

La filosofia i els estudis experimentals. Una reflexió crítica

Genoveva Martí*

Resum

La filosofia experimental és un corrent dins la filosofia contemporània que utilitza dades empíriques, provinents d'enquestes i estudis, per posar a prova tesis i teories filosòfiques. Els filòsofs experimentals consideren les reaccions del públic en general als experiments mentals, un recurs que els filòsofs sovint utilitzen per il·lustrar les conseqüències que es deriven de principis filosòfics quan aquests s'apliquen a situacions pràctiques, atès que són rellevants a l'hora de justificar o refutar una posició filosòfica. El meu propòsit és examinar críticament alguns dels supòsits de la filosofia experimental.

Paraules clau

Filosofia experimental, experiments mentals

Recepció original: 28 de setembre de 2020

Acceptació: 19 d'octubre de 2020

Publicació: 1 de juny de 2021

Introducció¹

El propòsit de la meua intervenció avui és reflexionar críticament sobre un moviment o corrent filosòfic que s'ha estès ràpidament en els darrers anys: l'anomenada filosofia experimental.

La filosofia experimental és un moviment que critica els mètodes que els filòsofs tradicionalment hem utilitzat per arribar a les tesis que mantenim i proposa mètodes diferents. És, tanmateix, un moviment molt heterogeni, els fonaments i objectius del qual no es poden caracteritzar en poques paraules. Hom pot trobar una força exhaustiva descripció de les varietats de filosofia experimental a l'article de la Stanford Encyclopedia of Philosophy, escrit per Joshua Knobe, un dels iniciadors del corrent.

A grans pinzellades la filosofia experimental utilitza dades empíriques, provinents d'enquestes fetes tant a filòsofs com a no-filòsofs, per extreure'n conclusions que es consideren rellevants per a la investigació i la discussió de teories filosòfiques. El cavall de batalla de la filosofia experimental han estat els 'experiments mentals' les situacions imaginades o concebudes, situacions usualment contrafactuals, tot i que poden ben bé estar basades en fets i històries reals, que molt sovint els filòsofs descrivim en la nostra argumentació per posar a prova les nostres teories, és a dir, per ressaltar la plausibilitat o, al contrari, la inadmissibilitat de les conseqüències d'algun principi o posició filosòfica, per

(*) Professora de Recerca d'ICREA (Institució Catalana de Recerca i Estudis Avançats) a la Facultat de Filosofia de la Universitat de Barcelona. Llicenciada en Filosofia per la Universitat de Barcelona i Doctora en Filosofia per la Universitat de Stanford. Ha estat professora a la Universitat de Washington, Seattle, la Universitat de Califòrnia, Riverside i la London School of Economics. La seva especialitat és la filosofia de llenguatge. Adreça electrònica: gmartí@ub.edu.

(1) Aquest text correspon a la lliçó inaugural del curs 2020-21 de la Societat Catalana de Filosofia, que va ser dictada el dia 22 de setembre de 2020. L'autora agraeix al Prof. Josep Monserrat i al Prof. Conrad Vilanou la invitació a impartir aquesta lliçó inaugural de la Societat Catalana de Filosofia, filial de l'Institut d'Estudis Catalans, que va tenir lloc a la sala Pere Coromines de la seu de l'Institut.

revelar contradiccions internes entre diversos principis o senzillament amb una intenció pedagògica, per aclarir els compromisos d'una tesi molt abstracta. Els filòsofs experimentals consideren que els filòsofs, diguem-ne, tradicionals, cada un de nosaltres ens recolzem exclusivament en les nostres pròpies reaccions i respostes a aquests experiments mentals, i defensen que, a l'hora de posar a prova una tesi, hauríem de recollir i tenir en compte les respostes no solament d'altres filòsofs, sinó també de la població en general.

La filosofia experimental és un moviment molt estès. Hi ha grups i societats de filosofia experimental als Estats Units (la Societat de Filosofia Experimental és part de l'American Philosophical Association), al Regne Unit, a Polònia, Alemanya, Portugal, Espanya i altres països.

