frontiers | Frontiers in Psychology

TYPE Methods
PUBLISHED 08 November 2022
DOI 10.3389/fpsyg.2022.724498

Check for updates

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University, Singapore

REVIEWED BY
Antonio Prieto,
National University of Distance Education (UNED), Spain
José Manuel Reales,
National University of Distance Education (UNED), Spain

*CORRESPONDENCE
Oliver Y. Chén
olivery.chen@bristol.ac.uk

SPECIALTY SECTION
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

RECEIVED 13 June 2021
ACCEPTED 13 September 2022
PUBLISHED 08 November 2022

CITATION
Chén OY, Phan H, Cao H, Qian T, Nagels G and de Vos M (2022) Probing potential priming: Defining, quantifying, and testing the causal priming effect using the potential outcomes framework. *Front. Psychol.* 13:724498. doi: 10.3389/fpsyg.2022.724498

COPYRIGHT
© 2022 Chén, Phan, Cao, Qian, Nagels and de Vos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Probing potential priming: Defining, quantifying, and testing the causal priming effect using the potential outcomes framework

Oliver Y. Chén[1]*, Huy Phan[2,3], Hengyi Cao[4,5,6], Tianchen Qian[7], Guy Nagels[8,9] and Maarten de Vos[10,11,12]

[1]Faculty of Social Sciences and Law, University of Bristol, Bristol, United Kingdom, [2]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom, [3]The Alan Turing Institute, London, United Kingdom, [4]Department of Psychology, Yale University, New Haven, CT, United States, [5]Center for Psychiatric Neuroscience, Feinstein Institutes for Medical Research, Manhasset, NY, United States, [6]Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY, United States, [7]Donald Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA, United States, [8]Department of Neurology, Universitair Ziekenhuis Brussel, Jette, Belgium, [9]Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom, [10]Faculty of Engineering Science, KU Leuven, Leuven, Belgium, [11]Faculty of Medicine, KU Leuven, Leuven, Belgium, [12]KU Leuven Institute for Artificial Intelligence, Leuven, Belgium

Having previously seen an item helps uncover the item another time, given a perceptual or cognitive cue. Oftentimes, however, it may be difficult to quantify or test the existence and size of a perceptual or cognitive effect, in general, and a priming effect, in particular. This is because to examine the existence of and quantify the effect, one needs to compare two outcomes: the outcome had one previously seen the item vs. the outcome had one not seen the item. But only one of the two outcomes is observable. Here, we argue that the potential outcomes framework is useful to define, quantify, and test the causal priming effect. To demonstrate its efficacy, we apply the framework to study the priming effect using data from a between-subjects study involving English word identification. In addition, we show that what has been used intuitively by experimentalists to assess the priming effect in the past has a sound mathematical foundation. Finally, we examine the links between the proposed method in studying priming and the multinomial processing tree (MPT) model, and how to extend the method to study experimental paradigms involving exclusion and inclusion instructional conditions.

KEYWORDS

priming effect, causal inference, potential outcomes framework, word fragment completion test, significant test, between-subjects study

Introduction

Imagine you are asked to fill in a fraction of a word, say _aze_ _e. Suppose the target word is gazette. What would your performance be if you have seen a list of words including gazette before the game? Intuitively, seeing a list of words containing the target answer improves one's performance. But, how could we formally test whether the improvement exists, and how could we quantify the amount of improvement?

More specifically, we call such a phenomenon where having been exposed to an item (e.g., viewing a word or an object) facilitates the subsequent recovery of the item based on a partial or reduced perceptual cue (e.g., viewing a partial word or a fragment of an object) *repetition priming* (Hayman and Tulving, 1989; Tulving and Schacter, 1990, 1992).

Neurobiologically, this (priming) effect on word identification is facilitated and carried out through activations in the brain involving memory and learning. Although the exact neural bases of priming are as of yet little known, several lines of evidence have hinted that priming is mediated by neural systems outside of the medial temporal and diencephalic regions (Tulving and Schacter, 1990), and that priming is related to changes in cortical modules that are involved in processing specific attributes of stimulus information (Squire, 1987). Neuropsychologically, posterior cortical areas in the right hemisphere seem to be associated with object identification (Warrington and Taylor, 1978); passive reading of familiar words produces selective bilateral activation in the extrastriate cortex, suggesting that visual identification (not necessarily understanding) of words has an anterior-occipital locus (Schwartz et al., 1980; Funnell, 1983; Satori and Job, 1987; Petersen et al., 1988).

Yet, despite neurobiological and neuropsychological advances, little do we know about how to formally test the existence of a priming effect, whether the effect is causal and if so, how to quantify it. The difficulty, in part, lies in the need to compare two scenarios where only one is observable. Specifically, to claim that there exists a priming effect (e.g., the effect of a word study on word identification), one must first quantify the outcomes of two scenarios (e.g., word identification accuracy after viewing the target words vs. the accuracy without viewing the target words) and then compare these two outcomes to draw a (statistical) conclusion. But, only one of these[1] is observable on each individual. How, then, could we compare an observable outcome with an unobservable one?

Here, linking Neyman and Rubin's works on causal inference and Tulving and Schacter's earlier works on priming, we aim to define, test, and quantify the causal priming effect using the *potential outcomes* framework (Neyman, 1923; Rubin, 1974, 1977, 1978). We demonstrate how to use this framework to study the priming effect by analyzing data from a between-subjects study (Hayman and Tulving, 1989). We also show that what has been previously used intuitively to study the priming effect has a sound mathematical foundation. But before we proceed, it is perhaps useful to discuss the reasons for choosing this framework, the relationship between priming and memory, and the convenience of studying priming using a word fragment completion test.

A brief introduction of causal inference

Let us begin by briefly introducing and comparing three useful approaches to study causation: Campbell's ("validity testing") framework, Pearl's (causal diagram) framework, and the Neyman–Rubin's (potential outcomes) framework.

Campbell's framework focuses on evaluating the validity of standard designs for experimentation in the social sciences and finding extraneous variables that may confound causal interpretations (Campbell, 1957).

The Neyman–Rubin framework focuses on the *magnitude* of the causal effect; it emphasizes the mathematical argument that can yield an analytical estimate of the causal effect. As only one of the two outcomes in the Neyman–Rubin framework can be observed from each individual, they are usually referred to as *potential outcomes* (Neyman, 1923; Rubin, 1974, 1977, 1978).

Pearl's framework introduces *directed* graphs into causal analysis, with nodes indicating variables (e.g., exposure and outcome) and edges indicating causal links (Pearl, 1993, 1995, 2001, 2009a). In addition, the $do(\cdot)$ operator[2] and the back-door and front-door criteria make some otherwise difficult causal effects identifiable (see later).

A comparison between Campbell's, Neyman–Rubin's, and Pearl's causal models

Similarities

Most psychologists are familiar with Campbell's method; perhaps few have had exposure to the Neyman–Rubin model (West and Thoemmes, 2010). In our view, however, Design 6 [3] in Campbell (1957) shows spirit of both Neyman–Rubin's

1 Either viewing the target words or not viewing them; one cannot unview the words that one had viewed.

2 Here, $do(X = x)$ means the model forces X to take the value of x; in other words, one sets X (*via* intervention) to be a constant value x.

3 Consider two experiments: $A\ X\ O_1$ *vs.* $A\ [\ ]\ O_2$, where A, X (or lack thereof), and O are ordered from left to right in time, and A, X, and O indicate a random sampling assignment, a treatment, and the outcome, respectively. In Campbell's approach, the presence of X on the left of O_1 means O_1 is the outcome of a group after receiving a treatment X, and the absence of X (*i.e.*, blank space) on the left of O_2 means O_2 is the outcome of another group without receiving a treatment.

potential outcomes framework[4] and Pearl's causal diagram[5]. As for Pearl's and Neyman–Rubin models, oftentimes, they are mathematically equivalent[6] (see Section 7.4.4 of Pearl, 2009b).

Differences

Compared with Campbell's approach, the Neyman–Rubin framework offers an analytical language for identifying and quantifying the causal effect. Compared with Neyman–Rubin formulations, Pearl's method is oftentimes easier for social scientists to understand and visualize the causal problems using vivid graphic representations. In certain cases[7], controlling for covariates using the Neyman–Rubin method may fail to identify a causal effect – a major criticism from the Pearl school. Furthermore, under the potential outcomes framework, there is a subtle difference between Neyman's null (where the null hypothesis considers zero average causal effect) and Fisher's null (where the null hypothesis considers zero individual causal effect) for many realistic situations, which may cause confusions (Ding, 2017). The Rubin school argues[8] that causation, especially causation involving directed causation and dynamic causation, cannot be simply explained by graphs. Pearl's method assumes that the $do(\cdot)$ operator itself does not perturb the (causal) system, about which some may caste doubts; in addition, oftentimes this assumption cannot be tested. For experimentalists, it is sometimes impractical to apply the $do(\cdot)$ operator to intervene certain variables such as gender and age. Finally, in practice, it may be difficult to obtain a complete picture of the causal diagram (e.g., the directed causal map of the brain network).

Weighing pros and cons and in light of priming research, in this article, we derive the potential causal priming framework in Neyman–Rubin's language and accompany graphs in Pearl's style to visualize causal relationships (see the Discussion section for future directions).

Remark 1. We encourage interested readers to compare, in detail, the potential outcomes framework with Campbell's framework (e.g., West and Thoemmes, 2010) and the potential outcomes framework with Pearl's framework [e.g., Gelman's blog post (Gelman, 2009) and Pearl's response under the post].

Remark 2. There are other fine works on causal inference; we refer our readers to them for further reading (Peters, 1941; Cochran and Chambers, 1965; Hill, 1965; Goldberger, 1972; Ding et al., 2016).

A brief discussion of memory

Different memory systems

Whereas the focus of the article is on priming, it is perhaps beneficial to familiarize oneself with the memory systems, in general. This is because on the one hand, priming is related to memory, and on the other hand, it is arguably independent of explicit and semantic memory (Tulving and Schacter, 1990). By stating explicit and semantic memory, one has already implied there exists some categorization of memory systems. Although we do not intend to and cannot fully examine the hypothesis regarding the number of memory systems present, a summary of a few key classifications of memory systems may help the readers to deal with priming conceptually. Tulving (1985) argued that there exist three types of memory systems: episodic (associated with self-knowing consciousness), semantic (associated with knowing consciousness), and procedural (associated with non-knowing consciousness). Cohen and Squire (1980) and Mishkin et al. (1984) argued that there are two types of memory systems: the former coined the two systems according to the concepts of "knowing how" and "knowing that," and the latter distinguished the habit system from the "memory" system. Others have proposed more specific classifications, arranged either hierarchically (Pribram, 1984)[9] or interactively without a fixed relationship to each other (Johnson, 1983). More specifically to priming, it is hypothesized that there exists a pre-semantic perceptual system [called the *perceptual representation system* (PRS)] that manages priming; the PRS operates independently of the explicit and semantic memory (Tulving and Schacter, 1990). In brief, the hypothesis of the PRS suggests that there is a dissociation between priming and explicit memory and that there is a dissociation between (pre-semantic) priming and semantic memory (Warrington and Taylor, 1978; Parker et al., 1983; Graf et al., 1984; Hashtroudi et al., 1984; Cermak et al., 1985; Light et al., 1986; Shimamura, 1986; Nissen et al., 1987; Kopelman and Corn, 1988; Parkin and Streete, 1988; Tulving and Schacter, 1990).

