DOCUMENTOS DEL CIECE

Reflexiones sobre causalidad, control y explicación en *Prospect Theory*

Marqués, Gustavo

Octubre 2016

Número 13 2016

Staff

Centro de Investigación en Epistemología de las Ciencias Económicas

Facultad de Ciencias Económicas
Universidad de Buenos Aires

Av. Córdoba 2122 1º p. Aula 111
(1120) Ciudad Autónoma de Buenos Aires
Argentina
Tel. (54-11) 4370-6152
Correo electrónico: ciece.fce@fce.uba.ar

ISSN: 1851-0922

Reflexiones sobre causalidad, control y explicación en *Prospect Theory*.

Gustavo Marqués (CIECE – IIEP)

Introducción

En este trabajo se continúa y profundiza la investigación iniciada en Marqués – Weisman (2015) acerca de la relevancia epistémica y práctica de *Prospect Theory* (PT). Se examina la capacidad de PT para identificar los factores causales de cierto tipo de sesgos en la toma de decisiones entre loterías. En particular, se analiza la causa del fenómeno de reversión de preferencias y la aptitud de PT para controlar (vía manipulación), predecir y explicar este fenómeno. Por último, se examina la relación entre causalidad y mecanismo en el caso de PT.

Palabras claves: prospect theory, explicación, predicción, causalidad, control, mecanismo.

1. Causalidad

Psillos (2004) considera que hay dos conceptos básicos de causalidad (toma esta perspectiva de Ned Hall)

1) causation as *dependence*
2) causation as *production.*

Elucida la perspectiva (1) de esta manera: "to say that *c* causes *e* is to say that *e* suitably depends on *c*", donde "dependencia" es concebida como dependencia *contrafáctica* ("if the cause hadn't happened, the effect wouldn't have happened"). Esta interpretación parece tener dos variantes. En primer lugar, se afirma que dado un escenario en que estén presentes c y e, *si* c es la causa de e, siempre que sustraemos c, e deja de ocurrir. Debiera pasar esto porque de lo contrario no sería cierto que "si c no hubiera ocurrido e no se hubiera presentado". Pero entonces la causalidad puede ser concebida como condición (estrictamente) *necesaria*. En segundo lugar, puede entenderse la dependencia como condición *suficiente*. Así concebida la causa debe ser *única*, pues si e tuviera más de una causa, digamos c y h, no podría afirmarse que si c hubiese estado ausente el efecto no hubiese ocurrido. Habiendo más de una causa suficiente del mismo evento ninguna de ellas es estrictamente necesaria para la producción del mismo.

La visión (2), por su parte, sostiene que "to say that *c* causes *e* is to say that something *in* the cause produces (brings about) the effect or that there is something (e.g., a mechanism) that links the cause and the effect". Esta es una visión "mecanística" de causalidad, en la que causa se define en términos de mecanismo. No es la visión que adjudicaremos a PT, aunque volveremos sobre este problema más adelante.

En este trabajo vamos a sostener que el empleo de PT para identificar framing effects satisface la visión "causation as dependence", en el sentido de que identifica *factores causales que contribuyen de manera determinante* o *suficiente* a la ocurrencia de su efecto (la reversión de preferencias en el rango considerado por la teoría). Asimismo, sostendremos que de esta manera PT aporta una *explicación* (no mecanísmica) de su efecto.

2. Cómo adjudicamos Causalidad

Hay al menos tres cuestiones importantes en torno a la causalidad: (1) *ontológica*: cómo caracterizar el tipo de dependencia que se presenta entre causa y efecto (en términos de condición necesaria, suficiente, o qué….); (2) *uso epistémico* del conocimiento causal: si poseyéramos conocimiento causal, qué uso epistémico podríamos hacer de él (i.e., para explicar, predecir y controlar) y (3) *atribución* de conexión causal: sobre la base de qué *evidencia* estamos legitimados a adjudicar a un evento eficacia causal sobre otro. Haremos un breve análisis de las dos primeras cuestiones y nos detendremos con más detalle en la tercera. Para facilitar la exposición de los tres aspectos partamos de la versión de concepción causal expuesta en Kincaid (2004).

(1) Según Kincaid, las leyes físicas de la teoría gravitacional de Newton captan fuerzas, las cuales son caracterizadas como *factores causales*.

> "What kind of thing in the world does Newton's law pick out? The most natural answer is that it identifies a force. What is a force? It is a causal factor. A force is *causal* in that it influences something. It is a *factor* in that it need not be the only influence present. Modern physics, for example, identifies electromagnetic and nuclear forces that can be present at the same time as gravity. So a paradigm case of a law is a force or causal factor" (Kincaid, 2004, p.170).

