Which Healthy Eating Nudges Work Best?

A Meta-Analysis of Field Experiments

Romain Cadario

IESEG

Pierre Chandon

INSEAD

Forthcoming in *Marketing Science*

September 21, 2018

Romain Cadario is Assistant Professor of Marketing at IÉSEG School of Management (LEM-CNRS UMR 9221), 1 Parvis de La Défense, 92044 Paris La Défense, France (e-mail: r.cadario@ieseg.fr). Pierre Chandon (pierre.chandon@insead.edu) is the L'Oréal Chaired Professor of Marketing, Innovation and Creativity at INSEAD, Boulevard de Constance, 77300 Fontainebleau, France, and the director of the INSEAD Sorbonne University Behavioral Lab. This authors wish to thank Janet Schwartz, Alexander Chernev, Maria Langlois, Yann Cornil, Gareth Hollands, and those who participated when the authors presented this research at the TCR, BSPA, SJDM, BDFG and BDRM conferences, the University of Toronto and the University of British Columbia. Helpful comments on various aspects of this research were provided by the editor and the review team.

Which Healthy Eating Nudges Work Best?

A Meta-Analysis of Field Experiments

Abstract:
We examine the effectiveness in field settings of seven healthy eating nudges, classified according to whether they are 1) cognitively-oriented, such as “descriptive nutritional labeling,” “evaluative nutritional labeling,” or “visibility enhancements”; 2) affectively-oriented, such as “hedonic enhancements or “healthy eating calls”; or 3) behaviorally-oriented, such as “convenience enhancements” or “size enhancements.” Our multivariate three-level meta-analysis of 299 effect sizes, controlling for eating behavior, population, and study characteristics, yields a standardized mean difference (Cohen’s *d*) of .23 (equivalent to -124 kcal/day). Effect sizes increase as the focus of the nudges shifts from cognition (*d*=.12, -64 kcal) to affect (*d*=.24, -129 kcal) to behavior (*d*=.39, -209 kcal). Interventions are more effective at reducing unhealthy eating than increasing healthy eating or reducing total eating. Effect sizes are larger in the US than in other countries; in restaurants or cafeterias than in grocery stores; and in studies including a control group. Effect sizes are similar for food selection vs. consumption, for children vs. adults, and are independent of study duration. Compared to the typical nudge study (*d*=.12), one implementing the best nudge scenario can expect a six-fold increase in effectiveness (to *d*=.74), with half due to switching from cognitively-oriented to behaviorally-oriented nudges.

Keywords:
Meta-analysis, health, food, field experiment, nudge, choice architecture.

1. Introduction

Unhealthy eating is a key risk factor in non-communicable diseases such as cardiovascular disorders and diabetes, which account for 63% of all deaths worldwide and will cost an estimated US$30 trillion in the next 20 years (Bloom et al. 2012). Traditional approaches to promote healthier eating include economic incentives such as soda taxes (for a recent review, see Afshin et al. 2017), and nutrition education (for a recent review, see Murimi et al. 2017).

More recently, interest has grown in nudge interventions as a spur to healthier eating. Disappointingly, existing meta-analyses have only found average effect sizes ranging from null or weak (e. g., Cecchini and Warin 2016; Littlewood et al. 2016; Long et al. 2015) to moderate (e.g., Arno and Thomas 2016; Hollands et al. 2015). However, these were based on a small number of studies (e.g., 19 for Long et al. 2015), specific foods (e.g., vegetables for Broers et al. 2017), specific settings (e.g., catering outlets for Nikolaou et al. 2014), or included online or laboratory studies (e.g., Sinclair et al. 2014) where effect sizes tend to be different than in field studies (Holden et al. 2016; Long et al. 2015). The literature still lacks a meta-analysis that includes all types of healthy eating nudges, classifies them into a conceptually-grounded framework, and studies their effectiveness in the field after controlling for the effects of important differences in eating behaviors, population, and study characteristics.

Nudges are defined by Thaler and Sunstein (2008, p.6) as "any aspect of the choice architecture that alters people's behavior in a predictable way (1) without forbidding any options or (2) significantly changing their economic incentives. Putting fruit at eye level counts as a nudge; banning junk food does not." By this definition, which has been adopted by influential review papers (e.g., Hollands et al. 2013; Skov et al. 2013), healthy eating nudges encompass a variety of simple, inexpensive, and freedom-preserving modifications to the choice environment such as nutrition labeling or portion size changes. Excluded, however, are traditional educational efforts such as cooking workshops in schools or nutrition pamphlets to parents (Wake 2018) which do not directly change the choice environment and are therefore complements of nudges rather than nudges per se (Sunstein 2018b). Price changes or sales promotions are also not considered nudges because they provide a direct economic incentive. Notwithstanding, in the general discussion we make a comparison between nudges and financial incentives by reanalyzing the data from a recent meta-analysis of price changes on healthy eating (Afshin et al. 2017).