Process dissociation model

Interposed between the classification of multiple memory systems and the study of priming is the need to separate the latter from other, for example, semantic and explicit

4 The letter *A* in Campbell's approach is equivalent to the randomization and matching mechanism in Neyman-Rubin's approach, or in Campbell's words: "*A* is the point of selection, the point of allocation of individuals to groups . . . At time *A* the groups were equal, even if not measured. . ."

5 Pearl's circles and arrows are equivalent, in spirit, to Campbell's letters and orders: in Campbell's notation, if *X* is placed on the left of *O*, it implies there is a directed arrow from *X* to *O*.

6 Namely, $P\{Y|do(Z) = z\} = P\{Y(z)\}$, where $P\{Y|do(Z) = z\}$ (in Pearl's language) means forcing Z to take the value z by removing all father nodes of Z, and $P\{Y(z)\}$ (in Neyman-Rubin's language) means the potential outcome of Y under z.

7 Suppose (1) U and W both cause X, (2) U and W cause T and Y, respectively, and (3) T causes Y. Using Pearl's method, the relationship from T to Y is causal. But using Neyman-Rubin's method, by controlling for X, the relationship from T to Y is not causal.

8 We attribute a part of the summary between Neyman-Rubin's and Pearl's methods to works from Peng Ding.

9 The discussion was on primates.

memory processes. This need is partly sprawled empirically from findings where patients with amnesia reported significantly worse explicit memory (intentional use of memory) than normal subjects but showed as large a priming effect (an arguably automatic, passive use of memory) as the normal subjects (Warrington and Weiskrantz, 1974; Graf et al., 1984; Cermak et al., 1985; see Shimamura, 1986 for a review). Practically, to separate and estimate the contribution of unconscious, automatic, controlled, and intentional processes, Jacoby (1991) proposed the *process dissociation framework* and argued its utility in studying perception, memory, and thought. The key point of the framework is to use regression models to separate the effect of (consciously controlled) recollection from that of (automatic) familiarity [see Experiment 3 in Jacoby (1991) for details].

The role of memory in encoding instructions

Participants in a priming study need to follow instructions. Working memory, the ability to maintain and process information (Baddeley and Hitch, 1974), plays an important role in encoding both spoken (Baddeley et al., 1984; Hanley and Broadbent, 1987) and written (Wright, 1978; Wright and Wilcox, 1978) instructions. Cognitive load (including intrinsic, extraneous, and germane loads) that connects instructional design to cognitive functions is related to working memory. The cognitive load consumes a part of the working memory, and particularly, with appropriate instructional design, the germane load positively affects learning (Cooper et al., 2001).

A brief introduction to the word fragment completion (WFC) test

The word fragment completion (WFC) test is widely used to assess priming. In general, the test consists of a study phase and a test phase. During the study phase, subjects are instructed to view a list of words, including target words (e.g., gazette) and non-target words (called buffers). The test phase starts after an interval (e.g., 2 h). During this phase, the subjects are randomly assigned into two groups; each group undertakes one of the following tasks: (1) uncovering studied words (e.g., gazette) given a cue and (2) uncovering non-studied words given a cue. Some fragment completion tests will include an additional test, which involves repeating the word identification of gazette either with the same cue or with a different cue (see examples in the Results section). In simple terms, priming is said to have occurred when the success rate of cue-based item identification after studying the item is higher than that of a non-studied item (see **Figure 1**).

Under the Neyman–Rubin's potential outcomes framework, the priming effect of receiving a word study, which consists of the target words (the exposure of interest), on word identification (the outcome) can be defined as follows:

> *It is the difference between the two potential outcomes: the first is the outcome had an individual received the word study which consists of the target words, and the second is the outcome had the same individual not received the word study (or received a word study which did not contain the target words).*

We restrict our focus on the priming effect during a non-semantic word completion test, although the framework can be extended to studying semantic tasks such as rating the pleasantness when viewing a word and giving its definition. This is, in part, because priming is not affected by semantic and non-semantic encoding (Tulving and Schacter, 1990). Similarly, as priming occurs in more complex studies such as visual object recognition (Schacter et al., 1990), the framework introduced in this article may also be useful to quantify these priming effects. Although we focus on modeling the causal effect in studies of implicit memory, it may shed some light on studies of explicit memory (see the Discussion section). Finally, we note that when the instructions were not implicit but explicit during a fragment completion test, the test should be, in spirit, considered more as a "cued recall test" than a "fragment completion test."

Method

Notations and definitions

We begin by defining the notations used throughout this article. We use Z to denote whether a word study concerning viewing a list of words (including target words, such as gazette, which we use as an example throughout this article, and non-target words, such as *vermouth*) is undertaken at time $t_0 = 0$ (see **Figure 1**). Specifically, $Z = 1$ means that a subject has undertaken a word study including the target words (and henceforth referred to as having undertaken a word study for simplicity), and $Z = 0$ means that the subject has not undertaken the word study (or have undertaken a word study with all non-target words, which, for simplicity, we will henceforth refer to as not having undertaken a word study). In this study, we consider that Z takes binary values (i.e., having vs. not having conducted a word study), although our approach can be extended to categorical Z that takes more than two values (e.g., word studies consisting of words with low, intermediate, and advanced level of complexity). The word complexity can be quantified by, for example, evaluating the combination of syllable shapes and word patterns. As such, a further extension of Z can take any value between 0 and 100 to indicate complexity of each word (see the Discussion section for continuous and time-dependent cases).

Let X_1 denote a cue (e.g., $X_1 =$ _aze_ _e) given during a WFC test at a time t_1 ($t_1 = t_0$, typically t_1 is 2 h after t_0) (see **Figure 1**). We write $Y(X = x, Z = z)$ as the outcome of

10.3389/fpsyg.2022.724498

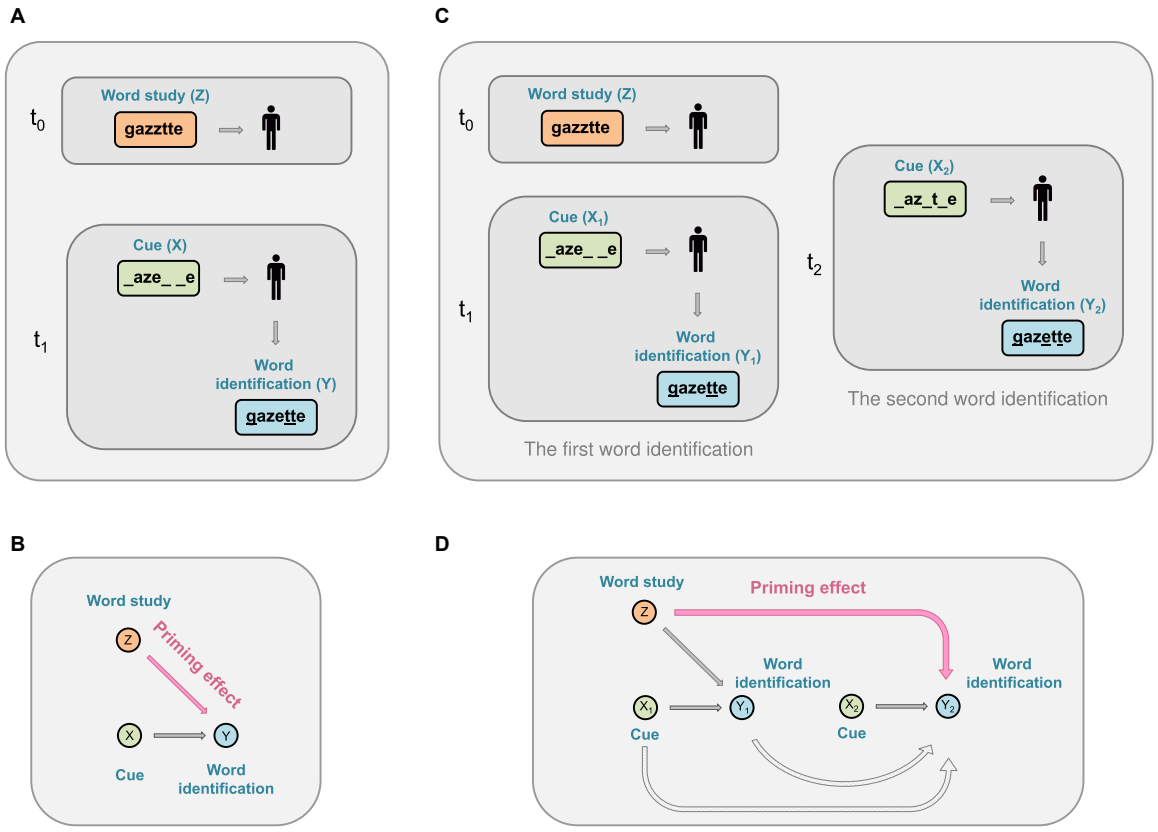


FIGURE 1
A schematic representation of the priming effect during a word completion experiment. **(A)** First, a subject views a list of words (including a target word gazette, or Z) during a word study. Next, a cue that consists of a fragment of the word (e.g., _aze_ _e, or X) is given to the subject. The subject is then asked to fill in the blanks. The subject may successfully uncover the target word or another word gazelle; or report a non-word gazeele, or an incomplete entry such as _azelle. **(B)** Priming effect refers to the phenomenon that being previously exposed to the word gazette (Z) primes (i.e., facilitates) the identification of the word (Y) given a partial or reduced cue (X). **(C)** A word study followed by two-word identifications. Left: The same word study and the subsequent cue-based identification, as in **(A)**. Right: An additional cue-based identification. In the figure, a different cue (e.g., _az_t_e, or X_2) is used during the second word identification for illustration purposes; two identical cues can also be used. **(D)** Priming effect where being previously exposed to the word gazette (Z) primes (i.e., facilitates) the identification of a word (Y_2) during a second word identification test given a partial or reduced cue _az_t_e (X_2).

word identification on the experiment unit (i.e., an individual participant), given that the unit received a word study $Z = z$ at t_0 and a cue $X = x$, where the upper case indicates a random variable and the lower case refers to its realized value. If there is an additional test, let X_2 denote the cue (e.g., X_2 = _az_t_e) given during the second word completion test at time t_2, where t_2 can be, for example, 2 h after t_1. By design, we have $t_2 > t_1 > t_0$. Between t_1 and t_2, participants can undertake tasks irrelevant to the experiment, such as taking a cognitive psychology class. We define Y_1 and Y_2 as the corresponding word identification outcomes given cues X_1 and X_2, respectively.