Així doncs, els filòsofs experimentals consideren que tradicionalment els filòsofs hem fet filosofia 'de butaca': asseguts en la solitud del nostre estudi, elaborem teories, o proposem principis i els sotmetem a judici consultant les nostres pròpies intuïcions i opinions sobre el que se'n deriva d'aquestes teories quan els seus principis s'apliquen en les situacions contrafactuals que imaginem en els nostres experiments mentals, assumint que les nostres pròpies reaccions són les correctes, les naturals, les que una persona que pensés críticament i seriosament sobre la situació en qüestió tindria. Els filòsofs experimentals veuen aquesta actitud com una mena d'arrogància del filòsof, i consideren que hauríem de 'aixecar-nos de la butaca' per comprovar si l'assumpció és correcta, comprovar si en general la gent comparteix la reacció del filòsof si expressen les mateixes actituds o emeten els mateixos veredictes quan consideren un experiment mental. Els resultats, val a dir, són moltes vegades contraris a les reaccions dels filòsofs. Ara bé, hauria això d'ésser sorprenent? Jo penso que no. Cal haver rebut una formació filosòfica, haver-se entrenat en fer moltes anàlisis rigoroses de les situacions sobre les que hom reflexiona, amb arguments sòlids i lògicament correctes, per tal d'emetre una opinió, que pot ser acceptada com a correcta, òbvia, de sentit comú, o molt discutible.

Hom no pot determinar democràticament, per exemple, si Modus Ponens és una regla correcta de raonament. De fet si atenem a alguns estudis psicològics, hauríem de concloure que no és una regla universal, doncs no figura en el catàleg de normes de raonament de moltes persones en un seguit de circumstàncies (Vegeu, per exemple, Blanchette i Leese, 2011). Els resultats d'un vot sobre la correcció de la regla de Modus Ponens ens podrien semblar interessants, curiosos, fins i tot podrien ésser útils per fer algun tipus de reflexió. Però no constitueixen una resposta a cap pregunta elemental sobre la fonamentació de la regla.

Jo he estat molt crítica amb el moviment de la filosofia experimental (Vegeu, per exemple, Martí, 2009 i 2012). Després mencionaré amb una mica més de detall algunes de les crítiques tant meves com d'altres filòsofs. Però de forma molt general: senzillament em sembla que la forma d'establir, posar a prova i justificar tesis filosòfiques no passa per fer una enquesta pel carrer o una votació a mà alçada.

Amb això no vull dir que la filosofia s'hagi de fer des de la torre de marfil, completament desconnectada del món real, de les creences, les preocupacions o les opinions de la societat que ens envolta: el punt important que vull ressaltar és que no són les creences o les reaccions d'una majoria o d'un grup important de la societat les que juguen un paper a l'hora de determinar si una tesi filosòfica és correcta o plausible.

Una altra cosa que és molt important deixar clara és que no estic negant en absolut la importància que la filosofia es recolzi en la ciència. Hom no pot posar en dubte que la pràctica i la reflexió filosòfiques sovint estan informades per la ciència empírica i pel saber en altres disciplines. Les tesis d'una persona que teoritza sobre la justícia i sobre els fonaments d'una societat justa, inevitablement toquen de prop qüestions que clarament cauen dins l'esfera del dret. Un filòsof que reflexioni sobre la mort, molt probablement voldrà entendre en profunditat i tenir en compte temes que pertanyen a la biologia o la investigació mèdica. Si hom teoritza sobre la causalitat inevitablement haurà de parar atenció a teories de la física. Molts dels temes que interessin als filòsofs requereixen una base sòlida en altres disciplines. Això, crec que queda clar, no ho qüestiono en cap moment.

Els filòsofs experimentals suggereixen que l'ús dels experiments mentals és una pràctica quasi bé exclusiva de la filosofia analítica contemporània i es presenten com a crítics d'aquesta tradició. Però el recurs de l'experiment mental des de sempre ha jugat un paper important en la pràctica filosòfica, amb propòsits i funcions molt diversos. Mencionaré alguns exemples que em permetran il·luminar algunes de les crítiques a la filosofia experimental.

Per exemple la primera formulació de l'imperatiu categòric que proposa Kant «actua solament d'acord amb màximes que puguis voler que es converteixin a la vegada en lleis universals» (Kant 1785, 4: 421) sota la interpretació d'Onora O'Neill (1975) i John Rawls (1980) involucra de forma crucial un experiment mental. Segons O'Neill i Rawls la formulació de Kant encapsula un procediment: formula la màxima que expressa els motius que governen l'acció. Expressa-la com una llei natural per tots els agents racionals i, finalment, valora si un món governat per la llei en qüestió és possible.