Veamos otro de sus ejemplos, que es aún más revelador: "Diet, for example, is a causal factor in health" (íbid., p. 170). Los factores, así concebidos, no son determinantes, pues por sí solos no garantizan un resultado, sino que *contribuyen* a la ocurrencia de un cierto efecto. Su concepción parece adaptarse bien a la visión de la causalidad como condición *necesaria*. Debe advertirse que es una noción débil, ya que es fácil presa de la conocida objeción de que la presencia de oxígeno debe considerarse causa del incendio (aunque éste haya sido intencional)[i].

(2) En muchos pasajes Kincaid parece sostener que para predecir y explicar (correctamente) se requiere disponer de conocimiento causal. Dice, por ejemplo,

"The key question thus is whether the social sciences provide causal claims that provide relatively extensive explanations and predictions". (ibid, p.174)

"Then what role do laws play in science? Perhaps many, but above all, science produces laws to explain and reliably predict the phenomena. That is precisely what identifying causal factors should allow us to do" (ibid, p. 72).

Puesta en estos términos su visión describe dos *usos epistémicos* principales que podría darse al conocimiento causal: predecir y explicar. Ello supone que se cuenta *previamente* con la capacidad para identificar las causas, algo que Kincaid no argumenta, ya que no especifica ninguna manera empírica para recoger evidencia destinada a identificar conexiones causales que sea independiente de la predicción y la explicación.

(3) Más interesante para nosotros es el problema de *cómo adjudicamos* naturaleza o capacidad causal a un enunciado L. Para Kincaid estamos legitimados a atribuir causalidad a L cuando tenemos capacidad de predecir y explicar por su intermedio.

"In general, our confidence that causal claims are true is a function of how widely they explain and predict. (Harold Kincaid, 2004, p.174)

"A claim to know a causal factor is dubious to the extent that it does not allow us to explain and predict" (ibid., p.72).

De este modo, para atribuir justificadamente causalidad a una pareja de eventos se requiere que podamos predecir y explicar por su intermedio. El concepto de causalidad se torna dependiente de estas dos nociones. ¿Por qué de ambas, podría uno preguntarse?

Antes que nada veamos si "predicción" y "explicación" son independientes entre sí o alguna de ellas resulta redundante. Dicho más explícitamente, dado un enunciado L, ¿pueden concebirse dos conjuntos diferentes de evidencia empírica, E1 y E2, tales que respalden respectivamente las pretensiones explicativas y predictivas de L? Debería poder precisarse la naturaleza particular de la evidencia que se necesita en cada caso, si se desea distinguir netamente entre explicación y predicción sobre esta base. Sin embargo, no parece ser posible trazar esta distinción por medios puramente empíricos.

Por otra parte, si todo lo que se pide cuando se pretende que L explica es que sus predicciones se confirmen en alto grado, entonces la noción de explicación se vuelve redundante: para adjudicar causalidad sólo se necesitarían los conceptos de predicción y confirmación. Pero este enfoque ha sido rechazado por quienes consideran que ciertas correlaciones espurias permiten predecir sin ser causales. Es probable que Kincaid suscriba la tesis de que atribuir capacidad predictiva a un enunciado resulta ser insuficiente para atribuirle poder causal. Ello explica que traiga a escena una segunda característica: su capacidad *explicativa*. Pareciera que lo que tiene en mente es que si L permite predecir *y* explicar (*ambas* cosas) entonces sí estaríamos autorizados a atribuirle fuerza causal.

En realidad, "L permite predecir (correctamente)" y "L permite explicar (correctamente)" son afirmaciones muy diferentes. La capacidad predictiva puede ser establecida de manera puramente

empírica[ii]; la capacidad explicativa, en cambio, es una *valoración epistémica* que los humanos hacemos de L. Podríamos decir que "c explica e" es función de un conjunto de argumentos, de los que forma parte sin duda la capacidad predictiva, pero que excede largamente a esta capacidad. Usualmente se requiere que L sea verdadera o se encuentre al menos bien testada y confirmada. Abordaremos luego esta cuestión en referencia a PT.

Guerring ofrece una respuesta diferente a la de Kincaid al problema de cómo puede ser establecido que el evento X causa el evento Y. Sostiene que este es el caso cuando se observa co-variación entre ambos factores en contextos experimentales especiales.

"… one might observe that in a properly conducted experiment (i.e., with a randomized treatment and isolated treatment and control groups) it is often possible to demonstrate that some factor causes a particular outcome even though the pathway remains mysterious (perhaps because it is not amenable to experimental manipulation). We do not hesitate to label these arguments as causal and as definitive (assuming proper experimental protocols have been followed and replications have been conducted in a variety of settings" (Guerring, 2010, p.1505).

Guerring no necesita aludir a la capacidad explicativa de la teoría porque considera que un procedimiento como el descrito es suficiente para desestimar la idea de que la correlación hallada es accidental. Desde esta perspectiva, la evidencia que se necesita para adjudicar legítimamente conexión causal a dos eventos puede ser puramente empírica.

3. Causalidad en PT: *loss aversion y framing effects*

Examinaremos ahora de qué manera el tratamiento de los *framing effects* que realiza PT satisface la visión "causation as dependence", con la salvedad de que "dependencia" *no* es entendida al modo contrafáctico.