To achieve our goal, we identify seven types of healthy eating nudges classified in three categories: cognitively-oriented, affectively-oriented, and behaviorally-oriented. Our framework also accounts for the type of eating behavior (food selection or actual consumption) and distinguishes between healthy and unhealthy eating. It also considers population characteristics such as age (children vs. adults), consumption setting (onsite cafeterias vs. offsite restaurants, cafes vs. grocery stores), and location of the study (US vs. other countries), as well as characteristics such as the duration of the study and its design. We test this framework with a three-level meta-analysis of 299 effect sizes from 90 articles and 96 field studies.

As shown in Table 1, our work contributes to the many useful existing meta-analyses in terms of (1) scale and scope, (2) method, and (3) categorization of predictors. In terms of scale and scope, we examine more than twice as many effect sizes as the largest existing meta-analysis. This is achieved despite focusing only on field experiments involving actual food choices (vs. perception, evaluation, or choice intentions) and conducted in field settings (onsite cafeterias, offsite eateries, or grocery stores) rather than in a laboratory or online. This allows us to offer guidance to restaurants, supermarket chains, and foodservice companies who want to help their customers eat more healthily but do not know which intervention will work best in their particular context. We also provide guidance for policymakers seeking to forecast the effects that these nudges would have in real-world settings.

Table 1: Comparing meta-analyses of healthy eating nudges

	Scale and scope			Method			Categorization of predictors			
Reference	Effect sizes (K)	Articles (N)	Setting	Hypo-theses	Accounting for repeated observations	Model	Intervention type	Consump-tion vs. selection	Healthy vs. un-healthy	Other control variables
This meta-analysis	299	90	Field only	Yes	Yes (3 levels)	Multivariate (16 *df*)	3 pure and 2 mixed types, (7 subtypes)	Yes	Yes	5 study & population characteristics
Arno and Thomas (2016)	42	36	Field & lab	No	No	Intercept only	1 (all together)	No	No	None
Broers et al. (2017)	14	12	Field & lab	No	No	Intercept only	1 (all together)	No	No	None
Cecchini and Warin (2016)	31	9	Field & lab	No	No	Univariate	1 (only labeling)	No	Yes	None
Holden et al. (2016)	56	20	Field & lab	No	No	Univariate	1 (only size changes)	Yes	No	Manipulation type, field vs. lab
Hollands et al. (2015)	135	69	Field & lab	No	No	Univariate	1 (only size changes)	Yes	Yes	Manipulation type, age, design
Littlewood et al. (2016)	20	14	Field & lab	No	No	Univariate	1 (only labeling)	Yes	No	None
Long et al. (2015)	23	19	Field & lab	No	No	Univariate	1 (only labeling)	No	No	Design
Nikolaou et al. (2014)	10	6	Field only	No	No	Univariate	1 (only labeling)	No	No	None
Robinson et al. (2014)	15	7	Field & lab	No	No	Intercept only	1 (only size changes)	No	No	None
Sinclair et al. (2014)	42	17	Field & lab	No	No	Univariate	2 (desc. vs. eval. label.)	Yes	No	None
Zlatevska et al. (2014)	104	30	Field & lab	No	No	Univariate	1 (only size changes)	No	Yes	Field vs. lab, age, sex, BMI

Methodologically, our meta-analysis differs from earlier ones on three levels. First, we formulate hypotheses about which healthy eating nudges work best, and the effects of eating behavior, the study population as well as study design. Second, to reduce the risk of confounds from univariate analyses, we employ a multivariate model incorporating all predictors simultaneously. Third, we include a three-level analysis to take into account the hierarchical structure of our data. Finally, as Table 1 shows, we use a more granular predictor structure compared to existing meta-analyses which either estimate the effect size of a single type of healthy eating nudge or compare the effect of a single difference (e.g., descriptive vs. evaluative labeling).