In the following, we always assume that each cue X corresponds to a single answer, and we drop the notational dependence of Y on the target word where there is no confusion. For example, $Y(X =$ _aze_ _e, $Z = 1) = 1$ means that the word identification is correct (e.g., the identified word is gazette, the target word, or gazelle, another correct answer[10]) given the cue X = _aze_ _e, after a word study Z consisting of a target word gazette. $Y(X =$ _aze_ _e, $Z = 1) = 0$ means that the word identification is incorrect given the cue $X =$ _aze_ _e, after a word study Z consisting of the target word gazette. Similarly, $Y(X =$ _aze_ _e, $Z = 0) = 1$ means that the word identification is correct (i.e., the identified word is gazette or gazelle) given the cue $X =$ _aze_ _e, had a word study Z not been conducted. $Y(X =$ _aze_ _e, $Z = 0) = 0$ means that the word identification is incorrect given the cue $X =$ _aze_ _e, had a word study Z not been conducted.

Definition 1.1 (Causal priming effect)

The causal priming effect on an experiment unit (i.e., a subject) given a cue $X = x$ is defined as

10 This indicates the study phase did not affect the WFC.

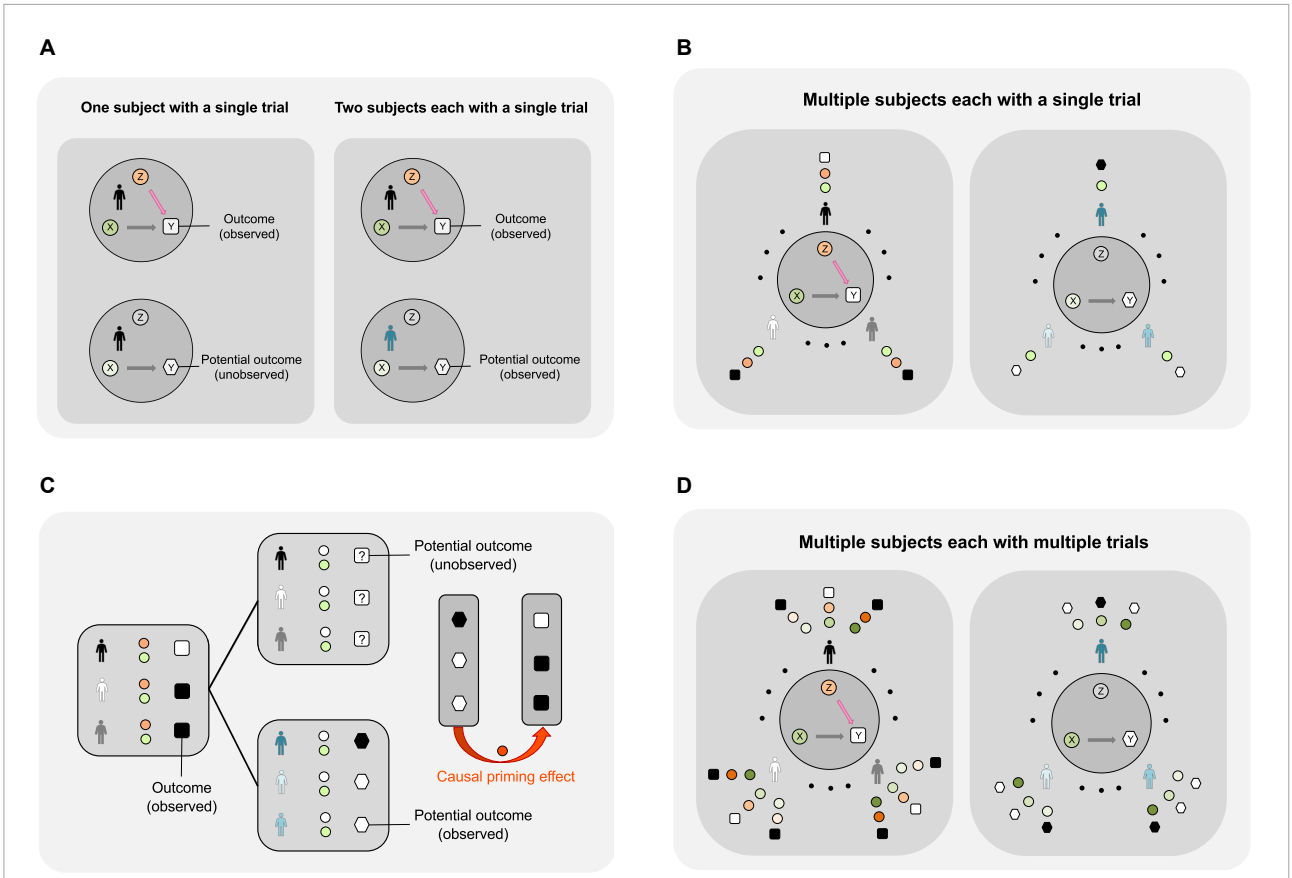


FIGURE 2
Using the potential outcomes framework to study the causal priming effect during a word completion experiment. **(A)** Identifying the causal priming effect with one trial. Top left: The figure describes a scenario where the subject first conducts a word study ($Z = 1$) and then aims to identify a target word based on a partial cue (X). The outcome (or Y) is observable. Bottom left: The figure describes a scenario where a subject does not conduct a word study ($Z = 0$) and aims to identify a target word based on a partial cue (X). If the subject has already participated in a word study, this (potential) outcome (or Y) under no word study is not observable. Top right: The same experiment as in the top left. *Bottom right*: Since the potential outcome of the subject (indicated by a human icon in black) is not observable, a different subject (indicated by a human icon in blue) is asked to identify the same target word based on the same partial cue (X), without a word study ($Z = 0$). If the two subjects are similar, then the causal priming effect is estimated by the difference between the outcomes (denoted by a letter Y in a square and a letter Y in a hexagon). **(B)** To reduce the possibility that a particular word may yield various priming effects on different subjects, multiple individuals are randomized to either conduct a word study ($Z = 1$) or not ($Z = 0$). The individuals with grayscale colors receive a word study; the individuals with bright colors do not receive a word study. For the left image, green circles represent various cues, orange circles represent word studies ($Z = 1$), and black or white squares represent the word study outcomes, where a black square indicates success ($Y = 1$) and a white square indicates failure ($Y = 0$). For the right image, green circles represent the cues corresponding to those on the left image, gray circles with the letter Z refer to having not conducted a word study ($Z = 0$), and black or white hexagons represent the word study outcomes, where a black hexagon indicates success ($Y = 1$) and a white hexagon indicates failure ($Y = 0$). **(C)** Average priming effect of a word study on word identification among multiple subjects using one trial. Left: Half of the subjects are randomized to perform the word identification experiments after a word study (indicated by orange circles). The black and white squares indicate the observed outcomes. Top middle: The potential outcomes of the same subjects (who have participated in a word study) had they not participated in a word study (indicated by blank circles). The squares with question marks indicate that these outcomes are not observable. Bottom middle: The remaining subjects perform the word identification experiments without a word study; the outcomes are observable and are indicated by black and white hexagons for successful and failed cases, respectively. Right: Due to randomization, the difference in the outcomes between the two groups (divided by the sample size) gives an estimate of the average priming effect of the word study on word identification in a sample. **(D)** The experiment can be further extended to multiple subjects with multiple words.

$\mathrm{Y}(X = \mathrm{x}, Z = 1) - \mathrm{Y}(X = \mathrm{x}, Z = 0)$; this quantifies the difference between the outcome Y from a study unit that has conducted a word study ($Z = 1$) versus the outcome Y from the same unit had no word study been conducted ($Z = 0$), given the same cue $X = \mathrm{x}$.

Only one of the two potential outcomes can be observed from each subject. In other words, the individual-level causal priming effect is non-identifiable. Therefore, a natural inquiry into the causal priming effect is to uncover the average priming effect across multiple subjects (see Figures 2A,B).

Definition 1.2 (Average priming effect of a single studied word)

Consider one target word in a wordlist that is viewed by a total of N subjects. Then, the *average priming effect* (APE) of a word study Z on a word identification Y given a cue X is defined as follows:

$$APE_N^{\{1\}} = \frac{1}{N}\left\{\sum_{i=1}^{N} Y_i\,(X = x,\ Z = 1) - Y_i\,(X = x,\ Z = 0)\right\}$$

where $\mathrm{Y}_i\,(X = \mathrm{x}, Z = \mathrm{z})$ indicates the word identification result of the i^{th} subject after given a cue $X = x$ and a word study Z ($Z = 1$ means after viewing a wordlist consisting of target words and $Z = 0$ means without a word study). The superscript $\{1\}$ indicates that it is the average priming effect for one target word, and the subscript N indicates that the effect is defined on N individuals.

Unfortunately, we cannot observe both $\mathrm{Y}\,(X = \mathrm{x}, Z = 1)$ and $\mathrm{Y}(X = \mathrm{x}, Z = 0)$ on the same subject. This is because after having assigned (or not assigned) a word study Z (e.g., $Z = 1$) and the word identification test result Y has been reported, we cannot go back to the time t_0 to assign a different Z (i.e., $Z = 0$). Certainly, one could experiment on the same unit in two trials (one with the word study $Z = 1$ and the other with $Z = 0$), which consists of a repeated-measures study (e.g., Challis and Brodbeck, 1992). The first study, however, may have a carryover or learning effect on the second. Therefore, we cannot ascertain that the priming effect is due to the word study Z or the information learned (e.g., the cue X, the word identification Y, or the study mechanism) from the first test (see the Discussion section for details).