En aquest cas, l'experiment mental té la funció de posar a prova la racionalitat d'una norma de conducta. Si la interpretació d'O'Neill i Rawls és correcta, la filosofia moral kantiana convida a fer experiments mentals. Però dubto que cap filòsof es deixés impressionar si una majoria del públic fos incapaç de detectar la contradicció inherent que una mala màxima, concebuda com a llei universal, provocaria (si no ho veuen, no ho veuen, la contradicció és objectiva). Això ens ha de fer pensar que hi ha temes (la validesa de Modus Ponens, la contradicció generada per l'aplicació d'una màxima) que depenen de qüestions objectives i que són completament independents del veredictes o de la reacció d'un grup o una majoria d'individus.

Altres experiments mentals tenen una funció diferent: estan dissenyats per concloure que una certa reacció davant la situació descrita és natural, de sentit comú. I són aquests tipus d'experiments mentals els que estan al punt de mira dels filòsofs experimentals, doncs si hom està disposat a defensar que una certa reacció és la més natural, perquè no comprovar si aquesta opinió és compartida? Examinem alguns exemples d'aquest tipus d'experiment mental.

Hume ens convida a considerar si una persona no seria capaç de formar la idea d'un cert to de blau malgrat no haver-lo experimentat mai:

Suposem que tots els tons d'aquest color, excepte aquell to en particular, se li mostren [a aquesta persona]... és obvi que percebrà un salt allà on el to en qüestió hauria d'ésser... I pregunto si no fora possible, sobre la base de la seva pròpia imaginació, ... donar-se a sí mateix la idea d'aquell to particular ... ? (Hume, 1748, Secció II: «On the origin of ideas», paràgraf 16).

És una funció típica dels experiments mentals la de suggerir un possible problema per una teoria contrastant el que es deriva de la teoria amb allò que sembla ser una resposta natural a una situació descrita. El que té d'interessant aquest cas és que Hume sembla utilitzar el mètode per indicar un possible problema per la seva pròpia teoria que les idees simples es deriven d'impressions.

També podem recordar la història del príncep i el sabater amb la que Locke il·lustra i defèn la seva teoria de la identitat personal: si la consciència, els records i el caràcter del príncep es transportessin al cos del sabater, la persona que anomenem 'príncep' persistiria encara que tingués el cos del sabater (Locke, 1694, Llibre II, capítol 27, secció 15). I Locke ens presenta el cas amb la seguretat que això és el que diríem si reflexionem sobre la situació. De fet ens diu explícitament: «tothom veu que seria la mateixa persona.»

No solament filòsofs occidentals utilitzen el recurs. En altres tradicions (aràbiga, xinesa) el recurs a l'experiment mental s'utilitza amb els mateixos propòsits.

Avicenna (980-1037) ens convida a imaginar un home enlairat, que mai ha tingut cap experiència sensorial, per argumentar que fins i tot en aquestes condicions aquest home tindria consciència de la seva pròpia existència. I podem anar més enrere i més lluny encara: Meng Zi, o Menci, que va viure al segle IV abans de la nostra era, utilitzava el conte d'un nen que està a punt de caure en un pou, per argumentar, contra Yang Zhu, que la naturalesa humana no està exclusivament motivada per consideracions egoistes: qualsevol que observés el nen a punt de caure al pou sentiria alarma i compassió. La compassió és per a Menci part de l'essència del que és ésser humà (Van Norden, 2008).

Aquests quatre exemples d'experiment mental apelen a la resposta 'òbvia', 'natural', 'de sentit comú'. Però què passa si no 'tothom ho veu'?

Sobre això podem fer algunes observacions:

En primer lloc, els desacords sobre la plausibilitat de les conseqüències de tesis filosòfiques no són un fenomen que enfronta el filòsof amb el públic en general. Els desacords existeixen també i primordialment entre filòsofs. No tots els filòsofs estan d'acord amb Locke quan reflexionen sobre l'exemple del príncep i el sabater. Els filòsofs sotmetem les nostres posicions a escrutini, a objeccions i a respostes. La imatge del filòsof a la torre d'ivori, pontificant mentre la resta de la humanitat tenen intuïcions molt diferents no es correspon amb la realitat de la pràctica filosòfica. Per obtenir opinions i intuïcions diferents, no cal fer enquestes. I de fet, alguns estudis que han fet els filòsofs experimentals, comparant les reaccions de filòsofs i del públic, indiquen l'existència de quasi bé tant desacord entre els experts com entre públic en general (Vegeu, per exemple, Culbertson i Gross, 2009).