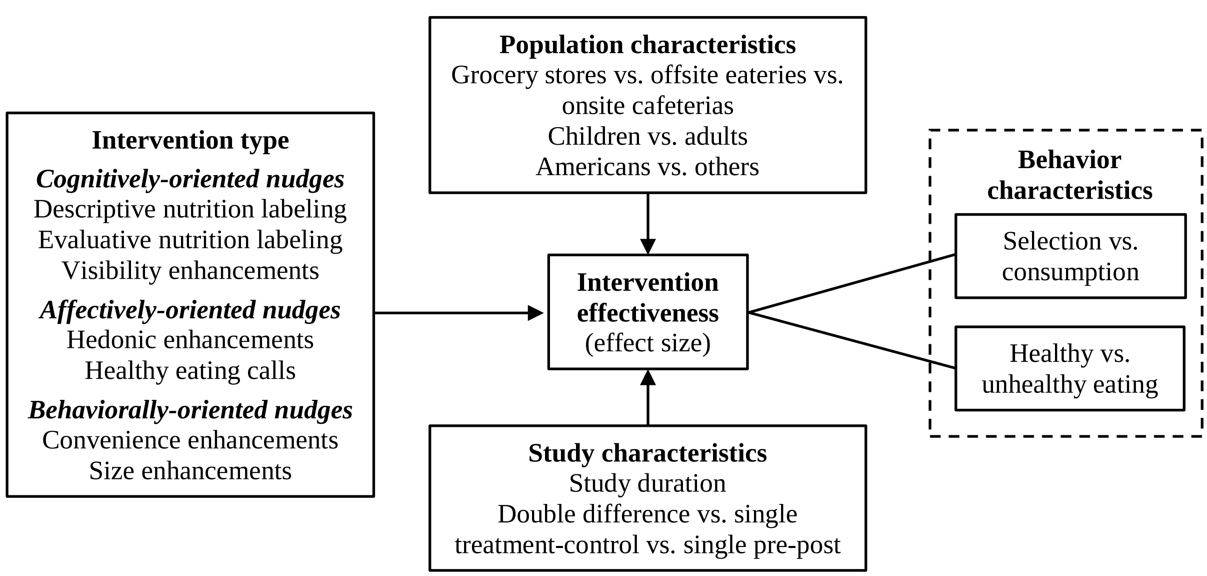
2. Conceptual framework

As shown in Table 2, the existing frameworks of healthy eating nudges are either based on the intervention instrument (e.g., a label, size of plate) or based on the hypothesized mechanisms of action (e.g., attention vs. social norm). Over the years, they have tended to focus on finer and finer distinctions. For example, Hollands et al. (2013) initially distinguished three types of nudges, whereas even the simplified version of their more recent TIPPME typology contains 18 categories (Hollands et al. 2017). Our framework keeps a fine level of analysis by distinguishing between seven types of nudges, which is useful for policymakers or practitioners who want to know the effect size of a particular intervention instrument, and then groups them into three theory-grounded categories, which allows us to make predictions about their effectiveness. As shown in Figure 1, our framework also differs from the existing literature by also taking into account not just the type of nudge but also the type of eating behavior (selection vs. consumption, healthy vs. unhealthy eating); population characteristics (location, age and country); as well as study characteristics (duration and design).

Table 2: A Comparison of frameworks of healthy eating nudges

Authors	Framework	# nudges	# levels	Basis for categorization	Prediction and test
Dolan et al. (2012)	MINDSPACE: Messenger, Incentives, Norms, Defaults, Salience, Priming, Affect, Commitments, Ego	9	1	Instrument	No
Ly et al. (2013)	A practitioner's guide to nudging	12	1	Mechanism	No
Chance et al. (2014)	4 Ps: Possibilities, Process, Persuasion, Person	4	1	Mechanism	No
BIT (2014)	EAST: Easy, Attractive, Social, Timely	4	1	Mechanism	No
Wansink (2015)	CAN: Convenience, Attractiveness, Normativeness	3	1	Mechanism	No
Kraak et al. (2017)	8 Ps: Place, Profile, Portion, Pricing, Promotion, healthy default Picks, Prompting or priming, Proximity	8	1	Instrument	No
Hollands et al. (2013)	TIPPME: Typology of Interventions in Proximal Physical Micro-Environments	9	1	Instrument	No
Hollands et al. (2017)	TIPPME: Updated version	18	1	Instrument	No
	This framework		2	Both	Yes
	Conceptual level: Cognitively-oriented, Affectively-oriented and Behaviorally-oriented	3		Mechanism	
	Nudge level: Descriptive nutrition labeling, Evaluative nutrition labeling, Visibility enhancements, Hedonic enhancements, Healthy eating calls, Convenience enhancements, Size enhancements.	7		Instrument	

