggja á skoðun til að skynsamlegt sé að telja samsvarandi skoðun sanna sé minni en 100%. Ástæðan er sú að annars væru nánast engar skoðanir skynsamlegar, enda ættum við ekki að leggja 100% trúnað á neitt nema hugsanlega það sem við getum komist að *a priori*, svo sem rökþæfingar.²⁷

Byrjum á að skoða rökin sem tengjast happdrættisþverstæðunni. Ímyndum okkur happdrætti með 1000 miðum en aðeins einum vinningsmiða. Segjum sem svo að við höfum afar áreiðanlegar heimildir fyrir því að þetta happdrætti sé eðlilegt að öllu leyti. Nánar tiltekið skulum við gefa okkur að það sé skynsamlegt að leggja 99,9% trúnað á að hver miði um sig sé tapmiði. En gerum jafnframt ráð fyrir því að það sé skynsamlegt að leggja 99,9% trúnað á að einhver einn af þessum miðum sé vinningsmiði, þótt við vitum auðvitað ekki hvaða miði það er. Samkvæmt reglu Lockes væri þá skynsamlegt að telja eftirfarandi fullyrðingar allar sannar:

26 Henry Kyburg setti fyrstur fram happdrættisþverstæðuna en David Makinson setti fram formálaþverstæðuna. Sjá Kyburg 1961 og Makinson 1965.

27 Margir sem fjallað hafa um þverstæðurnar taka fram að sá trúnaður sem þurfi að vera skynsamlegur til að samsvarandi skoðun sé skynsamleg þurfi að vera að minnsta kosti 50%. Ástæðan sem nefnd er þessu til stuðnings er að annars væri skynsamlegt að leggja meiri trúnað á neitun viðkomandi fullyrðingar en fullyrðinguna sjálfa og það virðist stangast á við að skynsamlegt sé að telja fullyrðinguna sanna. Ólíkt því sem stundum er gefið í skyn er hins vegar ekki strangt til tekið nauðsynlegt að gera ráð fyrir þessari túlkun á reglu Lockes til að færa fram þau rök sem hér verða til umræðu.

m_1 : Miði nr. 1 er tapmiði.

m_2 : Miði nr. 2 er tapmiði.

...

m_{1000} : Miði nr. 1000 er tapmiði.

v : Einn af þessum miðum er vinningsmiði (ekki tapmiði).

En takið nú eftir því að síðasta fullyrðingin er í raun jafngild því að segja að ein af fullyrðingunum þar fyrir ofan sé ósönn. Á táknmáli rökfræðinnar myndum við segja að $v \equiv \neg m_1 \vee \neg m_2 \vee \dots \vee \neg m_{1000}$. Þessar fullyrðingar eru því *ósamrýmanlegar*. Regla Lockes leiðir sem sagt til þess að réttlætanlegt geti verið að telja að ósamrýmanlegar fullyrðingar séu allar sannar.

Rökin sem sett eru fram á þessum grunni kveða á um að regla Lockes geti ekki verið rétt vegna þess að hún hafi þessa fjarstæðukenndu afleiðingu. Setja má rökin fram með almennum hætti svona (þar sem m er einhver jákvæð tala sem er lægri en 100):

H1. Samkvæmt reglu Lockes er skynsamlegt að hafa þá skoðun að A sé sönn ef skynsamlegt er að leggja meiri en $m\%$ trúnað á A .

H2. Til er safn ósamrýmanlegra fullyrðinga $\{A_1, \dots, A_n\}$ sem er þannig að skynsamlegt er að leggja meiri en $m\%$ trúnað á hverja fullyrðingu um sig.

H3. Það getur ekki verið skynsamlegt að hafa ósamrýmanlegar skoðanir.

H4. Regla Lockes er röng.