En segon lloc, vull tornar ara al comentari d'abans: la filosofia no està desconnectada del món que ens envolta. Tant si es dedica a l'ètica, la teoria de la referència o la identitat personal, el filòsof és part d'una comunitat. El filòsof observa; Menci en cert sentit és més experimental que els filòsofs experimentals: ell fa observacions sobre certes reaccions dels éssers humans i les utilitza en la seva filosofia (i el mateix és de suposar havia fet Yang Zhu, ressaltant reaccions diferents per defensar conclusions oposades a les de Menci).

Podríem dir que en certes àrees de la filosofia, no totes naturalment, el filòsof, inicialment, és una mena d'antropòleg que practica el mètode d'observació participativa. Però,

naturalment el filòsof no es queda aquí. No és la funció de la filosofia simplement descriure les reaccions humanes. I aquí entro en el tercer punt crític, el més rellevant des del meu punt de vista. I és que les dades per elles mateixes no constitueixen una refutació o una confirmació d'una teoria filosòfica. Les dades, en tot cas, són el primer pas per la reflexió i l'anàlisi crítica, per l'argument rigorós, per a la consideració de possibilitats diverses, per aquest sistema d'equilibris variables, de tortuós 'estira i arronsa' que constitueix el camí d'arribada a una posició determinada en filosofia.

Què faria Menci si la seva història generés reaccions majoritàriament oposades? Tancar la barraca i acceptar la posició de Yang Zhu? Crec que no. És ben sabut, per exemple, que algunes de les reaccions a l'experiment mental del tramvia desbocat són, si més no, psicològicament interessants. En la història original del tramvia desbocat (originalment proposada per Philippa Foot al 1967), hem d'imaginar que ens trobem sols en un tramvia sense frens que viatja a tota velocitat. Si no fem res, el tramvia atropellarà a cinc persones innocents. Però podem desviar-lo i en aquest cas el tramvia atropellarà a una persona (Foot, 1978). La història s'utilitza en pràcticament totes les classes introductòries d'ètica. Ens ajuda a posar en evidència els conflictes entre diferents principis (és pitjor provocar un mal o permetre que succeeixi? És sempre millor beneficiar al major nombre de persones?). En una versió un tant perversa de la història del tramvia sense frens, hem d'imaginar que no estem dins el tramvia sinó observant l'escena i podem activar un mecanisme pel qual una persona que està a prop de les vies cau davant del tramvia salvant així les cinc persones al final del tram. Diversos estudis mostren que quan se'ls presenta aquesta versió, si la persona que hom pot fer caure davant el tramvia sembla ser una persona pobre, bruta i en parracs, molta més gent troba acceptable accionar el mecanisme que si la persona en qüestió es descriu com una persona neta i ben vestida (Vegeu, per exemple, Bleske-Rechek, 2010). Què n'hem de fer d'això com a filòsofs? Capítular i acceptar que és acceptable moralment actuar d'una forma o una altra en la situació descrita? Crec que no paga la pena fer més elaboracions.

En resum: les dades empíriques que recullen els filòsofs experimentals sí que poden ésser útils, però en moltes ocasions, ja les tenim; i en qualsevol cas, només serien l'inici de la reflexió i el debat filosòfic. Si resulta que molta gent tenen, o no tenen, les mateixes reaccions als exemples de Hume, Locke, Avicenna o Menci, aquesta dada és el principi, no el final de la discussió.

També, i això em porta a una altra crítica, les dades recollides en les enquestes dels filòsofs experimentals sovint ni tan sols són útils per començar a fer la reflexió filosòfica per a la qual ells pensen que són rellevants. Així doncs ens trobem amb una quantitat substancial de publicacions en revistes de filosofia i en revistes interdisciplinars de psicologia i filosofia que no tenen cap rellevància pel problema concret que hom suposa que guia la investigació. Hi ha més d'un exemple, però aquí voldria mencionar un de molt recent que em sembla molt clar. Fa poc s'ha publicat un estudi en què els autors es proposen comprovar si la història del vaixell de Teseu és percebuda universalment com un 'puzle', un trencaclosques filosòfic sobre els principis que governen la identitat dels artefactes. L'experiment mental del vaixell de Teseu, tal com el presenta Hobbes (1655, 2, 11, 7) és força interessant. Considerem un vaixell al qual se li van fent malbé els taulons del folre que progressivament es van substituint per nous. Un procés de renovació que s'allarga durant molts anys fins que pràcticament no queda cap tauló original.