La psicología cognitiva ha obtenido importantes resultados experimentales que muestran que en muchas circunstancias los individuos adoptan decisiones sobre la base de un repertorio de disposiciones y heurísticas. Ejemplos de disposiciones son "loss aversion", "myopia", "adaptation", "saliency", "focus illusion" and "mental accounts". Las heurísticas incluyen a "default rule", "fifty – fifty" and "1/n rule" (Benartzi and Thaler, 2001, 2007).

Un resultado remarcable de este enfoque es que cuando un individuo enfrenta una situación en la que su riqueza (wealth) cambia respecto de un estado dado, su *utilidad* (medida en términos absolutos) es mayor cuando el cambio es representado en términos de pérdidas que en términos de ganancias. Este hecho, que refleja el impacto emocional producido por una alteración en el nivel de riqueza dado, es llamado *loss aversion*[iii]. Se ha mostrado que *loss aversion* influencia un amplio espectro de decisiones generando algunos patrones de conducta típicos. Dos de los más importantes son el llamado '*endowment effect*' y la diferente disposición a asumir riesgos cuando se enfrentan descripciones alternativas de una misma situación de elección (i.e., *framing effects*). El primer patrón surge en circunstancias de intercambio. Los individuos que padecen *loss aversion*

valúan más un bien cuando ya disponen de él que cuando no se encuentra en su poder. El fenómeno de *endowment effect* ha sido señalado inicialmente por Thaler (1980) y ha sido posteriormente demostrado en muy diferentes circunstancias experimentales (Knetch, 1989; Kahneman et.al., 1991).

El Segundo fenómeno es aún más impactante. Tversky and Kahneman (1981, 1986) mostraron que *loss aversion* es uno de los factores causales que generan reversión de preferencias en condiciones de incertidumbre. Esto ocurre cuando la misma situación de elección es descrita alternativamente en términos de ganancias y pérdidas. Consideremos por ejemplo el caso de la así llamada *Asian disease* (Kahneman, 2003a, pp. 1458).

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Descritas de esta manera la mayoría de los individuos eligen el Programa A por sobre el B, mostrando aversión al riesgo. Pero consideremos ahora el siguiente par de opciones:

If Program A' is adopted, 400 people will die.

If Program B' is adopted, there is a one third probability that nobody will die and a two-thirds probability that 600 people will die.

Cuando las opciones son descritas de este modo la mayor parte de la gente prefieren el Programa B' al Programa A', aunque es fácil advertir que A es equivalente a A', y que B es equivalente a B'. En ambos casos se describen las *mismas* opciones. La única diferencia consiste en que el primer caso las opciones son descritas en términos de vidas salvadas (ganadas) y en el segundo en términos de vidas perdidas. El cambio terminológico debería ser irrelevante para individuos completamente racionales, pero evidentemente no lo es para los sujetos bajo estudio, ya que sus preferencias en el seno del primer marco de referencia son revertidas cuando dicho marco se transforma del modo indicado. El modo en que las opciones son descritas parece ser más importante para ellos que las opciones mismas. Este descubrimiento representa una ampliación del dominio de aplicación de *loss aversion*, que muestra que en estas circunstancias influencia las actitudes acerca del riesgo. La teoría que proporciona este resultado es conocida como *Prospect Theory* (PT).

Claramente, *Prospect Theory* revela la existencia de conexión causal tanto si se toman en cuenta los estándares de Kincaid como los de Guerring. Satisface plenamente el requisito de co-variación y arroja predicciones acerca de qué tipo de cambios en el punto de referencia producirá qué cambios en el ranking de preferencias sobre loterías de una mayoría de individuos. Una virtud adicional de PT es que el factor causal identificado (*the frame* o el punto de referencia) puede ser *impuesto* a los tomadores de decisiones, lo que permite *intervenir* sobre el marco y *manipular* las decisiones de la mayoría de los individuos. Los interventores alteran el marco (punto de referencia) y hacen que los tomadores de decisiones elijan de una u otra manera, según lo esperado[iv]. En esto consiste el Paternalismo Libertario. La capacidad para producir el efecto deseado de manera sistemática es otro indicador –aún más fuerte que la mera predicción exitosa- de que se posee conocimiento causal relevante. Por ello sostenemos que estamos legitimados a decir que PT ha identificado una causa suficiente o determinante del efecto investigado (i.e., la reversión de preferencias).

Es remarcable que estos resultados no solo se cumplen en situaciones experimentales sino también en aplicaciones a condiciones ordinarias (Benartzi and Thaler, 2007¸ Benartzi and Thaler, 2001). Esto significa que en un buen número de circunstancias diferentes PT no afronta el llamado problema de validez externa. Sin embargo, una restricción importante a este desempeño exitoso es que en sentido estricto PT no anticipa cambios en la conducta individual, sino en las decisiones de *la mayoría* de individuos pertenecientes a un grupo de referencia. No obstante esto, muchos economistas creen este resultado es todo lo que la teoría económica es capaz de proporcionar (y todo lo que se necesita). En suma, PT permite manipular y controlar un factor causal central en cierto tipo de decisiones tomadas por una mayoría de individuos bajo condiciones de incertidumbre.