Hvernig má bregðast við þessum rökum? H1 leiðir af því hvernig regla Lockes er skilgreind og henni verður því ekki haggð. Rökin fyrir H2 eru sömuleiðis traust: happdrættisdæmið sem við tókum (og reyndar mörg önnur dæmi að auki) sýnir fram á að hún sé sönn með óyggjandi hætti. Þeir sem vilja halda í reglu Lockes þurfa því að hafna H3.

Áður en við skoðum þessi viðbrögð betur, skulum við snúa okkur að formálaþverstæðunni. Ímyndum okkur að þú sért nýbúinn að skrifa fræðibók, til dæmis um þorskastríðin, sem inniheldur 1000 óskyldar staðhæfingar. Þú ert ábyrgur fræðimaður og hefur því góð og haldbær rök fyrir hverri fullyrðingu, þótt þú viðurkennir að vísu að þú getir *hugsanlega* haft rangt fyrir þér um hverja og eina þeirra. Segjum sem svo að þetta þýði að það sé skynsamlegt af þér að leggja 99,9% trúnað á hverja fullyrðingu um sig. Þar með telst það skynsamlegt samkvæmt reglu Lockes að hafa þá skoðun að hver og ein fullyrðing sé sönn. En svo áttar þú þig á því að samkvæmt hefðbundinni afleiðslurökfræði leiðir af því að hver og ein af þessum fullyrðingum sé sönn að þær séu *allar* sannar. Engu að síður er *ekki skynsamlegt* að leggja mikinn trúnað á að svo sé (enda afar ólíklegt að þú hafir ekki farið með mál einhvers staðar í bókinni, ef til vill fyrir algjöra slysnis eða óheppni). Raunar gildir, samkvæmt þeirri meginkeningu bayesískrar þekkingarfræði að trúnaður eigi að vera í samræmi við lögmál líkindafræðinnar, að þú eigir

að leggja um 36,8% trúnað á þetta.²⁸ Ef við gefum okkur að þetta sé undir þeim mörkum sem regla Lockes kveður á um að sé „nægilega“ mikill trúnaður þá leiðir af þessu að ekki sé skynsamlegt að mynda sér skoðanir með því að leiða þær út frá öðrum skynsamlegum skoðunum samkvæmt reglum rökfræðinnar.

Líkt og áður kveða rökin sem sett eru fram á þessum grunni á um að regla Lockes geti ekki verið rétt vegna þess að hún hafi fjarstæðukennda afleiðingu. Setja má rökin fram með almennum hætti svona:

- F₁. Samkvæmt reglu Lockes er aðeins skynsamlegt að hafa þá skoðun að A sé sönn ef skynsamlegt er að leggja meiri en m% trúnað á A.
- F₂. Til er safn fullyrðinga $\{A_1, \dots, A_n\}$ sem er þannig að skynsamlegt er að leggja meiri en m% trúnað á hverja fullyrðingu um sig en minni en m% trúnað á fullyrðinguna $(A_1 \& \dots \& A_n)$, sem er rökleg afleiðing þess.
- F₃. Ef það er er skynsamlegt að telja A_1, A_2, \dots og A_n sannar þá er einnig skynsamlegt að telja að allar röklegar afleiðingar þeirra séu sannar, þar á meðal $(A_1 \& \dots \& A_n)$.

F₄. Regla Lockes er röng.

Um möguleg viðbrögð við þessum rökum gildir það sama og um rökin sem byggjast á happdrættisþverstæðunni. Fyrsta forsendan, F₁, er sönn samkvæmt skilgreiningunni á reglu Lockes. F₂ leiðir svo af formáladæminu sjálfu. Eftir stendur F₃, sem kveður í raun á um að skoðanir eigi að samrýmast þeirri reglu rökfræðinnar sem nefnist „og-innleiðing“.