La idea és que en cada pas del procés en què hom substitueix un tauló, el vaixell continua essent el mateix i així fins al final del procés. D'una altra banda considerem un vaixell al qual se li desmunten tots els taulons per netejar-los i es tornen a muntar. També en aquest cas diríem que el vaixell desmuntat i tornat a muntar és el mateix. El problema interessant sorgeix quan combinem les dues històries com va fer Hobbes, quan imaginem que tenim el vaixell renovat, però a la vegada algun ha anat guardant els taulons envellits i munta un vaixell. Quin és el vaixell de Teseu llavors? Sembla que, considerant les històries per separat, tant un com l'altre, el vaixell renovat i el vaixell tornat a muntar, poden ésser considerats el vaixell de Teseu, però no poden ser el mateix, doncs està clar que són dos vaixells diferents.

Aquest és el trencaclosques que molts de nosaltres hem utilitzat en classes introductòries de filosofia per presentar els problemes associats als principis que governen, o semblen governar, la identitat dels objectes. Hi ha hagut molta discussió en la literatura sobre la història del vaixell de Teseu. Alguns filòsofs argumenten que el vaixell renovat és el vaixell de Teseu, sobre la base d'arguments de la prioritat de la identitat de forma sobre la identitat de matèria. Altres, al contrari, argumenten que el vaixell reconstruït amb les taules originals és l'autèntic vaixell de Teseu, donant prioritat a uns altres principis d'identificació. Altres filòsofs consideren que la història és autènticament paradoxal i que mostra que els principis d'identificació que funcionen prou bé en situacions del dia a dia, poden portar a contradiccions. En tots els casos, el problema és concebut com un autèntic problema amb o sense solució.

Bé, un grup nombrós de filòsofs (Rose et al., 2020) van decidir determinar si la història del vaixell de Teseu era també percebuda com un problema, un trencaclosques, per la població en general, en diferents països i diferent cultures. Cosa que, en principi, sembla prou interessant.

Què van fer aquests autors? Van dissenyar una vinyeta amb una versió de la història de Hobbes, la van donar a llegir als participants en l'estudi i els van preguntar: «Segons tu, quin és el vaixell de Teseu, el que s'ha anat renovant o el tornat a muntar?» Vora 60% dels participants contesten que és el renovat, i més del 30% contesten que és el tornat a muntar. Com que la minoria és prou significativa, els autors conclouen que la història del vaixell de Teseu és efectivament un puzzle filosòfic.

L'estudi va aparèixer recentment al volum 3 de la col·lecció Oxford Studies in Experimental Philosophy. És un article amb més de quaranta autors i és part d'un projecte subvencionat per la Templeton Foundation amb més de 2 milions d'euros.

Clar, el fet que hi hagi un desacord entre diverses persones sobre la pregunta, no significa que la història del vaixell de Teseu sigui paradoxal, problemàtica, un trencaclosques. El que fa que una història, o la descripció d'una situació constitueixi un puzzle no és el desacord, sinó l'ambivalència personal. El que ens indica que el problema del Tramvia és un problema és precisament el fet que quan reflexionem sobre les alternatives, fins i tot si acabem prenent una decisió, podem veure la dificultat, la tensió, perquè els principis que porten a cada una de les alternatives són plausibles. En la història del vaixell de Teseu, el principi segons el qual els canvis graduals no alteren la identitat d'un objecte, és plausible; el principi segons el qual desmuntar i muntar un objecte no altera la seva identitat és plausible (si no fos així, no portaríem pas el rellotge que ens va llegar l'àvia a netejar-lo a la joieria). Però ens trobem amb una història que ens empenya en dues direccions oposades. És aquesta ambivalència el que indica la presència d'un cas difícil, no el

fet que una part de la població diu una cosa i una part diu una altra. En l'estudi que han fet aquests autors, no hi ha cap indicació de l'existència d'ambivalència ni cap indicació que hagin intentat mesurar-la.