Reiss (2007) desarrolla un punto de vista que parece afectar seriamente a nuestro argumento. Dice que el poder de manipular un efecto (intervenir y producir su ocurrencia) en un tiempo t0, no asegura que se haya identificado a la causa de dicho efecto en t0 (debido al llamado problema de la causación común).

"Suppose we believe *X* to cause *Y* and *I* to be an intervention, both in Woodward's sense. But let the real structure be such that the correlation is because of a common cause *Z*, and let *I* affect *Z* or, alternatively, *X* and *Y* independently, perhaps such that *X* changes before *Y* changes, so that it looks as if *X* causes *Y*. There is nothing wrong with this from a policy point of view. But it would be mistaken to explain the change in *Y* by citing the change of *X"*.

En consecuencia, su postura implica que la capacidad que exhibe PT de manipular y generar a voluntad un efecto no es una garantía de que haya identificado su causa, salvo que se descarte la

posible acción de un factor común Z como el señalado. No podríamos afirmar que es el cambio en el marco lo que causa la reversión de preferencias, si no podemos garantizar (con seguridad y certeza) la inexistencia de una causa común o intermedia que podría ser en última instancia la responsable de la reversión de preferencias.

Creo que lo más sensato es desestimar la crítica misma (formulada en abstracto), o, más precisamente, invertir la carga de la prueba: si alguien descree que X (el factor manipulado) es la (verdadera) causa del efecto Y, debe indicar cuáles son las razones de su creencia, y si fueran razonables éstas podrían ser investigadas. Hay que negarse a admitir la relevancia de las objeciones basadas en la mera posibilidad. La posibilidad del genio maligno siempre está latente! La objeción merece ser tomada en cuenta sólo si se cuenta con evidencia atendible de que un factor semejante pudiera encontrarse presente. Por otra parte, ¿es posible respaldar empíricamente la atribución de eficiencia causal a PT de mejor manera que ésta? ¿Cómo? ¿Qué otra evidencia sería superior o más adecuada para ello?

Somos falibles y la certeza es inalcanzable. Hay que digerirlo. Pero algunos no lo hacen: es la búsqueda de certeza lo que hace que diversas nociones de causalidad se construyan en el marco de escenarios especiales, menos complejos que los reales, y en los cuales, por construcción, se asegura la inexistencia de todas aquellas circunstancias que podrían perturbar la aplicación pura del concepto.

4. Credibilidad y causalidad

Los modelos económicos convencionales contienen suposiciones irrealistas: agentes dotados de conocimiento perfecto, capacidades computacionales ilimitadas y conocimiento de los estados futuros, economías sin costos de transacción, etc. Muchos especialistas y estudiantes de economía se preguntan cuál es el nexo que tales construcciones teóricas pueden tener con aspectos de las economías reales (Mäki, 2009). Una manera de interpretar esta discusión es que lo que se halla en debate es la capacidad de modelos "irrealistas" para identificar las *causas* que contribuyen a generar los fenómenos económicos. Una vasta literatura aborda el problema de cómo pueden ser usados dichos modelos para entender y explicar aspectos salientes de las economías reales (Sugden, 2000, 2008; Cartwright, 2007 ; Alexandrova, 2008).

Sugden (2000) ofrece una perspectiva iluminadora de este asunto. Para ello hace distingue entre el "real world" (W), que está, por decirlo así, allá afuera, y el "model world" (M), el mundo de representaciones teoréticas que es considerado *paralelo* a nuestro mundo ordinario de eventos económicos.

"On this view, the model is not so much an abstraction from reality as a parallel reality. The model world is not constructed by starting with the real world and stripping out complicating factors: although the model world is *simpler* than the real world, the one is not a *simplification* of the other" (Sugden, 2000, p.25)

Pese a su mutua independencia, si algunas condiciones especiales son satisfechas sería legítimo extrapolar a W aquello que ha sido hallado en M. Supongamos que sabemos que en W se ha verificado la existencia de un patrón de eventos económicos R. Sugden piensa que el problema de cuál es la causa de R puede ser respondido construyendo un modelo M* en el que un mecanismo C postulado implique las características principales del patrón R. Esto constituye una típica explicación de la forma "qué haría posible R". La posibilidad aludida no es meramente *lógica*, porque según Sugden, si *ciertas condiciones* son satisfechas resultaría legítimo conjeturar que C describe el mecanismo *causal* que genera R en W. Más precisamente, Sugden sostiene que un modelo M* revela que podría haber un mecanismo C causando R en W en la medida en que M* (y el mecanismo que contiene) es *creíble*. Ofrece dos intentos principales por elucidar credibilidad. El primer enfoque hace uso de herramientas propias del análisis literario y funda la credibilidad en un tipo de conocimiento intuitivo ya disponible (i.e., pre – modélico).