Það er ljóst af ofangreindu að þeir sem ætla sér að halda í reglu Lockes þurfa að hafna bæði H₃ og F₃. Það fæli í sér að skynsamlegar skoðanir gætu stangast á við þær kröfur sem hefðbundin afleiðslurökfræði virðist gera til þess hvaða skoðanir sé skynsamlegt að hafa á hverjum tíma. Nánar tiltekið þýðir þetta að ekki væri hægt að gera eftirfarandi kröfu:

Afleiðslukrafa fyrir skoðanir: Mengi þeirra fullyrðinga sem skynsamur einstaklingur er reiðubúinn að telja sannar skal vera sjálfu sér samkvæmt og lokað undir rökfræðilega afleiðingu.²⁹

Þessi krafa segir í raun ekkert annað en að hefðbundin afleiðslurökfræði setji skynsamlegum skoðunum skorður. Ef við höfnum þessu þurfum við að fara að velta því fyrir okkur hvort hefðbundin afleiðslurökfræði hafi nokkuð fram að

28 Vegna þess að fullyrðingarnar eru óskyldar má reikna líkurnar á því að þær séu allar sannar með því að margfalda saman líkurnar á hverri og einni. Líkurnar eru því $0,999^{1000} \approx 0,368 = 36,8\%$.

29 Sagt er að mengi sé lokað undir tiltekna aðgerð ef það að beita aðgerðinni á stök mengisins gefur annað stak í menginu. Til dæmis er mengi náttúrulegra talna lokað undir samlagningu vegna þess að það að leggja saman tvær eða fleiri náttúrulegar tölur gefur alltaf aðra náttúrulega tölu. Á hinn bóginn er mengi náttúrulegra talna ekki lokað undir deilingu eins og til dæmis sjá má af því að 1 og 2 eru náttúrulegar tölur en $1/2 = 0,5$ er það ekki.

færa í þekkingarfræðilegu tilliti.³⁰ Í því ljósi gæti virst vænlegra að nálgast happdrættis- og formálafverstæðurnar með því einfaldlega að hafna reglu Lockes. Þetta myndi að vísu þýða að tengslin milli skoðana og trúnaðar væru torræðari en þau virðast vera við fyrstu sýn, en á móti kemur að við þyrftum engar áhyggjur að hafa af þverstæðunum tveimur.

6. Skoðanir, trúnaður og samþykki

Í síðasta hluta voru sett fram rök sem áttu að sýna að regla Lockes leiddi til þess að hafna þyrfti þekkingarfræðilegu hlutverki hefðbundinnar afleiðslurökfræði. Í þessum hluta mun ég hins vegar freista þess að sýna að við þurfum ekki að ganga að þessum afarkostum. Nánar tiltekið mun ég færa rök fyrir því að hægt sé að halda í reglu Lockes fyrir skoðanir og trúnað en viðurkenna um leið þekkingarfræðilegt mikilvægi afleiðslurökfræði. Að vísu munum við þurfa að hafna afleiðslurökfræðinni fyrir skoðanir, en á móti kemur að sams konar afleiðslukrafa reynist gilda um annars konar afstöðu til fullyrðinga.

Í þekkingarfræði er hugtökunum „skoðun“ og „trúnaður“ báðum ætlað að lýsa hugarástandi sem beinist með einum eða öðrum hætti að fullyrðingum. Í báðum tilvikum má segja að hugarástandið felist í því að *finnast fullyrðingin vera sönn*.³¹ Það gleymist þó stundum að til er annars konar hugarástand sem skoða má út frá þekkingarfræðilegu sjónarhorni. Eitt slíkt hugarástand er það sem breski heimspekingurinn L. Jonathan Cohen kallar *samþykki* (e. *acceptance*).³² Samkvæmt Cohen samþykkir maður fullyrðingu þá og því aðeins að maður taki því sem gefnu í tilteknu samhengi að fullyrðingin sé sönn. Eins og Cohen bendir á þýðir þetta að mögulegt er að samþykkja fullyrðingar sem maður telur ekki að séu sannar og sem maður leggur lítinn sem engan trúnað á. Til dæmis er algengt að lögfræðingar samþykki að skjólstæðingar þeirra séu saklausir – með öðrum orðum taka þeir því sem gefnu í lögmannsstörfum sínum að skjólstæðingarnir séu saklausir. Það geta lögfræðingarnir gert jafnvel þótt þeir séu ekki endilega þeirra skoðunar að skjólstæðingarnir séu saklausir, og þótt þeir leggi lítinn sem engan trúnað á að svo sé.