Si teníem un interès en descobrir si la història del vaixell de Teseu és percebuda com problemàtica (i, insisteixo, per fer això només cal explicar-la en una classe introductòria de filosofia o en una sessió del Barcelona Pensa), o si teníem interès en saber si hi havia alguna diferència cultural en la percepció de la història com un puzzle, per exemple, aquest estudi no ens aporta cap dada.

Aquesta és la conclusió que defensem Marta Campdelacreu, Ramón García-Moya, Enrico Terrone i jo mateixa en un article que fa poc hem escrit: «How to test the Ship of Theseus» (manuscrit).

Tanmateix, i perquè crec que hom pot aprendre coses importants de tots els corrents i doctrines filosòfiques, voldria acabar fent una reflexió sobre un estudi en concret, de fet l'estudi que se situa al bell inici del moviment experimental; pel que fa a aquest estudi, crec, les dades que ens aporta són útils i ens poden portar a qüestionar algunes tesis que, en general, considerem ben establertes.

Per discutir l'estudi primer hem de repassar algunes qüestions molt bàsiques.

Hi ha una distinció entre allò que tenim la intenció d'aconseguir, les conseqüències de les nostres accions que són deliberades o intencionals i les conseqüències que tot i ésser previstes i conegudes, no són part d'allò que tenim intenció d'aconseguir. Per exemple, si jo moc la mà per saludar a una persona, moure la mà és una acció intencional. Sé perfectament que quan moc la mà mouré les molècules d'aire al seu voltant, però aquesta és una conseqüència, coneguda per mi, de la meua acció. Moure les molècules no és el que tinc intenció de fer. És, diguem-ne, un efecte colateral. La diferència s'ha mirat de captar de formes diverses (moure les molècules no és intencional perquè l'acció que faries seria del mateix tipus encara que estiguessis a la Lluna...). Pels meus propòsits aquesta no és una qüestió important. El que ens interessa és que hi ha aquesta distinció i que és una distinció ben establerta i acceptada. Seguint de nou a Philippa Foot (1978, p. 20) una persona no té la intenció d'aconseguir directament o indirectament les conseqüències de les seves accions voluntàries quan aquestes conseqüències no són ni la finalitat de la seva acció, ni els mitjans per obtenir-la.

I la distinció no és merament ociosa, com es diu a vegades «una distinció sense una diferència». És rellevant filosòficament i també per efectes pràctics. El principi del doble efecte es basa en aquesta distinció i ha estat una doctrina, correcta o incorrecta, d'importància cabdal en molts àmbits. Segons la doctrina del doble efecte és a vegades permisible provocar resultats que no són els resultats intencionals o pretesos, sinó simplement resultats previstos, encara que siguin contraris al bé d'algú. El principi ha estat amplament discutit en debats sobre l'avortament i sobre l'eutanàsia. De nou la meua discussió no se centra sobre si el principi del doble efecte és o no correcte. L'única qüestió important és que hom pot distingir entre les conseqüències merament previstes i les conseqüències intencionals d'una acció i que aquesta distinció té rellevància filosòfica.

Segona distinció: les conseqüències d'algunes accions poden ser bones o poden ser dolentes, independentment de si són les conseqüències que teníem la intenció de produir o no. És a dir el valor de qualsevol cosa que passa (bona, dolenta, neutral) és independent de si teníem o no teníem la intenció de produir-la. Si premo un botó i això causa

un desastre nuclear, la conseqüència és catastròfica, si soluciono el problema del canvi climàtic la conseqüència és molt bona, independentment de si jo tenia la intenció de causar aquestes conseqüències.

Per tant, aquí tenim una tesi molt simple: si una acció és o no és intencionada no depèn de si les conseqüències de l'acció són bones o dolentes, ja siguin conseqüències preteses, o merament previstes.

I bé, aquí està l'estudi: Joshua Knobe (actualment professor de Lingüística, Filosofia i Psicologia a Yale) va presentar un experiment mental per comprovar si aquesta tesi que filosòficament sembla òbvia guiava les reaccions del públic en general. Knobe (2003) va presentar als participants en el seu estudi una història aproximadament com la següent:

La persona que presideix una companyia ha decidit implementar un nou programa. És un programa que reportarà molts diners a la companyia, i això és el que interessa a aquesta persona (i al consell d'administració, naturalment). La directora sap que el programa tindrà una certa conseqüència X, però no té cap interès en X. Tot el que l'interessa és l'efectivitat i rendiment econòmic del programa.