Estimating average priming effect involving one target word in a 2*K* trial study

Consider a sample of $2K$ subjects, where half of the subjects undertake a word study and half do not (see **Figure 2B**). Let S_Z denote the indices of the subjects who undertake the word study, and let S_{NZ} denote the indices of the subjects who do not. Let *DPE* denote the difference between the average observed word identification accuracy of the S_Z group and the average observed word identification accuracy of the S_{NZ} group, as follows:

$$DPE_{2K}^{\{1\}} = \frac{1}{K}\left\{\sum_{i\in S_Z} \mathrm{Y}_i\,(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 1)\right\} - \frac{1}{K}\left\{\sum_{i\in S_{NZ}} \mathrm{Y}_i\,(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 0)\right\}.$$

Following the definition of the *APE* in *Definition 1.2*, the average (causal) priming effect across a sample of $2K$ individuals involving one target word is defined as follows (see **Figure 2C**):

$$APE_{2K}^{\{1\}} = \frac{1}{2K}\left\{\sum_{i=1}^{2K} \mathrm{Y}_i\,(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 1) - \mathrm{Y}_i(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 0)\right\}.$$

Since $APE_{2K}^{\{1\}}$ is not observable and $DPE_{2K}^{\{1\}}$ is, one would ask if $DPE_{2K}^{\{1\}}$ is close to $APE_{2K}^{\{1\}}$. The answer depends on two factors: (a) how well matched are subjects who conduct the word study and those who do not; (b) if the word study is randomly assigned. We examine these factors in detail as follows:

First, if the S_Z group and the S_{NZ} group are perfectly matched[11] [i.e., for every subject in the S_Z group who receives a word study, there is a subject in the S_{NZ} group who does not receive a word study; and these two (matched) subjects would perform identically if a word study were conducted or if a word study were not conducted[12]], then $DPE_{2K}^{\{1\}} = APE_{2K}^{\{1\}}$. This holds whether the word study Z is randomly assigned or not (Rubin, 1974). Second, if the two groups are not perfectly matched, but before the tests, investigators have controlled all the variables that would affect the performance (e.g., only consider subjects with the same age, gender, and education background), then $DPE_{2K}^{\{1\}}$ is close to $APE_{2K}^{\{1\}}$ (i.e., the subjects are as if matched). Third, if the word study Z is randomly assigned, even if there are unmatched subjects (e.g., subjects have significant different language proficiency). For example, English speakers may perform better than non-English speakers in a word completion test in English; the random assignment is going to balance, in expectation, all observed and unobserved factors that would impact the word identification. To put it more concretely, by randomly assigning a word study to individuals, some English speakers would receive a word study (the rest of the English speakers would not receive one), and some non-English speakers would receive a word study (the rest non-English speakers would not receive one). As a result, the individuals who receive a word test consist of both English and non-English speakers, and the individuals who do not receive a word test also consist of both English and non-English speakers; thus, the bias due to language efficiency is reduced. Randomization becomes increasingly effective when the sample size N increases (Scheffe, 1959; Rubin, 1974; Wu and Hamada, 2000; Hinkelmann and Kempthorne, 2005).

Although matching or randomization makes $DPE_{2K}^{\{1\}}$ a suitable estimator for $APE_{2K}^{\{1\}}$, it remains important to generalize

11 The definition of "match" here is more restricted than it is in the context of propensity score matching, as we consider the matched pair to have identical potential outcomes. We use this term for illustration of causal effect rather than estimation.

12 We need both potential outcomes to be equal; not just the potential outcome under treatment (*i.e.*, $Z = 1$, namely had a word study been conducted).

it to any $2K$ sample. To that end, we defined the expected priming effect (EPE) (i.e., the expectation of $DPE_{2K}^{\{1\}}$) as follows:

$$EPE_{2K}^{\{1\}} = \mathbb{E}\left\{\frac{1}{K}\sum_{i\in S_Z} I_{\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=1)=1} - \frac{1}{K}\sum_{i\in S_{NZ}} I_{\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=0)=1}\right\}$$

where $\mathbb{E}$ indicates the expectation operation.

Since $\mathrm{Y}_i = 0$ or 1, then $DPE_{2K}^{\{1\}} = \frac{1}{K}\sum_{i\in S_Z} I_{\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=1)=1} - \frac{1}{K}\sum_{i\in S_{NZ}} I_{\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=0)=1}$, where $I_{\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=\mathrm{z})=1}$ is an indicator function[13] that takes value 1 if $\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=\mathrm{z}) = 1$, and takes value 0 if $\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=\mathrm{z}) = 0$. Then $EPE_{2K}^{\{1\}}$ reduces to

$$EPE_{2K}^{\{1\}} = \mathbb{P}\{\mathrm{Y}(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=1) = 1\} - \mathbb{P}\{\mathrm{Y}(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=0) = 1\} \quad (1)$$

where $\mathbb{P}\{\mathrm{Y}(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=\mathrm{z}) = \mathrm{y}\}$ denotes the probability of the word identification Y equals to y (y $= 0$ or 1) given the cue $X = \mathrm{x}$ and the word study Z equals to z ($z = 0$ or 1).

In simple terms (see Remark 3), $EPE_{2K}^{\{1\}}$ means that the expected priming effect estimated from $2K$ subjects regarding one word is the difference between the probability of correctly identifying the target word for all subjects who have taken the word study ($Z = 1$) and the probability of correctly identifying the target word for those who have not participated in the word study ($Z = 0$).

Estimating priming effect involving multiple target words

The variability of individual memory affects the individual priming effect (a treatment of which is to estimate the average priming effect across subjects, as outlined in *Definition 1.2*) and so does the variability of words. Hence, the *APE* estimated using a complicated, uncommon, and non-word is likely to differ from the *APE* estimated using a simple and common word; this is true even when words of similar complexity are considered (because even when we only focus on, say, words of intermediate complexity, there are, potentially, differences in syllable shapes and word patterns). A natural treatment is to conduct tests on multiple words and estimate the average priming effect over these words across subjects.

Definition 1.3 (Average priming effect across multiple studied words)

Consider a wordlist consisting of M target words in a study consisting of a total of N subjects (see **Figure 2D**). Define X_{ij} as a cue given to the i^{th} subject associated with the j^{th} target word. Define Y_{ij} as the outcome of the corresponding word identification. Then, the *average priming effect* of the word study Z across M words on multiple word identifications is defined as follows:

$$APE_N^{\{M\}} = \frac{1}{NM}\left\{\sum_{i=1}^{N}\sum_{j=1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=1\right) - \mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=0\right)\right\}$$

where $\mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=\mathrm{z}\right)$ indicates the word identification result of the j^{th} target word from the i^{th} subject after given the cue $\mathrm{X}_{ij} = \mathrm{x}_{ij}$ and the word study Z ($Z = 1$ means after viewing a wordlist consisting of target words and $Z = 0$ means without the word study). The superscript $\{M\}$ indicates that it is the average priming effect for M ($M \geq 2$) target words, and the subscript N indicates that the estimate is obtained from a sample of N subjects.

Again, $APE_N^{\{M\}}$ is not observable. The observable DPE in a study consisting of $N = 2K$ subjects and M target words (between the group given a word study and the group not given a word study) is as follows:

$$DPE_{2K}^{\{M\}} = \frac{1}{KM}\left\{\sum_{i\in S_Z}\sum_{j=1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=1\right)\right\} - \frac{1}{KM}\left\{\sum_{i\in S_{NZ}}\sum_{j=1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=0\right)\right\}. \quad (2)$$

Similar to a $2K$ trial study concerning one target word, the expected priming effect reduces to

$$EPE_{2K}^{\{M\}} = \frac{1}{M}\sum_{j=1}^{M} EPE_{2K,j}^{\{1\}} \quad (3)$$

where $EPE_{2K,j}^{\{1\}}$ refers to $EPE_{2K}^{\{1\}}$ for the j^{th} word.

In simple terms, $EPE_{2K}^{\{M\}}$ means that the expected priming effect estimated from $2K$ subjects across M words is the difference between the probability of corrected identifying *each* of the M target words for all subjects who have participated in the word study ($Z = 1$) and the probability of correctly identifying the corresponding word for all subjects who have not participated in the word study ($Z = 0$) averaged over M words. For simplicity, let us denote $\mathbb{P}\{\mathrm{Y}(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=1) = 1\}$ as p_1 and $\mathbb{P}\{\mathrm{Y}(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=0) = 1\}$ as p_0, which can be estimated by $\frac{1}{KM}\left\{\sum_{i\in S_Z}\sum_{j=1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=1\right)\right\}$ and $\frac{1}{KM}\left\{\sum_{i\in S_{NZ}}\sum_{j=1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij}=\mathrm{x}_{ij},\ \mathrm{Z}=0\right)\right\}$, respectively.

Remark 3. Eqs. (1, 3) are analytical solutions to estimating the priming effect for one target word and M target words, respectively. They have been used intuitively by experimentalists; the aforementioned arguments demonstrate the mathematical validity of such usages in practice.

13 Random variables related to individuals are assumed to be independent and identically distributed (*i.i.d.*). Thus, we can write the indicator function $I_{\mathrm{Y}_i(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=\mathrm{z})=1}$ as $I_{\mathrm{Y}(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=\mathrm{z})=1}$.

Connecting the potential outcomes framework with multinomial processing tree (MPT) models in studying priming

It turns out that the potential outcome framework-based priming study discussed here can be linked to the priming study using the multinomial processing tree (MPT) model (Batchelder and Riefer, 1999; Erdfelder et al., 2009). To see this, consider a word fragment completion test using an MPT diagram (see **Figure 3A**).

Here, let us denote A and $1-A$ as the probabilities of correctly and incorrectly, respectively, identifying the words without a word study (i.e., given $Z = 0$). Let us denote B and $1-B$ as the probabilities of storing (consciously or unconsciously) and not storing, respectively, the studied word after a word study (i.e., given $Z = 1$). Let C and $1-C$ be the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are stored in the memory; let D and $1-D$ be the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are not stored in the memory.

Naturally, $p_1 = BC + D(1-B)$ and $p_0 = A$, where p_1 and p_0 are defined previously, and the priming effect estimated using the potential outcomes framework is $p_1 - p_0$. We have $p_1 - p_0 = \{BC + D(1-B)\} - A = B(C-D) + (D-A)$. Note that (1) $B(C-D)$ is the product of consciously or unconsciously storing information from the word study into the memory (i.e., B) and the improvement[14] of word identification accuracy, thanks to the stored information [i.e., $(C-D)$]; (2) $(D-A)$ gives the difference between the probability of correctly uncovering words after a word study, even though no information from the word study has been added into the memory (to correctly identify words, one, therefore, has to either actively retrieve existing knowledge or use guessing), and the probability of uncovering words without a word study (which also relies on either existing knowledge or guessing).

Furthermore, it is not unfair to assume that A equals or is very close to D. Suppose $A = D$, then the relationship between the priming effect identified using the potential outcomes framework and the MPT model simplifies to $p_1 - p_0 = B(C-D)$. In other words, the potential priming effect (i.e., $p_1 - p_0$) chiefly depends on two factors: first, the consciously or unconsciously stored memory from the word study (i.e., B); second, the improvement of word identification accuracy, thanks to the stored information (i.e., $(C-D)$).