Með þennan greinarmun á skoðunum/trúnaði og samþykki að vopni skulum við nú snúa okkur aftur að þekkingarfræðilegu gildi afleiðslurökfræði. Ljóst er að við erum oft á tíðum í þannig aðstæðum að ætlast er til þess að þær fullyrðingar sem við samþykkjum séu í samræmi við hefðbundna afleiðslurökfræði. Dæmin eru mýmörg, en í stað þess að sækja vatnið yfir bæjarlækin skulum við taka dæmi af fræðilegri heimspekingrein líkt og þessari. Við myndum flest ef ekki öll líta á

30 Sumir heimspekingar sem hafa fjallað um þetta virðast setta sig við þessa niðurstöðu. Sjá til dæmis Kolodny 2007.

31 Þetta þarf ljóslega að útfæra með ólíkum hætti fyrir skoðun annars vegar og trúnað hins vegar. Til dæmis mætti segja að það að hafa þá skoðun að A sé sönn felist í því að hafa einhvers konar tilfinningu fyrir því að A sé sönn, en að það að leggja x% trúnað á A felist í því að hafa x% tilfinningu af sama tagi fyrir því að A sé sönn.

32 Cohen 1992. Fleiri hafa sett fram hugmyndir um að eitthvað sem megi kalla samþykki hafi mikið þekkingarfræðilegt gildi, þar á meðal tveir bayesískir þekkingarfræðingar. Sjá Maher 1993 og Kaplan 1996.

það sem alvarlegan galla á slíkri grein ef finna mætti tvær eða fleiri fullyrðingar í greininni sem stönguðust á. Að sama skapi myndum við telja það verulegan galla á greininni ef leiða mætti mjög ótrúverðuga fullyrðingu af því sem sagt er í greininni. Í þessu samhengi gerum við því kröfu um að höfundurinn samþykki einungis fullyrðingar sem samræmast kröfum hefðbundinnar afleiðslurökfræði.

Ef þetta er rétt má álykta sem svo að gera megi eins konar afleiðslukröfu til samþykkis jafnvel þótt þessi krafa eigi ekki við um skoðanir. Þessa kröfu væri hægt að orða svona fyrir tiltekið samhengi S:³³

Afleiðslukrafa fyrir S-samþykki: Mengi þeirra fullyrðinga sem skynsamur einstaklingur er reiðubúinn að samþykkja í tilteknu samhengi S skal vera sjálfu sér samkvæmt og lokað undir rökfræðilega afleiðingu.

Þetta skiptir máli fyrir reglu Lockes og þar með fyrir bayesíska þekkingarfræði vegna þess að þetta þýðir að það að hafna afleiðslukröfunni fyrir skoðanir felur ekki í sér að hefðbundin afleiðslurökfræði hafi engu þekkingarfræðilegu hlutverki að gegna. Það er að vísu rétt að ef við föllumst á reglu Lockes hefur afleiðslurökfræði lítið sem ekkert að segja um það hvaða skoðanir við eigum að hafa, en á móti kemur að til er annars konar afstaða – samþykki – sem lýtur lögmálum hefðbundinnar afleiðslurökfræði í vissu samhengi. Í ljósi þess að skoðun og samþykki eru náskyld fyrirbæri (hvort tveggja er dæmi um afstöðu gagnvart fullyrðingum) sem oft fara saman (við samþykkjum yfirleitt það sem við teljum satt, og öfugt) væri ekki jafn fráleitt og virst gæti í fyrstu að bregðast við happdrættis- og formálapverstæðunum með því einfaldlega að hafna bæði H₃ og F₃, og þar með einnig afleiðslukröfunni fyrir skoðanir.³⁴