Per un grup dels voluntaris participants X es va descriure com una conseqüència catastròfica pel medi ambient. I per l'altre grup X es va descriure com una conseqüència beneficiosa pel medi ambient.

La pregunta pels dos grups, la mateixa: Creus que X és una conseqüència pretesa intencionadament?

El grup per al qual X havia estat descrita com una conseqüència beneficiosa pel medi ambient va respondre que no, com és d'esperar: tal com està descrita la situació, X és una conseqüència prevista però no pretesa, és un efecte col·lateral de l'aplicació del programa, que és intencional, per a aconseguir diners, que és la conseqüència que hom té la intenció d'aconseguir. Fins aquí tot quadra.

Però el grup pel qual X havia estat descrita com catastròfica, va respondre que X era una conseqüència pretesa, és a dir que el director tenia la intenció de provocar aquesta conseqüència, aquesta catàstrofe.

Això és una sorpresa, doncs si analitzem amb cura com hem fet abans la diferència entre allò que pretenem, allò que intencionalment volem provocar, i altres conseqüències conegudes però no preteses de les nostres accions, l'opinió sobre si la directora de la companyia provoca aquestes conseqüències, beneficioses o desastroses, intencionadament no hauria de canviar pel fet que siguin bones o dolentes.

Val a dir que aquest experiment s'ha fet posteriorment d'altres formes, per tal d'eliminar factors que podrien haver influenciat els participants, eliminant per exemple la referència a grans companyies i els seus directors executius que, potser d'entrada, són sospitosos de males intencions. Però els resultats semblen prou sòlids.

Bé, què n'hem de fer d'això? La inclinació inicial, crec, és pensar que els participants no han entès aquesta diferència tan important entre conseqüències preteses i resultats col·laterals que no són ni pretesos ni mitjans per a aconseguir l'objectiu pretès, i que per tant potser els ho hem d'explicar amb més claredat, convidant-los a fer un curs introductor de filosofia on es parli de filosofia de l'acció, volició i intencionalitat. Jo, com a mínim, desitjaria que els meus alumnes no cometessin aquest error.

Però d'una altra banda, aquests resultats també fan pensar. Fan pensar si no podríem potser argumentar que quan les conseqüències de les nostres accions tenen un gran impacte, i un resultat catastròfic és de gran impacte, potser no és possible que el resultat, essent conegut, no sigui a la vegada pretès.

Si hom no evita l'esdeveniment d'un mal conegut, fins a quin punt no podem dir que la seva inacció és voluntària i, per tant, volgudes, i preteses les seves conseqüències?

L'exemple de Knobe se centra en una acció puntual, aïllant-la de l'entramat d'accions i conseqüències que són part de plans i projectes de la vida real. Potser si tinguéssim una descripció més propera a la vida real podríem avaluar millor la situació.

Curiosament Michael Bratman, un filòsof de l'acció, clàssic, no experimental, l'any 1984 ja havia publicat un assaig en el que posava en dubte que la distinció entre efecte col·lateral no pretès, i resultat intencional es pogués mantenir sempre nítidament. Però Bratman és precisament un filòsof que pensa en termes de plans, projectes i objectius, que integren múltiples accions connectades.

En qualsevol cas, l'experiment d'en Joshua Knobe té interès, ens pot fer pensar, però vull ressaltar dues coses: (i) que això ja se li havia ocorregut a un altre filòsof sense fer cap enquesta, i (ii) que, en qualsevol cas, les dades ens poden servir per iniciar una reflexió i una discussió, però per sí soles no proven res sobre la teoria de l'acció i la teoria de la intencionalitat, per sí soles i sense anàlisi crítica no contribueixen a confirmar ni a refutar cap tesi filosòfica.

I aquesta és la lliçó que m'agradaria endur-me a casa.