The aforementioned argument can be extended to studying multiple, successive word identification phases. We leave this to our readers as an exercise.

14 It is natural to assume $C = D$; otherwise, we can replace "improvement" with "the difference."

Extending the framework to experimental paradigms with exclusion and inclusion instructional conditions

Consider a three-phase study (see **Figure 3B**), where two sets of different items are shown during Phases 1 and 2 for learning purposes, and during Phase 3, the participants are given a list of items consisting of items that have appeared during Phases 1 and 2 and distractor items that have not appeared before. Subsequently, they are asked to classify them into either "old" or "new" following an inclusion instruction or an exclusion instruction (Buchner et al., 1995). Under an inclusion instruction, the participants need to call an item *old* if it has appeared in either Phase 1 or 2 and call a distractor item *new*. Under an exclusion instruction, the participants need to call an item *old* only if it has appeared in Phase 2, and *new* otherwise.

The framework proposed in this article can also be modified to study the experimental paradigm with exclusion and inclusion instructional conditions. To demonstrate this, let us define Z_1 and Z_2 as two lists of items during Phases 1 and 2, respectively. Let X and X' denote the outcomes of the identification during Phase 3 under inclusion and exclusion instructions, where their realizations are either {new} or {old} for each given item. Following similar arguments as before, we define the potential difference between the results from the inclusion and exclusion instructions as follows:

$$\frac{1}{NM}\left\{\sum_{i=1}^{N}\sum_{j=1}^{M} \mathrm{Y}_{ij}(\mathrm{X}_{ij} = \mathrm{x}_{ij}, Z_1 = 1, Z_2 = 1) - \mathrm{Y}_{ij}(X'_{ij} = x'_{ij}, Z_1 = 1, Z_2 = 1)\right\}.$$

Note that here, it is assumed that the inclusion and exclusion instructions are given to the same participants in a group. This is not ideal as repeating Phase 3 under different conditions may bias the results. Using the potential outcome framework, this scenario can be estimated as follows:

$$\frac{1}{KM}\left\{\sum_{i\in S_I}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij}, Z_1 = 1, Z_2 = 1\right)\right\} - \frac{1}{KM}\left\{\sum_{i\in S_E}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(X'_{ij} = x'_{ij}, Z_1 = 1, Z_2 = 1\right)\right\}$$

where the two parts (before and after the minus sign) are estimated from subjects in groups S_I (following the inclusion instruction) and S_E (following the exclusion instruction), respectively. Note that the aforementioned result equals to $p_{i-}p_e$ in Buchner et al. (1995), which quantifies the probability of consciously recollecting a Phase 1 item.

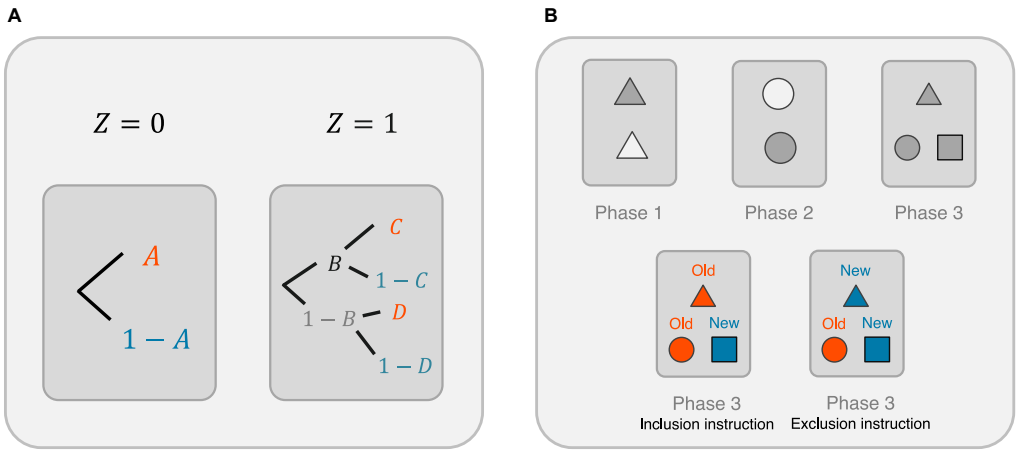


FIGURE 3
Linking the potential outcomes-based priming research with two prominent quantitative psychological methods. **(A)** Multinomial processing tree (MPT) model for studying the priming effect. Here, A and $1-A$ are the probabilities of correctly and incorrectly, respectively, identifying the words without a word study (i.e., given $Z = 0$); B and $1-B$ are the probabilities of storing (consciously or unconsciously) and not storing, respectively, the studied word after a word study (i.e., given $Z = 1$); C and $1-C$ are the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are stored in the memory; D and $1-D$ are the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are not stored in the memory. **(B)** A three-phase experiment with exclusion and inclusion instructional conditions. Top left: A set of items is shown during the Phase 1 study. Top middle: Another set of items (different from those in Phase 1) is shown during Phase 2. Top right: During Phase 3, the participants are given a list of items consisting of those who have appeared during Phases 1 and 2 and distractor items that have not appeared before. Subsequently, they are asked to classify them into either "old" or "new" following an inclusion instruction or an exclusion instruction. Bottom left: Under the inclusion instruction, participants need to call an item old if it has appeared in either Phase 1 or 2 and call a distractor item new. Bottom right: Under the exclusion instruction, participants need to call an item old only if it has appeared in Phase 2, and new otherwise.

Testing the significance of the priming effect

Returning to the priming study, although the focus of this article so far has been to define and quantify the causal priming effect, it may also be important for investigators to test whether a detected causal effect is significant. For example, consider 100 subjects who have undertaken the word study and 100 people who have not undertaken the word study. Suppose the estimated expected priming effect (*EPE*) is 0.1; is 0.1 in a sample of 200 subjects significant (from 0, where 0 indicates no priming effect)? What if the estimated *EPE* is 0.05?

One way to answer this question is to conduct a hypothesis test on whether the estimated priming effect is significant; that is, to verify the (alternative) hypothesis that the *EPE* is significantly greater than zero. Thanks to **Eqs.** 1, 3, the *EPE* can be written in terms of probability and can therefore be examined using a proportion test (Ott and Longnecker, 1980; Bickel and Doksum, 2000).

Formally, the test statistic is defined as follows:

$$z = \frac{DPE}{\sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{N_1}+\frac{1}{N_0}\right)}} \tag{4}$$

where $\hat{p}_1 = \frac{1}{KM}\sum_{i\in S_Z}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 1\right)$,
$\hat{p}_0 = \frac{1}{KM}\sum_{i\in S_{NZ}}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 0\right)$,
$DPE = \hat{p}_1 - \hat{p}_0$, $\hat{p} = \frac{1}{2KM}\sum_{i\in S_Z\cup S_{NZ}}\sum_{j=1}^{M} \mathrm{Y}_{ij}(\mathrm{X}_{ij} = \mathrm{x}_{ij}, Z = \{0,1\})$, and $N_1 = N_2 = KM$ [15]. Here, the *DPE* is the empirical estimate of the *EPE* obtained from **Eq.** 2 and $\hat{p}$ is the pooled probability (from both groups) of correct word identification (in other words, the overall probability of correctly identifying a word when the group that undertaken a word study and the group that did not undertake a word study are combined). The hat symbol, for example, in $\hat{p}_1$ is an estimate of p_1.

One can then compare the p-value associated with the z score to evaluate the significance. Note that the aforementioned z-test is the same as a Chi-square test, where the z-statistic is equal to the square root of the Chi-square statistic, and the p-values of the two tests are identical. When the word studies are multivariate (e.g., there are more than two types of word study), continuous (e.g., the word study involves words with different degrees of complexity), or time-dependent (e.g., several tests are carried out with large time intervals in between), more advanced statistical tests can be used (see the Discussion section for details).

Subsequently, the 100 $(1-\alpha)$ percent confidence interval (Wilson, 1927; Newcombe,

15 Readers can relatively easily extend it to more complicated cases involving unbalanced groups, where the numbers of subjects and/or target words in two groups are unequal.

1998) for the estimated priming effect is as follows:

$$\left(DPE - z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{N_1}+\frac{\hat{p}_0\left(1-\hat{p}_0\right)}{N_0}},\right.$$

$$\left.DPE + z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{N_1}+\frac{\hat{p}_0\left(1-\hat{p}_0\right)}{N_0}}\right).$$

Estimating priming effects with covariates

Although the word study Z is the primary factor that affects the outcome Y, it remains possible that there exist additional variables (denoted as W) that, if not considered, may bias the estimation of the causal priming effect. These variables could either have a causal relationship with the word identification outcome Y (e.g., take W as intelligence) or are spuriously (i.e., by chance) correlated with the outcome in the sample (e.g., one's height). Randomization only ensures that in expectation, the covariates are balanced between the two treatment groups. There, however, could still be chance imbalances in the covariates between the two treatment groups; in this case, adjusting for the covariates will increase the signal-to-noise ratio[16] and make the priming effect more likely to be detected, if exists.

For example, take W as one's IQ, which may affect word identification. Consider 20 subjects with a mean IQ of 100 (10 with IQ larger than 100 and 10 with IQ less than 100). Certainly, we could create two splits with each split containing five individuals with above-average IQ and five with below-average IQ. Our point is that sometimes, such a balanced sample is difficult to obtain, and thus, protective measures need to be taken instead (a good example here is the proficiency in English language – another variable that may affect word identification). In practice, however, it is difficult and costly for researchers to collect samples that contain subjects that are perfectly matched. Thus, we proceed here assuming such a (not completely matched) case occurs. For example, if we are to randomly assign a word study ($Z = 1$) to 10 subjects and no word study ($Z = 0$) to another 10 subjects, the group with the word study may contain eight subjects with above-average IQ and the other group with two subjects with above-average IQ. Then, the result using **Eq.** 2 could potentially over-estimate the priming effect since there are more people with above-average IQ in the word study group.