"Credibility in models is, I think, rather like credibility in 'realistic' novels. In a realistic novel, the characters and locations are imaginary, but the author has to convince us that they are credible – that there could be people and places like those in the novel. …We judge the author has failed if we find a person acting out of character, or if we find an anachronism in a historical novel: these are things that *couldn´t* have happened" (Sugden, 2000, p. 25).

En un Segundo intento va más allá de la analogía entre modelos y novelas y ofrece dos criterios de credibilidad. El primero es coherencia *interna* (que es restringida a las relaciones entre los componentes del modelo). No debe ser entendida como mera consistencia lógica, cuya importancia se descuenta, sino en un sentido más "material", que refiere a sus contenidos.

"The assumptions of a good model (must) cohere in the broader sense that they fit naturally together". For instance, a model cannot mix hypothesis of "hyper-rationality" in one context and "bounded rationality" in another. This sort of coherence is needed for credibility: "If a model lacks coherence, its results cannot be seen to follow naturally from a clear conception of how the world might be; this prompts the suspicion that the assumptions have been cobbled together to generate predetermined results" (Sugden, 2000, p. 26).

El Segundo criterio es más ambicioso e incumbe a *la relación entre el mundo modelo y la realidad*. Sostiene que los modelos "must also cohere with what is known about causal processes in the real world. Thus, Akerlof's assumptions that prices tend to their market-clearing levels is justified by evidence from a wide range of 'natural' and laboratory markets. Schelling's assumptions that many people have at least mildly segregationist preferences is justified by psychological and sociological evidence, and coheres with common intuition and experience" (Sugden, 2000, p. 26).

La insistencia de Sugden en la independencia del mundo modelo respecto de los asuntos de la vida real y el énfasis puesto en su naturaleza "paralela" oscurece el hecho de que es precisamente el uso de conocimiento ya disponible (proveniente seguramente de un mix de observación, sentido común, experiencia de la vida y saberes propios de otras disciplinas sociales) lo que introduce credibilidad en un modelo. La clase de modelos a los que Sugden refiere como "creíbles"

adquieren este estatuto porque incorporan algunas características centrales de las economías y de los agentes *que sabemos que existen o pueden existir en los mercados reales*. Este saber no es resultado del modelaje, es un pre-requisito de su credibilidad.

La distinción entre modelos creíbles y no-creíbles es importante y merece ser tomada en cuenta y desarrollada. Destaquemos dos aspectos. En primer lugar, la importancia asignada a la credibilidad desafía el dictum de Friedman según el cual "theories should not be judged by the accuracy of their assumptions". Más aún, contradice explícitamente la tesis de que "truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense)". (Friedman, 1953, p. 14). En segundo lugar, su perspectiva coincide con la de muchos otros autores que asignan un papel central al *background knowledge* en la construcción de modelos. En particular, presupone que muchas causas de eventos económicos o sociales son conocidas mediante procedimientos pre-modélicos, y que es precisamente cuando los individuos o sus contextos son modelados invocando dichas causas que los modelos resultantes se tornan creíbles.

Sin embargo, Sugden parece limitar arbitrariamente los componentes del background knowledge, restringiéndolo a sentido común y conocimiento práctico. En mi opinión el éxito de PT muestra que es posible ampliar la noción de credibilidad (plausibilidad) de modelos económicos incorporando *conocimiento causal hallado por otras disciplinas sociales en condiciones experimentales*. En particular, PT proporciona recursos para anclar el modelaje en economía en conocimiento psicológico dotado de soporte experimental. Satisface las demandas de autores como Darrell and Maier-Rigaud (2012, p. 292) quienes criticaron a la teoría neoclásica "for its refusal to integrate social scientific research, especially from social psychology and sociology", y denunciaron el fracaso de la economía convencional "to design models that take into account key elements that drive economic outcomes in real-world markets. Half a century of research that conveniently disregarded essential institutional and behavioural characteristics of the markets...".

5. Explicación vía mecanismos

Otro de los blancos de la argumentación de Reiss (2007) es la pretensión de que para explicar se requiere la identificación de un mecanismo responsable del efecto. Hedstrom and Swedberg, Glennan y muchos otros han defendido una idea semejante[v].

Reiss sostiene que es verdad que si identificamos la causa actuante tenemos una explicación. Pero para identificar dicha causa no basta con manipular exitosamente los efectos, porque se presenta el problema de la causa común. Acordamos con Reiss en que reclamar un mecanismo no ayuda en nada a resolver *este* problema. Pero discrepamos con él en que para estar legitimados para aseverar conexión causal entre dos factores deberíamos primero exhibir un argumento concluyente que descarte la posibilidad de que se encuentre presente una causa común. ¿Por qué deberíamos estar obligados a proporcionar un argumento semejante en ausencia de una razón atendible para ello? Si no se exponen razones concretas para sospechar que existe una causa común actuando, no hay razones para dudar seriamente de que es el cambio de marco (frame) lo que –vía *loss aversion*- causa la reversión de preferencias.