Referències

- Avicena (Ibn Sina, 980-1037) *On the Soul*. Traduït per J. McGinnis i D.C. Reisman. Hackett (2007).
- Blanchette, I., Leese, J. (2011) «The effect of negative emotion on deductive reasoning: examining the contribution of physiological arousal». *Journal of Experimental Psychology*, 58(3), p. 235-246.
- Bleske-Rechek, A. (2010) «Evolution and the trolley problem: people save five over one unless the one is young, genetically related, or a romantic partner». *Journal of Social, Evolutionary, and Cultural Psychology*, 2010, 4(3), p. 115-127.
- Bratman, M. (1984) «Two Faces of Intention». *The Philosophical Review*, 93 (3), p. 375-405.
- Campdelacreu, M., García-Moya, R., Martí, G., Terrone, E. (manuscrit) «How to test the Ship of Theseus».
- Culbertson, J., Gross, S. (2009) «Are Linguists Better Subjects?». *British Journal for the Philosophy of Science*, 60, p. 721-736.
- Foot, P. (1978) «The Problem of abortion and the doctrine of double-effect». *Oxford Review*, 1967. Reimprès a *Virtues and Vices*, 1978. Blackwell.
- Hobbes, T. (1655) *De Corpore*. Traducció del llibre I per A.P. Martinich. Nova York, Abaris Books, 1981.
- Hume, D. (1748) *An enquiry concerning human understanding*. Edició de L.A. Selby-Bigge.
- Kant, I. (1785) *Groundwork of the Metaphysics of Morals*. Traducció i edició de M. Gregor.

- Knobe, J. (2003) «Intentional action and side effects in ordinary language». *Analysis*, 63 (279), p. 190-194.
- Locke, J. (1694) *An essay concerning human understanding*, 2a edició, 1694. Edició d'A. C. Fraser.
- Martí, G. (2009) «Against Semantic Multiculturalism». *Analysis*, 69, p. 42-48.
- Martí, G. (2012) «Empirical Data and the Theory of Reference». W. P. Kabasenche, M. O'Rourke i M. H. Slater (eds.) *Reference and Referring. Topics in Contemporary Philosophy*. MIT Press, p. 63-82.
- O'Neill, O. (1975) *Acting on Principle*. New York, Columbia University Press.
- Rawls, J. (1980) «Kantian constructivism in moral theory». *Journal of Philosophy*, 77, p. 515-572.
- Rose, D. et al. (2020) «The Ship of Theseus Puzzle». T. Lombrozo, J. Knobe i S. Nichols (eds.), *Oxford Studies in Experimental Philosophy*. Volum 3. Oxford University Press.
- Van Norden, B.W., traductor (2008) *Mengzi: With Selections from Traditional Commentaries*. Indianapolis, Hackett Publishing.

La filosofía y los estudios experimentales. Una reflexión crítica

Resumen: La filosofía experimental es una corriente dentro de la filosofía contemporánea que utiliza datos empíricos, provenientes de encuestas y estudios, para poner a prueba tesis y teorías filosóficas. Los filósofos experimentales consideran las reacciones del público en general a los experimentos mentales, un recurso que los filósofos a menudo utilizan para ilustrar las consecuencias que se derivan de principios filosóficos cuando estos se aplican a situaciones prácticas, ya que son relevantes a la hora de justificar o refutar una posición filosófica. Mi propósito es examinar críticamente algunos de los supuestos de la filosofía experimental.

Palabras clave: Filosofía experimental, experimentos mentales

La philosophie et les études expérimentales. Une réflexion critique

Résumé : La philosophie expérimentale est un courant au sein de la philosophie contemporaine qui utilise des données empiriques, provenant d'enquêtes et d'études, pour mettre à l'épreuve des thèses et des théories philosophiques. Les philosophes expérimentaux étudient les réactions du public en général face aux expériences mentales, une ressource que les philosophes utilisent souvent pour illustrer les conséquences dérivées de principes philosophiques, quand ceux-ci s'appliquent à des situations pratiques, compte tenu de leur pertinence pour justifier ou réfuter une position philosophique. Mon objectif est d'examiner sous un angle critique certaines hypothèses de la philosophie expérimentale.

Mots clés: Philosophie expérimentale, expériences mentales

Philosophy and experimental studies. A critical reflection

Abstract: Experimental philosophy is a current within contemporary philosophy that uses empirical data from surveys and studies to test philosophical theories and theses. Experimental philosophers consider the general public's reactions to thought experiments. This device is often used by philosophers to illustrate the consequences resulting from philosophical principles when they are applied to practical situations, as these experiments are relevant in justifying or refuting a philosophical position. The aim is to critically examine some of the assumptions of experimental philosophy.

Keywords: Experimental philosophy, thought experiments