The effect of an additional variable can be adjusted in a logistic regression model. Specifically, consider

$$logit(\mathbb{P}\{\mathrm{Y}_i\,(\mathrm{X}=\mathrm{x},\ \mathrm{Z}=z_i)\ =1\}) = \beta_0+\beta_z z_i+\beta_w w_i$$

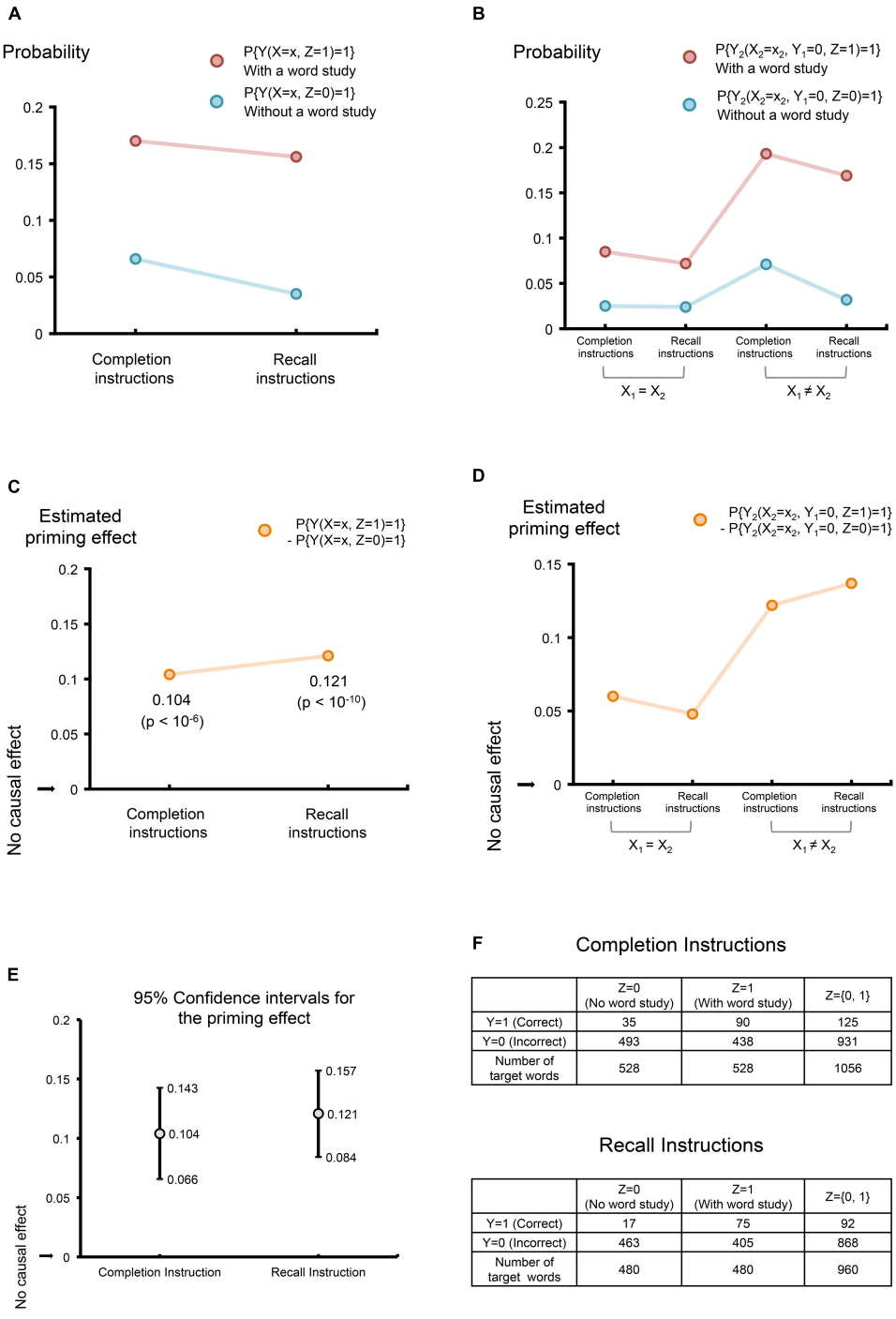
where z_i indicates whether the i^{th} subject receives a word study or not, w_i is the IQ for the i^{th} subject, β_0 is the estimated intercept, and β_z and β_w are the estimated parameters for z_i and w_i, respectively. The estimated β_z then indicates the priming effect from Z, when it is adjusted for the IQ effect (W). Specifically, controlling (i.e., removing) the effect from IQ to word identification Y, β_z quantifies the priming effect: the probability of correctly identifying a word increases $\frac{e^{\beta_z}}{1+e^{\beta_z}}$ when an individual conducts a word study versus not conducting a word study. Again, the logistic formula is stated for a word study considering one target word with the same cue $X = x$ and can be relatively easily extended to a study considering multiple words and multivariate covariates.

In the following, we will perform data analysis using data from a between-subjects study (Hayman and Tulving, 1989) to demonstrate how to use the framework to study the potential causal priming effect. The advantage of using a between-subject study is that the priming effect can be evaluated when the same experiment cannot be run on the same subjects more than one time; it may also reduce the likelihood of carryover or learning effect in a repeated-measures design (see section "Discussion").

Results

Consider a between-subjects WFC test. A total of 84 students enrolled in a second-year psychology course at the University of Toronto were randomly divided into two groups (one with the last name A-K and the other with the last name L-Z). A set of 48 target English words of intermediate difficulty was selected from a word pool and divided into two wordlists (A and B), with 24 target words in each list. An additional 64 (non-target) English words were used as buffer words. During the study phase, the first group studied wordlist A and the second wordlist B; the wordlist B thus served as non-studied words for the first group, and wordlist A served as non-studied words for the second group. During the test phase, there were two test instructions: the subjects with completion instructions were asked to complete the fragment with any word that comes to mind; subjects with recall instructions were asked to complete the fragment only with studied words. All subjects are randomized into four groups, each to take two tests. Specifically, participants in Group 1 ($N = 22$) conducted two tests under the completion instructions with the same fragment cues during the two tests; participants in Group 2 ($N = 22$) conducted two tests under the completion instructions with different fragment cues during the two tests; participants in Group 3 ($N = 20$) conducted two tests under the recall instructions with the same fragment cues during the two tests; participants in Group 4 ($N = 20$) conducted two tests under the recall instructions with different fragment cues during the two tests. Full data description is available in Experiment 2 in Hayman and Tulving (1989) with study data summarized in **Figure 4F**.

16 Statistically speaking, adjusting for covariates in a randomized study improves the precision (reduces the variance) of the treatment effect estimator (see Moore et al., 2011; Qian et al., 2018).



F Completion Instructions

	Z=0 (No word study)	Z=1 (With word study)	Z={0, 1}
Y=1 (Correct)	35	90	125
Y=0 (Incorrect)	493	438	931
Number of target words	528	528	1056

Recall Instructions

	Z=0 (No word study)	Z=1 (With word study)	Z={0, 1}
Y=1 (Correct)	17	75	92
Y=0 (Incorrect)	463	405	868
Number of target words	480	480	960

FIGURE 4
Estimating the causal priming effect. **(A)** Probabilities of correctly uncovering target words given fractional cues, under the completion instructions and the recall instructions. The color red is used to indicate experiments involving a word study including target words; the color blue is used to indicate experiments involving a word study without target words (abbreviated as "without a word study"). **(B)** The probabilities of correctly uncovering target words during a second cue-based test, given that the first test failed to uncover the same word. Four experimental sceneries were considered, combining two instruction strategies: the completion instructions and the recall instructions, and two types of cues: the same cues and different cues during two tests. X_1 and X_2 refer to cues from the first test and the second test, respectively; $X_1 = X_2$ indicates the same cues were used in the two tests; $X_1 \neq X_2$ indicates different cues were used in the two tests. Again, the color red is used to indicate experiments involving a word study including target words; the color blue is used to indicate experiments involving a word study without target words (abbreviated as "without a word study"). **(C)** The estimated causal priming effects and their *p*-values correspond to **(A)**. **(D)** The estimated causal priming effects correspond to **(B)**. **(E)** The estimated 95% confidence intervals for the priming effect in **(C)**. **(F)** Data used in estimating the priming effect corresponding to **(A,C,E)**. Data adapted by permission from RightsLink Permissions American Psychological Association "Is priming in fragment completion based on a 'traceless' memory system?" by Hayman and Tulving (1989), American Psychological Association.

The potential priming effect of the WFC test is displayed in Figure 4. Using **Eqs.** 2, 4, the *DPE* under completion instruction is 0.104 ($z = 5.23$, $p = 10^{-6}$), with a 95% confidence interval (0.066, 0.143); the *DPE* under recall instructions is 0.121 ($z = 6.36$, $p = 10^{-10}$), with a 95% confidence interval (0.084, 0.157). The corresponding estimated probabilities (of correct word identification with or without a word study), priming effects and their confidence intervals are shown in Figures 4A,C,E, respectively.

Next, we consider the *DPEs* where the word study primes the identification of a target word in the second test, given the identification of the same target word failed during the first test (see Figures 4B,D). Although the number of the words that failed to be identified during the first test was not reported in Hayman and Tulving (1989) (hence we cannot compute the exact *p*-values), readers could follow the previous example and use **Eq.** 4 in their research when data are available. Nevertheless, we will report the *DPEs* without *p*-values. There are two reasons for this. First, since the total number of the target words studied is large [i.e., 480 words (20 subjects each with 24 target words) and 528 words (22 subjects each with 24 target words) in our case], a positive *DPE* is likely to yield a significance non-zero priming effect. Second, it allows us to numerically compare the priming effects under different experimentation strategies. Specifically, using the same cues in the two tests, the *DPE* under the completion instructions is 0.06; the *DPE* under the recall instructions is 0.048; meanwhile, using the different cues in the two tests, the *DPE* under the completion instructions is 0.122; the *DPE* under the recall instructions is 0.137 (see Figure 4D). The much stronger priming effect observed in both experiments where different cues are provided suggests that given a failed attempt using one cue during the first test, information has potentially been learned by combining the first cue and a *different* cue during the second test.

Extensions, limitations, future directions, and final remarks

In this article, we defined, quantified, and tested the priming effect using the *potential outcomes framework*. Although we only considered cases involving a binary exposure (having a word study versus not having a word study), the framework can be extended to categorical exposures (e.g., we can code an exposure that does not consist of a word study as $Z = 0$, one that consists of a word study including short words as $Z = 1$, and one that consists of a word study including long words as $Z = 2$). In addition, the framework can be extended to continuous exposures (e.g., when a word study consists of words with different degrees of complexity, we can allow Z to take any value between 0 and 100 to indicate complexity of each word). Furthermore, it can be extended to time-dependent exposures [e.g., we can write word studies conducted at different times as $Z(t)$, for each time t]. Finally, it can also be extended to cases where several exposures are concerned (e.g., let $Z_1 =$ reading a list of words, $Z_2 =$ viewing a list of non-word symbols, and $Z_3 =$ listening to a list of words), where the priming effect for each exposure can be estimated when the other exposures are controlled. For example, when estimating the priming effect of symbol recognition ($Z_2 = 1$ versus $Z_2 = 0$), one could fix Z_1 and Z_3; namely, the priming effect can be estimated, for example, using $Y(X = x, Z_1 = 0, Z_2 = 1, Z_3 = 0) - Y(X = x, Z_1 = 0, Z_2 = 0, Z_3 = 0)$, where bold X indicates all cues used for three studies.