Niðurstaða mín af þessum útdrúðum um samþykki er því sú að hægt sé að bjarga reglu Lockes frá happdrættis- og formálapverstæðunum án þess að gefa upp á bátinn þá hugmynd að hefðbundin afleiðslurökfræði hafi þekkingarfræðilegu hlutverki að gegna. Hugmyndin er í stuttu máli sú að hlutverk hefðbundinnar afleiðslurökfræði líti ekki að skoðunum heldur að samþykki. Á hinn bóginn gildir regla Lockes eingöngu um skoðanir, enda er ljóst af ofangreindu að það hvort skynsamlegt sé að taka einhverju sem gefnu í tilteknu samhengi er óháð því hvort skynsamlegt sé að leggja tiltekinn trúnað á það. Því má segja að greinarmunurinn á samþykki annars vegar, og skoðunum og trúnaði hins vegar, bendi eindregið til þess að togstreitan sem virðist vera til staðar á milli reglu Lockes og þekkingarfræðilegs gildis hefðbundinnar afleiðslurökfræði byggist í raun á því að rugla saman tvönn konar afstöðu sem hægt er að taka til fullyrðinga.

33 Þessi krafa er hér sett fram fyrir ótilgreint samhengi S, til dæmis samhengi af því tagi sem tekið var dæmi um hér að ofan. Ég tel nokkuð ljóst að afleiðslukröfur af þessu tagi gilda ekki fyrir hvaða samhengi sem er. Til dæmis gerum við nokkuð augljóslega ekki þá kröfu til skáldsagna að þær fylgi afleiðslurökfræði með þessum hætti.

34 Ég hef fjallað nánar um hvernig nota má greinarmuninn á samþykki og skoðunum til að leysa formálapverstæðuna í nýbirtri grein, „Deductive Cogency, Understanding, and Acceptance“, Finnur Dellsén 2017.

7. Lokaorð

Hér hafa verið rakin helstu grunnatriði bayesískrar þekkingarfræði, svo sem hugmyndin um að fást við trúnað í stað skoðana, tengsl líkindafræði og þess trúnaðar sem fullkomlega skynsamur einstaklingur leggur á ólíkar fullyrðingar, hugmyndin um bayesíska skilyrðingu, og það hvernig kenningar bayesískrar þekkingarfræði eru rökstuddar með hjálp svokallaðra veðmálsraka. Ljóst er að bayesísk þekkingarfræði er ólík hefðbundinni þekkingarfræði að mörgu leyti, sérstaklega með tilliti til þess að bayesísk þekkingarfræði fjallar um trúnað (sem getur tekið óendanleg mörg gildi) en hefðbundin þekkingarfræði fjallar um skoðanir (sem aðeins geta tekið tvö gildi). Til að freista þess að tengja bayesíska þekkingarfræði við hefðbundna þekkingarfræði skoðuðum við svo áhrifamikla hugmynd sem tengir þetta tvennt saman – reglu Lockes. Í ljós kom að regla Lockes er vandkvæðum bundin vegna tveggja þverstæðna sem virðast sýna að reglan gangi gegn hefðbundinni afleiðslurökfræði. Í síðasta hlutanum reifaði ég hins vegar hugmynd sem myndi að miklu eða öllu leyti mæta þessum vanda með því að leggja til að þekkingarfræðilegt gildi afleiðslurökfræði lúti að samþykki en ekki skoðunum.³⁵