Un problema diferente es el de si es necesario exhibir un mecanismo para aseverar que PT explica este fenómeno. No tenemos nada que decir *en general* respecto de la conexión entre explicación y mecanismos, pero sostenemos que en *este* caso la explicación tiene lugar sin que se conozca o necesite exhibir mecanismo alguno. No descartamos que tal mecanismo exista (y pensamos que sería valioso indagar si lo hay, y en ese caso en qué consistiría). Pero sugerimos que la necesidad de invocar un mecanismo se presenta en otro nivel, ante una pregunta diferente. Dicho con mayor precisión, proponemos distinguir entre las siguientes dos preguntas:

Pregunta 1: Por qué, dadas las condiciones C, la mayor parte del grupo G eligió A por sobre B (B por sobre A)?

Respuesta: debido a la descripción particular con que se le ofrecieron las opciones en cada caso.

Pregunta 2: Por qué describir las opciones A y B de manera X/Y hace que la mayor parte del grupo elija A cuando la descripción es X (y B cuando es Y)?

Respuesta 2: PT no lo dice. Se necesitaría mostrar un mecanismo (probablemente psicológico o neurobiológico)

En términos más generales, para explicar por qué la mayoría de un grupo de individuos, en condiciones C relevantes para PT, eligieron de manera X, basta con aludir al factor causal determinante (el "frame"). Para explicar por qué ese "frame" (esa manera de describir las opciones) hace que esa mayoría elija A y no B, se necesita exhibir un mecanismo. Más sucintamente, el fenómeno B se explica invocando la causa A (el *frame*, en nuestro caso), y la conexión (causal) entre A y B se explica invocando un mecanismo. Como dicen Ylikovsky y Hedstrom:

"A simple causal claim tells us about counterfactual dependency: It tells us
what would have happened if the cause had been different. The mechanism tells us why
the counterfactual dependency holds and ties the *relata* of the counterfactual to the knowledge
about entities and relations underlying it".

Que "causalidad" sea elucidada por los autores en términos de "dependencia contrafáctica" no afecta nuestro planteo. Lo central es que juzgan, acertadamente, que la conexión causal no se auto-explica, sino que necesita ser explicada en términos de mecanismos. Una explicación de este tipo *abre la caja negra*. Al igual que los enunciados legales, los enunciados causales son también de tipo caja negra. Por ello se necesita "relacionar el contrafáctico (i.e., el enunciado causal) al conocimiento de entidades y relaciones subyacentes".

6. Mecanismos como elucidación de causalidad y explicación

El concepto de "mecanismo" es objetado porque a pesar de que se lo introduce como un sustituto o un complemento de la noción de causa, no es sencillo definirlo de manera independiente (i.e.,

de modo que no presuponga causalidad). Glennan propone a los mecanismos como la manera de abrir la caja negra de las adjudicaciones causales (una manera de exhibir la "secret conection" a que aludía Hume). Desde esta perspectiva, "mecanismo" es una noción más básica que "causalidad". Sin embargo, si uno renuncia a la causalidad no es sencillo especificar qué tipo de conexión existe entre las partes de un mecanismo. Pero si se desea distinguir a los mecanismos de las meras secuencias de eventos, se necesita que exista algún tipo de conexión, "generativa" o "productiva", por llamarla de alguna manera, que sea la responsable de la producción del efecto. Si esta conexión entre las partes es concebida como causal, reaparece la "secret conection", y si no lo es debe elucidársela (una tarea que hasta donde sé sigue pendiente).

Lo dicho sugiere que el concepto de mecanismo está débilmente construido: o deja sin precisar el tipo particular de conexión que existe entre sus diversas partes, o es causa-dependiente: esas conexiones son elucidadas en términos de causa y efecto. En ambos casos la elucidación fracasa: o nos quedamos con las manos vacías o incurrimos en circularidad. Probablemente por ello dice Cartwright que la noción de "causa" es endémica. Quiere decir que es ineliminable (irreducible a otra cosa). Si tiene razón estamos condenados a la circularidad: todo lo que se puede hacer es reducir ciertas relaciones causales a otras relaciones causales más básicas (Bunge, 2004).

Pero supongamos que el concepto de mecanismo fuera suficientemente claro e independiente del concepto de causalidad, y que en principio pudiéramos describir en sus términos la secuencia de eventos que conduce desde la causa a su efecto. Ello exhibiría la "secret conection" aludida por Hume y resolvería el problema de la causalidad. Es decir, abriría la caja negra. Sin embargo, aunque siempre es *deseable* contar con una descripción mecanísmica de una relación causal, *no* siempre es *razonable* exigirla. Glennan tiene razón cuando expresa que el

"analysis of causal connections in terms of mechanisms is only meaningful when there are ways (even if indirect) of acquiring knowledge of their parts and the interactions between them" (Glenann, 1996).