Eq. 4 is used to test the significance of the priming effect with binary exposures. When the exposures are multivariate, continuous, or time-dependent, the test can be carried out by first arranging the exposure and outcome as explanatory and dependent variables in a regression setting, and then testing the exposure effect by examining the significance of the (regression) parameters. For example, when there are three types of word studies (no word study, a word study with short words, and a word study with long words), one can consider a regression model with a block design, where each block consists of subjects from one of the three groups. The estimated regression parameter for the block variable then indicates the priming effect between two paired groups. When the exposure is time-dependent[17], one could refer to functional regression models, wherein $Z(t)$ and $Y(t)$ are treated as functional regressors and responses, respectively (Ramsay and Silverman, 1997).

The proposed method aimed at providing a framework that could estimate and validate analytically priming in between-subjects designs. It nonetheless has a few limitations. First, we demonstrated the method using data from previous experiments (that are not primarily intended to evaluate the priming effect but to assess the independence of successive tests). Inevitably, this restricted our arguments; future research may verify and expand our analysis to general priming research. Future studies may also extend to cases with a larger sample, and non-twin studies need to examine covariant control under the potential outcomes framework (see section "Estimating Priming Effects With Covariates") and its utility on providing an estimated priming effect that is less biased. In parallel, future research may further consider twin studies where the subjects are nearly perfectly matched. Second, the method we introduced rests on the Neyman–Rubin potential outcomes framework. There is, however, on the one hand, not as of yet a consensus that one causal framework is better than others, although we have discussed the advantages of the Neyman–Rubin framework in estimating potential priming effect (especially its mathematical representations). On the other hand, we recognize that despite

17 For example, at each time t, a study $Z(t)$ is assigned to individuals, and a word identification $Y(t)$ is observed. This is particularly useful when the sample consists of subjects whose implicit memory degenerates over time.

differences and disagreements [e.g., see Gelman's blog post (Gelman, 2009) and discussion under the post], there is some commonality between Neyman–Rubin's and Pearl's frameworks (Section 7.4.5 of Pearl, 2009b), and there exists "a happy symbiosis between graphs and counterfactual notation" (Section 7.4.4 of Pearl, 2009b). In this study, while we present the arguments using the Neyman–Rubin model, we have adopted Pearl's diagram representation (although without graphic notations) to illustrate the experiments. We do so without implying that one framework is superior to the other. Future studies may theoretically compare the Neyman–Rubin approach with Pearl's approach in detail for studying potential causal priming (e.g., their mathematical or empirical equivalence or difference). Further research may also incorporate Campbell's approach to identify potential threats that may impair the validity of inferences made on the estimated priming effect.

Sometimes, investigators studying the priming effect may observe post-treatment variables (i.e., variables obtained after the word study Z is assigned). Examples of post-treatment variables are (a) a measure of subjects' compliance to the originally assigned word study – a subject chooses not to take the word study after it is assigned; (b) in studies with a long time interval between two priming tests, whether or not the subject drops out is a post-treatment variable (missingness of outcome); (c) in longitudinal (priming) studies involving patients with severe amnesia, the outcome can be censored (i.e., not recorded due to death); (d) in studies investigating priming effects for patients with brain disorders, surrogate variables of disease progression and fluctuation, such as the degree of memory loss, are post-treatment variables. The estimators provided in this article can only be used to adjust for pre-treatment variables; if one adjusts for post-treatment directly using the framework outlined in this article, the estimated effects are no longer causal (Frangakis and Rubin, 2002).

It is worthwhile noting that besides the potential outcome framework (by comparing outcomes on randomly selected or matched subjects, or subjects with covariates adjusted), priming can also be estimated using a repeated-measures design, in which *all* subjects are exposed first to half of the target words and then another half of the target words (e.g., see Challis and Brodbeck, 1992). The priming effect can then be estimated as the difference between the proportion of fragments of studied words completed and the proportion of fragments of non-studied words completed for the same subject (and averaged across all subjects). Instead of matching two groups of subjects as proposed in this study, the key to using the repeated-measures designs is to match the length, frequency, etc., of the words and randomize the words employed. Whereas this indeed provides an alternative (and potentially convenient[18]) approach to assess the priming effect, and we welcome future research to compare this approach with the potential outcomes framework; a key concern with this method is the carryover or learning effect. The carryover or learning effect here is not necessarily the phenomenon where after studying the same (or similar) words multiple times, the earlier word study and identification may improve the same subject's later word identification; rather, it also includes the phenomenon where the experiment mechanism of the first repeated-measures study may improve learning during the second repeated-measures study. We have seen such a carryover or learning effect during a smartphone-based cognitive test, where even though different tests (e.g., drawing different shapes) were given to the same subjects over time, their performance improved. For the WFC test, it may be possible that the subjects learned some rules (despite not being informed) during the first half of the experiment or became more focused during the second half either because they had guessed the approximate rule or because they had realized that the word study may be an important part (since, for example, two wordlists had been given sequentially) to their performance of the experiment. Future studies could examine the existence of such a learning or carryover effect, and if exists, whether and how it would affect estimating the priming effect.

There are times where even randomization becomes impossible. For example, suppose one is interested in studying how a new medicine affects priming; in this case, we have two potential causes: a word study (Z) and medication (Med). It is unethical to assign a group of 45-year-old healthy subjects to take a new drug to investigate whether the drug improves priming at 50. In addition, there is likely another source, say, the socioeconomic status (which may be related to the affordability of new drugs) or genetics (if there is a family history of memory problems, one may be more willing to take the drug), that may be associated with taking the drug and/or developing memory problems at 50. Similarly, it would be difficult to estimate the effect of taking the drug on improving priming by comparing the performance of an individual at 50 who had taken the drug with his or her performance at 50 had he or she not taken the drug. To solve these issues, the propensity score matching (PSM) estimates the treatment effect by comparing the outcomes of the subjects under treatment (e.g., taking the drug) with a set of "matched" subjects without treatment (e.g., having not taken the drug) (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 1999, 2002; Caliendo and Kopeinig, 2008). More concretely, one could first compute the propensity score of A's and B's taking the drug based on their gender, economic, social, genetic, and demographic backgrounds, and choose two individuals C and D from a group of 50-year-olds who had not taken the drug but have propensity scores (of taking the drug during their younger years) closest to A's and B's, respectively. Subsequently, A and C will receive a word study, and B and D will not. Following

18 For example, when there are no matched samples (but see Propensity Score Matching (PSM) and covariates adjustment discussed in this paper), and that the carryover or learning effect is ignorable, a repeated-measures design is attractive.

the previous notations, we have $Y_A\left(X_j, Z = 1, Med = 1\right)$, $Y_B(X_j, Z = 0, Med = 1)$, $Y_C(X_j, Z = 1, Med = 0)$, and $Y_D\left(X_j, Z = 0, Med = 0\right)$, for $1 \leq j \leq M$, where M target words are considered. Then, the priming effects for the group taking the drug and the one not taking the drug are $DPE_{Med=1} = \frac{1}{M}\left\{\sum_{j=1}^{M} Y_A\left(X_j, Z = 1, Med = 1\right) - \sum_{j=1}^{M} Y_B\left(X_j, Z = 0, Med = 1\right)\right\}$ and $DPE_{Med=0} = \frac{1}{M}\left\{\sum_{j=1}^{M} Y_C\left(X_j, Z = 1, Med = 0\right) - \sum_{j=1}^{M} Y_D\left(X_j, Z = 0, Med = 0\right)\right\}$, respectively. Subsequently, we can estimate the drug effect on priming using $DPE_{Med=1} - DPE_{Med=0}$. Note that for simplicity, only one individual is considered for each of the 2×2 factors; one can relatively easily extend the above to include multiple subjects in each group.

Although we have throughout focused on a type of non-semantic priming, other studies have reported that new semantic knowledge can be acquired among (even) patients with amnesia. For example, the learning of specified target words in meaningful texts, statements of facts about people and places, specified target words as parts of meaningful sentences, new computer-related vocabulary, computer commands, semantic interpretations of ambiguous descriptions of situations and events, and production of words to cues consisting of the initial letters of words (see Hayman et al., 1993 for a summary of studies). Future studies should independently verify the extent to which the framework introduced in this article can be used to estimate causal semantic priming. A beginning can, perhaps, be made by reporting the individual ratings of the meaningfulness of the target words (e.g., during a word study, every participant is to rate on a scale of 0–10, the meaningless of each studied word), and subsequently treating the ratings as covariates.

In conclusion, we define, quantify, and test the causal priming effect using the *potential outcomes* framework. Applying data from a between-subjects word completion test, we demonstrate that the framework identifies a significant priming effect from a word study to cue-based word identification, under both completion and recall instructions; the priming effect under the recall instructions is more significant than that under the completion instructions. Furthermore, when there are two consecutive tests, the framework shows that even if the word identification failed during the first test, there is likely a priming effect from the initial word study to the second word identification, regardless of the type of instructions and whether the same or different cues are used in the two tests. In addition, there is a stronger priming effect in experiments where different cues are provided, suggesting that given a failed attempt using one cue during the first test, additional information may have been learned by combining the first cue and a different cue during the second test. Finally, our explorations show that what has been intuitively used by scholars to estimate the priming effect in the past has a meaningful mathematical basis.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: Hayman and Tulving (1989).

Ethics statement

The studies involving human participants were reviewed and approved by the University of Toronto. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

OC designed the potential causal priming framework (based on Neyman and Rubin's works on causal inference and Tulving and Schacter's works on priming) and performed the analysis. HP provided computational support. HC and GN provided neurobiological and psychological interpretations. TQ provided statistical support. MV provided funding, support, and guidance. OC wrote the manuscript, with comments from all other authors. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to Semir Zeki for his useful comments on earlier versions of the article. We thank the two reviewers for their insightful comments and suggestions, which have significantly improved the range, depth, and quality of the article.

Conflict of interest

Author GN is Medical Director of Neurology at, and minority shareholder of, icometrix. OC had consulted for F. Hoffmann-La Roche.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1

Baddeley, A., Lewis, V., and Vallar, G. (1984). Exploring the articulatory loop. *Q. J. Exp. Psychol. Sect. A* 36, 233–252. doi: 10.1080/14640748408402157

Batchelder, W. H., and Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* 6, 57–86. doi: 10.3758/BF03210812

Bickel, P. J., and Doksum, K. A. (2000). *Mathematical statistics: Basic ideas and selected topics, Vol I*, 2nd Edn, Boca Raton, FL: CRC Press.