Heimildir

- Christensen, David. 1991. Clever Bookies and Coherent Beliefs. *The Philosophical Review* 100, 229–247.
- Christensen, David. 1996. Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers. *The Journal of Philosophy* 93, 450–479.
- Cohen, L. Jonathan. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.
- de Finetti, Bruno. 1974–5. *Theory of Probability*. New York: Wiley.
- Erlendur Jónsson. 2008. *Hvað eru vísindi?* Reykjavík: Háskólaútgáfan.
- Finnur Dellsén. 2015. Tvö viðhorf til vísindalegrar þekkingar – eða eitt? *Ritid* 15, 135–155.
- Finnur Dellsén. 2016. Realism and the Absence of Rivals. *Synthese*, DOI: 10.1007/s11229-016-1059-3.
- Finnur Dellsén. 2017. Deductive Cogency, Understanding, and Acceptance. *Synthese*, DOI: 10.1007/s11229-017-1365-4.
- Foley, Richard. 1992. The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly* 29, 111–124.
- Goldman, Alvin I. 2001. Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research* 63, 85–110.
- Horwich, Paul. 1982. *Probability and Evidence*. Cambridge: Cambridge University Press.
- Jeffrey, Richard C. 1970. Dracula Meets Wolfman: Acceptance vs. Partial Belief. *Induction, Acceptance, and Rational Belief* (bls. 157–185). Ritstj. Marshall Swain. Boston: D. Reidel.
- Joyce, James M. 2005. How Probabilities Reflect Evidence. *Philosophical Perspectives* 19, 153–178.

35 Ég vil þakka Hlyni Orra Stefánssyni og nafnlausum ritrýni *Hugar* fyrir mjög gagnlegar athugasemdir og uppástungur. Vinna við þessa grein var styrkt af Starfslaunasjóði sjálfstætt starfandi fræðimanna sem hluti af verkefninu *Vísindi, samfélag og gagnrýnin hugsun* (styrknúmer 15566). Ég vil einnig þakka Reykjavíkúrákadémiunni og meðlimum þess fyrir ýmiss konar stuðning við verkefnið.

- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.
- Kaplan, Mark. 1985. It's Not What You Know that Counts. *The Journal of Philosophy* 82, 350–363.
- Kaplan, Mark. 1996. *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.
- Kolodny, Niko. 2007. How Does Coherence Matter? *Proceedings of the Aristotelian Society* 117, 229–263.
- Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.
- Kyburg, Henry E. 1961. *Probability and the Logic of Rational Belief*. Middletown: Wesleyan University Press.
- Maher, Patrick. 1988. Prediction, Accommodation, and the Logic of Discovery. *Proceedings of the Philosophy of Science Association* (1. bindi, bls. 273–285). Ritsjt. Arthur Fine og Jarrett Leplin. East Lansing: Philosophy of Science Association.
- Maher, Patrick. 1993. *Betting on Theories*. Cambridge: Cambridge University Press.
- Makinson, David. 1965. The Paradox of the Preface. *Analysis* 25, 205–207.
- Sober, Elliot. 2009. Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection with Fossils, Fishing, Fine-tuning, and Firing Squads. *Philosophical Studies* 143, 63–90.
- Strevens, Michael. 2009. Objective Evidence and Absence: Comment on Sober. *Philosophical Studies* 143, 91–100.
- Tversky, Amos og Daniel Kahneman. 1983. Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90, 293–315.
- van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.

Abstract

From Belief to Credence and Back Again: An Overview of Bayesian Epistemology

This paper discusses the delicate relationship between traditional epistemology and the increasingly influential probabilistic (or ‘Bayesian’) approach to epistemology. The paper introduces some of the key ideas of probabilistic epistemology, including credences (degrees of belief), Bayes’ theorem, conditionalization, and Dutch Book arguments. The tension between traditional and probabilistic epistemology is brought out by considering the lottery and preface paradoxes as they relate to rational (binary) belief and credence respectively. It is then argued that this tension can be alleviated by rejecting the requirement that rational (binary) beliefs must be consistent and closed under logical entailment. Instead, it is suggested that the requirements of consistency and closure apply to a different type of binary propositional attitude, viz. acceptance.