Dado que, según nuestro conocimiento, ni la psicología ni la neurobiología han proporcionado nada semejante, parece razonable no exigir la invocación de un mecanismo en el caso de los framing effects. Creemos que la evidencia favorable con que éstos cuentan es suficiente para asignar papel causal al frame.

Pero Glenann (1996) no se contenta con postular que "causalidad" debe ser elucidada en términos de "mecanismo", sino que impone una condición adicional muy fuerte: sostiene que una conexión entre eventos sea considerada causal cuando es decidida *científicamente* a través de la postulación de un mecanismo. No se trata pues simplemente de postular la acción de un mecanismo, sino que tal pretensión debe estar fundad en *conocimiento especial* (*científico*, se lo conciba del modo que sea).

Hay que notar que esta exigencia implica que no podríamos hablar de causalidad no sólo en el caso de los framing effects, sino en muchos otros en que naturalmente diríamos que disponemos

de una explicación causal de los sucesos. Recordar la ilustración paradigmática de Kincaid en ocasión de su visita al CIECE en 2014: se tira una pelota de *beisball* contra un vidrio y éste se rompe. Lo hacemos una, dos, n veces y sucede lo mismo. Pareciera que tenemos una conexión causal. ¿Carecemos de ella si no disponemos de conocimiento del mecanismo del mecanismo físico que culmina en la rotura del vidrio? Ello haría que la atribución de causalidad fuera dependiente del conocimiento científico. ¿Es necesario? ¿Es razonable? Podría argumentarse que no. De otra manera el lenguaje causal debería erradicarse de los asuntos de la vida diaria. Es más, los modelos creíbles de Suden, a los que referimos anteriormente, fracasarían en captar factores causales. Y en idéntica situación se hallaría buena parte de la evidencia causal empleada en el ámbito jurídico, ya que tampoco es de carácter científico.

Que las asignaciones causales disponibles puedan ser refinadas en términos más precisos y bajo condiciones más rigurosas es un objetivo deseable. Pero esto no implica que se les niegue carácter causal, pues significaría que sólo estaríamos legitimados a predicar causalidad de relaciones entre entidades que consideramos irreductibles a términos de otras entidades. La causalidad se volvería una rareza en este mundo, pues nada garantiza que las teorías científicas actuales, aún las mejor confirmadas, no son precisables y mejorables en términos aún más específicos.

Por otra parte, disponer de conocimiento científico está muy bien, pero no debe perderse de vista que es más bien una excepción que la regla. Más que exigirlo como condición necesaria para atribuir causalidad se lo debería considerar como un plus: como un medio para enriquecer el conocimiento de relaciones causales ya identificadas. A mi juicio Glennan peca en este punto de una buena dosis de cientificismo. Del mismo modo, parecería excesivo que, al igual que en el caso de la causalidad, todas las nociones epistémicas centrales (predicción, explicación, comprensión, etc.), solo valieran en tanto y en cuanto se usen en el marco del lenguaje especial de alguna de las ciencias particulares. ¿Debe el historiador, el experto o aún el hombre corriente renunciar al uso de estos conceptos en su *metier*? La exigencia de Glennan retrotrae a las prescripciones de Hempel (1942) para la historia.

Por último, cabe advertir que si se aceptara la exigencia de Glennan, los modelos económicos construidos sobre la base de PT estarían en mejores condiciones que los modelos creíbles a la Sugden, ya que aquellos incorporan factores causales del comportamiento descubiertos por la psicología cognitiva en situaciones de control experimental.

Conclusiones

Se han sostenido que dadas ciertas condiciones, Prospect Theory identifica un factor causal determinante del fenómeno de reversión de preferencias y de las decisiones que adopta en condiciones de incertidumbre la mayoría de los individuos. Sobre la base de este conocimiento, PT permite predecir, explicar y manipular (controlar) este amplio rango de decisiones. Puede hacer esto sin necesidad de invocar o identificar mecanismo alguno. Esta capacidad de PT permite ampliar la noción de modelos creíbles (plausibles) de Robert Sugden, incorporando factores causales del comportamiento que no pertenecen al conocimiento ordinario o al conocimiento

experto, fundado en la práctica, sino que ha sido adquirido científicamente, mediante la contribución de otras disciplinas sociales en situaciones de control experimental.

Bibliografía

Alexandrova, A., (2008), Making Models Count, Philosophy of Science, 75 (July 2008) pp. 383–404.

Benartzi, Sh. and Thaler, R., Naive Diversification Strategies in Defined Contribution Saving Plans, *The American Economic Review*, Vol. 91, No. 1. (Mar., 2001), pp. 79-98.

Benartzi, Sh. and Thaler, R., "Heuristics and Biases in Retirement Savings Behavior", Journal of Economic Perspectives, Vol. 21, nº3, Summer 2007, pp. 81-104.