Buchner, A., Erdfelder, E., and Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *J. Exp. Psychol. Gen.* 124, 137–160. doi: 10.1037/0096-3445.124.2.137

Caliendo, M., and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *J. Econ. Surv.* 22, 31–72. doi: 10.1111/j.1467-6419.2007.00527.x

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychol. Bull.* 54, 297–312. doi: 10.1037/h0040950

Cermak, L. S., Talbot, N., Chandler, K., and Wolbarst, L. R. (1985). The perceptual priming phenomenon in amnesia. *Neuropsychologia* 23, 615–622. doi: 10.1016/0028-3932(85)90063-6

Challis, B. H., and Brodbeck, D. R. (1992). Level of processing affects priming in word fragment completion. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 595–607. doi: 10.1037/0278-7393.18.3.595

Cochran, W. G., and Chambers, S. P. (1965). The planning of observational studies of human populations. *J. R. Stat. Soc. Ser. A* 128, 234–266. doi: 10.2307/2344179

Cohen, N. J., and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science* 210, 207–210. doi: 10.1126/science.7414331

Cooper, G., Tindall-Ford, S., Chandler, P., and Sweller, J. (2001). Learning by imagining. *J. Exp. Psychol. Appl.* 7, 68–82. doi: 10.1037/1076-898X.7.1.68

Dehejia, R. H., and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J. Am. Stat. Assoc.* 94, 1053–1062. doi: 10.1080/01621459.1999.10473858

Dehejia, R. H., and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Rev. Econ. Stat.* 84, 151–161. doi: 10.1162/003465302317331982

Ding, P. (2017). A paradox from randomization-based causal inference. *Stat. Sci.* 32, 331–345. doi: 10.1214/16-STS571

Ding, P., Feller, A., and Miratrix, L. (2016). Randomization inference for treatment effect variation. *J. R. Stat. Soc. Ser. B* 78, 655–671. doi: 10.1111/rssb.12124

Erdfelder, E., Auer, T. S., Hilbig, B. E., Aßfalg, A., Moshagen, M., and Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychol.* 217, 108–124. doi: 10.1027/0044-3409.217.3.108

Frangakis, C. E., and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* 58, 21–29. doi: 10.1111/j.0006-341X.2002.00021.x

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *Br. J. Psychol.* 74, 159–180. doi: 10.1111/j.2044-8295.1983.tb01851.x

Gelman, A. (2009). *Resolving disputes between J. Pearl and D. Rubin on causal inference*. Available online at: https://statmodeling.stat.columbia.edu/2009/07/05/disputes_about/ (accessed September 26, 2021).

Goldberger, A. S. (1972). *Selection bias in evaluating treatment effects: Some formal illustrations*. Discussion Paper 123–172. Madison, WI: Institute for Research on Poverty.

Graf, P., Squire, L. R., and Mandler, G. (1984). The information that amnesic patients do not forget. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 164–178. doi: 10.1037/0278-7393.10.1.164

Hanley, J. R., and Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *Br. J. Psychol.* 78, 287–297. doi: 10.1111/j.2044-8295.1987.tb02247.x

Hashtroudi, S., Parker, E. S., DeLisi, L. E., Wyatt, R. J., and Mutter, S. A. (1984). Intact retention in acute alcohol amnesia. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 156–163. doi: 10.1037/0278-7393.10.1.156

Hayman, C. A. G., and Tulving, E. (1989). Is priming in fragment completion based on a "traceless" memory system? *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 941–956. doi: 10.1037/0278-7393.15.5.941

Hayman, C. A. G., Macdonald, C. A., and Tulving, E. (1993). The role of repetition and associative interference in new semantic learning in amnesia: A case experiment. *J. Cogn. Neurosci.* 5, 375–389. doi: 10.1162/jocn.1993.5.4.375

Hill, A. B. (1965). The environment and disease: Association or causation? *Proc. R. Soc. Med.* 58, 295–300. doi: 10.1177/003591576505800503

Hinkelmann, K., and Kempthorne, O. (2005). *Design and analysis of experiments*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/0471709948

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *J. Mem. Lang.* 30, 513–541. doi: 10.1016/0749-596X(91)90025-F

Johnson, M. K. (1983). "A multiple-entry, modular memory system," in *The psychology of learning and motivation*, ed. G. H. Bower (New York, NY: Academic Press), 81–123. doi: 10.1016/S0079-7421(08)60097-3

Kopelman, M. D., and Corn, T. H. (1988). Cholinergic "blockade" as a model for cholinergic depletion: A comparison of the memory deficits with those of alzheimer-type dementia and the alcoholic korsakoff syndrome. *Brain* 111, 1079–1110. doi: 10.1093/brain/111.5.1079

Light, L. L., Singh, A., and Capps, J. L. (1986). Dissociation of memory and awareness in young and older adults. *J. Clin. Exp. Neuropsychol.* 8, 62–74. doi: 10.1080/01688638608401297

Mishkin, M., Malamut, B., and Bachevalier, J. (1984). "Memories and habits: Two neural systems," in *The neurobiology of learning and memory*, eds G. Lynch, J. L. McGaugh, and N. M. Weinberger (New York, NY: Guilford press), 65–77.

Moore, K. L., Neugebauer, R., Valappil, T., and van der Laan, M. J. (2011). Robust extraction of covariate information to improve estimation efficiency in randomized trials. *Stat. Med.* 30, 2389–2408. doi: 10.1002/sim.4301

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat. Med.* 17, 873–890. doi: 10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I

Neyman, J. (1923). On the application of probability theory to agricultural experiments Appeared (in Polish). *Rocz. Nauk Rolniczych*. 10, 1–51.

Nissen, M. J., Knopman, D. S., and Schacter, D. L. (1987). Neurochemical dissociation of memory systems. *Neurology* 37, 789–789. doi: 10.1212/WNL.37.5.789

Ott, R. L., and Longnecker, M. T. (1980). *An introduction to statistical methods and data analysis*, 6th Edn. Boston, MA: Cengage.

Parker, E. S., Schoenberg, R., Schwartz, B. L., and Tulving, E. (1983). "Memories on the rising and falling blood-alcohol curve," in *Paper presented at the twenty-fourth annual meeting of the Psychonomic Society*, San Diego, CA.

Parkin, A. J., and Streete, S. (1988). Implicit and explicit memory in young children and adults. *Br. J. Psychol.* 79, 361–369. doi: 10.1111/j.2044-8295.1988.tb02295.x

Pearl, J. (1993). Comment: Graphical models, causality and intervention. *Statist. Sci.* 8, 266–269. doi: 10.1214/ss/1177010894

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* 82, 669–688. doi: 10.1093/biomet/82.4.669

Pearl, J. (2001). "Direct and indirect effects," in *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, (San Francisco, CA: Morgan Kaufmann Publishers Inc), 411–420.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Stat. Surv.* 3, 96–146. doi: 10.1214/09-SS057

Pearl, J. (2009b). *Causality: Models, reasoning, and inference*, 2nd Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803161

Peters, C. C. (1941). A method of matching groups for experiment with no loss of population. *J. Educ. Res.* 34, 606–612. doi: 10.1080/00220671.1941.10881036

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–589. doi: 10.1038/331585a0

Pribram, K. H. (1984). "Brain systems and cognitive learning process," in *Animal cognition*, eds H. L. Roitblat, T. G. Bever, and H. S. Terrace (Mahwah, NJ: Lawrence Erlbaum Associates. Inc), 627–656.

Qian, T., Rosenblum, M., and Qiu, H. (2018). *Improving power in group sequential, randomized trials by adjusting for prognostic baseline variables and short-term outcomes*. Baltimore, MD: Johns Hopkins University, Dept. Available online at: https://biostats.bepress.com/jhubiostat/paper285/ (accessed Septermber 21, 2022).

Ramsay, J. O., and Silverman, B. W. (1997). *Functional data analysis*. New York, NY: Springer-Verlag New York. doi: 10.1007/978-1-4757-7107-7

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55. doi: 10.1093/biomet/70.1.41

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66, 688–701. doi: 10.1037/h0037350

Rubin, D. B. (1977). Assignment to a treatment group on the basis of a covariate. *J. Educ. Stat.* 2, 1–26. doi: 10.3102/10769986002001001

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Stat.* 6, 34–58. doi: 10.1214/aos/1176344064

Sartori, G., Masterson, J., and Job, R. (1987). "Direct-route reading and the locus of lexical decision," in *The cognitive neuropsychology of language*, eds M. Coltheart, G. Sartori, and R. Job (Mahwah, NJ: Lawrence Erlbaum Associates. Inc), 59–77.

Schacter, D. L., Cooper, L. A., and Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *J. Exp. Psychol. Gen.* 119, 5–24. doi: 10.1037/0096-3445.119.1.5

Scheffe, H. (1959). *The analysis of variance*. New York, NY: Wiley.

Schwartz, M. F., Saffran, E. M., and Marin, O. S. M. (1980). "Fractionating the reading process in dementia: Evidence for word-specific print-to-sound association," in *Deep dyslexia*, eds M. Coltheart, J. C. Marshall, and K. E. Patterson (London: Routledge & Kegan Paul) 259–269.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Q. J. Exp. Psychol. Sect. A* 38, 619–644. doi: 10.1080/14640748608401617

Squire, L. R. (1987). *Memory and brain*. Oxford: Oxford University Press.

Tulving, E. (1985). How many memory systems are there? *Am. Psychol.* 40, 385–398. doi: 10.1037/0003-066X.40.4.385

Tulving, E., and Schacter, D. L. (1990). Priming and human memory systems. *Science* 247, 301–306. doi: 10.1126/science.2296719

Tulving, E., and Schacter, D. L. (1992). "Priming and memory systems," in *Neuroscience year: Supplement 2 to the encyclopedia of neuroscience*, eds B. Smith and G. Adelman (Boston, MA: Birkhäuser Boston).

Warrington, E. K., and Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia* 12, 419–428. doi: 10.1016/0028-3932(74)90072-4

Warrington, E., and Taylor, A. (1978). Two categorical stages of object recognition. *Perception* 7, 695–705. doi: 10.1068/p070695

West, S. G., and Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychol. Methods* 15, 18–37. doi: 10.1037/a0015917

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* 22, 209–212. doi: 10.1080/01621459.1927.10502953

Wright, P. (1978). Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. *Instr. Sci.* 7, 249–312. doi: 10.1007/BF00120935

Wright, P., and Wilcox, P. (1978). "Following instructions: An exploratory trisection of imperatives," in *Studies in the perception of language*, eds W. J. M. Levelt and G. B. F. d'Arcais (New York, NY: John Wiley & Sons), 129–153.

Wu, C. F. J., and Hamada, M. (2000). *Experiments: Planning, analysis, and parameter design optimization*. New York, NY: John Wiley & Sons.