Bunge (2004), 'How does it work? The search for explanatory mechanisms', Philosophy of the Social Sciences, 34 (2), 182–210.

Cartwright, N. (2007), *Hunting Causes and Using Them –Approaches in Philosophy and Economics*, Cambridge, Cambridge University Press.

Darrell, Arnold, and Maier-Rigaud, Frank P., The Enduring Relevance of the Model Platonism Critique for Economics and Public Policy, *Journal of Institutional Economics, 2012, Vol. 8(3), 289-294.*

Friedman, M., (1953), "The Methodology of Positive Economics". In: Essays in Positive Economics, Chicago: University of Chicago Press.

Glennan, S. (1996), 'Mechanisms and the nature of causation', Erkenntnis, 44,
(1), pp. 49–71.

(2008), Mechanisms. In S. Psillos and M. Curd, eds. *The Routledge Companion to Philosophy of Science*. Abingdon: Routledge, pp. 376-384.

Guerring, John, (2010), Causal Mechanisms: Yes, But…., *Comparative Political Studies 43(11)*

Hedström, Peter and Richard Swedberg (1998a), Social Mechanisms. An Analytical Approach to Social Theory, Cambridge, UK: Cambridge University Press.

Kahneman, D. (2003a), "Maps of Bounded Rationality: Psychology for Behavioral Economics", AER, Vol. 93, Nº 5, pp.1449–1475.

_____(2003b), "A psychological Perspective on Economics", AER, Vol. 93, Nº 2, pp.162-168.

Hedstrom, P., and Ylikosky, P., Causal Mechanisms in the Social Sciences, Annual Review of Sociology, 2010.

Kahneman, D. and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision Under Risk". Econometrica, 47, pp.263-291.

Kahneman, D., Knetsch, J.L. and Thaler, R.H. (1991), "The Endowment Effect, Loss Aversion, and Status Quo Bias". Journal of Economic Literature, Volume 5, Nº 1, pp.193–206.

Kincaid (2004). 'Are There Laws in the Social Sciences? Yes', in C. Hitchcock, Contemporary Debates in Philosophy of Science. Oxford: Blackwell, 168–187.

Knetsch, J.L., (1989), "The Endowment Effect and Evidence of Nonreversible Indifference Curves". The American Economic Review, vol. 79, nº 5, pp.1277–1284.

Mäki, U., (2009), "Realistic realism about unrealistic models", en The Oxford handbook in the philosophy of economics, ed. by Kincaid and Ross, Oxford University Press.

Marqués, G., Weisman, D., (2015), "A Criterion for realism, with an application to behavioral economics models", The Journal of Philosophical Economics, Volume IX Issue 1, (Autumn 2015). ISSN 1843-2298.

Psillos, S., (2004), A Glimpse of the *Secret Connexion:* Harmonizing Mechanisms with Counterfactuals.

Sugden, R. (2000) "Credible Worlds: the status of theoretical models in economics" *Journal of Economic Methodology* 7/1: 1-31.

_____ (2008): Credible worlds, capacities and mechanisms.

Thaler, Richard (1980). "Toward a positive theory of consumer choice". *Journal of Economic Behavior & Organization* **1** (1): 39–60.

Tversky, A. and Kahneman, D. (1981), "The Framing of Decisions and the Psychology of Choice". Science, New Series, Vol. 211, Nº 4481, pp.453–458.

_____(1986), "Rational Choice and the Framing of Decisions". The Journal of Business, Vol. 59, Nº 4, Part 2, pp.S251–S278.

[i] Como se verá, la noción de causalidad apropiada para PT intenta ser más fuerte, debido a que tiene que recoger como parte de su contenido que el cambio de marco (frame) introducido exógenamente es determinante (suficiente) para el cambio de decisión. Pero entrampada en la polaridad condición necesaria-suficiente no parece ser sencillo dilucidar en qué consiste (realmente) la conexión causal. Pero podemos dejar a un lado este problema "esencialista", y abordar las otras dos cuestiones, más propiamente epistemológicas.

[ii] La capacidad de predecir correctamente no necesita remitir a otras nociones. Se basa o funda sólo en la obtención de los resultados anticipados. Es una capacidad más básica o más fundante que la capacidad de explicar o contar con conocimiento causal.

[iii] PT affords a quantitative measure of loss aversion, which ordinarily ranges between 2, 25 and 2, 50.

[iv] Nuestro análisis satisface las exigencias del *empirismo*: la identificación de la conexión causal se basa en evidencia empírica. En particular, se discrepa de la tesis de que la "isolation" en el marco de un modelo puede ser suficiente para confirmar causal claims. La confirmación de que L es un "causal claim" es (debe ser) empírica.

[v] "The basic idea of a mechanism-based explanation is quite simple: At its core, it implies that proper explanations should detail the cogs and wheels of the causal process through which the outcome to be explained was brought about" (Hedstrom and Ylikosky, 2010).