Figure 1: Conceptual framework



2.1. Intervention type

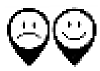
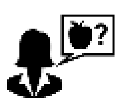
Conceptual level. We draw on the classic tripartite classification of mental activities into cognition, affect, and behavior (or conation), which dates back to eighteenth-century German philosophy (Hilgard 1980). The trilogy of mind has long been adopted in psychology and marketing to understand consumer behavior and predict the effectiveness of marketing actions (Barry and Howard 1990; Breckler 1984; Hanssens et al. 2014; Oliver 1999; Srinivasan et al. 2010). As shown in Figure 1, we distinguish between 1) cognitively-oriented interventions that seek to influence what consumers know; 2) affectively-oriented interventions that seek to influence how consumers feel, without necessarily changing what they know; and 3) behaviorally-oriented interventions that seek to influence what consumers do (i.e., their motor responses), without necessarily changing what they know or how they feel. Within each type, we further distinguish subtypes that share similar characteristics and have been tested by enough studies to enable a meaningful meta-analysis. This subcategorization is based on existing classifications, such as the distinction between descriptive and evaluative nutritional labeling (Fernandes et al. 2016; Sinclair et al. 2014). We acknowledge that the cognitive-affective-behavioral categorization is not iron-clad and that it is possible for some nudges to have features that straddle multiple categories. In the discussion we examine how changes in the categorization affect the results.

Cognitively-oriented interventions. As described in Table 3, we identify three types of cognitively-oriented interventions. The first type, "descriptive nutritional labeling," provides calorie count or information about other nutrients, be it on menus or menu boards in restaurants, or on labels on the food packaging or near the foods in self-service cafeteria and grocery stores. The second type, "evaluative nutritional labeling," typically (but not always) provides nutrition information but also helps consumers interpret it through color coding (e.g., red, yellow, green as nutritive value increases) or by adding special symbols or marks (e.g., heart-healthy logos or smileys on menus). Although the third type, “visibility enhancement,” does not directly provide health or nutrition information, it is a cognitively-oriented intervention because it informs consumers of the availability of healthy options by increasing their visibility on grocery or cafeteria shelves (e.g., placing healthy options at eye level and unhealthy options on the bottom shelf) or on restaurant menus (e.g., placing healthy options on the first page and burying unhealthy ones in the middle). Because people typically look at only a subset of the options available to them (Chandon et al. 2009), making healthy options more visible and unhealthy options less visible thus changes the information about the range of healthiness or nutrition options they can choose from.

Affectively-oriented interventions. The first type of affectively-oriented interventions, which we call “hedonic enhancements,” seeks to increase the hedonic appeal of healthy options by using vivid hedonic descriptions (e.g., “Twisted citrus-glazed carrots”) or attractive displays, photos, or containers (e.g., “pyramids of fruits”). To date, no field experiment has sought to reduce hedonic enhancements expectations for unhealthy options by using disparaging descriptions or unattractive photos. These interventions are affectively-oriented because, rather than focusing on informing consumers about the nutritional quality of food options or their likely health impact, they focus on the more affectively-oriented hedonic consequences of eating the food.

The second type of affectively-oriented interventions, “healthy eating calls,” directly encourages people to better. This can be done by placing signs or stickers (e.g., “Make a fresh choice,” or “Have a tossed salad for lunch!”) or by asking foodservice staff to verbally encourage people to choose a healthy option (e.g., asking “which vegetable would you like to have for lunch?” when children can choose none) or to change their unhealthy choices (e.g., “Your meal doesn’t look like a balanced meal” or “Would you like to take half a portion of your side dish?”).

Table 3: Categorization of nudge interventions

Type		**Target: Healthy eating** (k = 180)	**Target: Unhealthy eating** (k = 79)
Descriptive nutritional labeling (k = 34)		Calorie or nutrition labeling (Auchincloss et al. 2013; Bollinger et al. 2011; Brissette et al. 2013; Chu et al. 2009; Downs et al. 2013; Dubbert et al. 1984; Dumanovsky et al. 2011; Elbel et al. 2011; Elbel et al. 2009; Elbel et al. 2013; Ellison et al. 2013; Finkelstein et al. 2011; Krieger et al. 2013; Pulos and Leng 2010; Roberto et al. 2010; Tandon et al. 2011; Vanderlee and Hammond 2014; Vasiljevic et al. 2018; Webb et al. 2011)	
Evaluative nutritional labeling (k = 43)		Green stickers, smileys, "heart healthy" logos (Cawley et al. 2015; Ensaff et al. 2015; Gaigi et al. 2015; Hoefkens et al. 2011; Kiesel and Villas-Boas 2013; Levin 1996; Levy et al. 2012; Mazza et al. 2017; Ogawa et al. 2011; Olstad et al. 2015; Reicks et al. 2012; Thorndike et al. 2014; Thorndike et al. 2012)	Red stickers next to unhealthier options (Crockett et al. 2014; Hoefkens et al. 2011; Levy et al. 2012; Olstad et al. 2015; Shah et al. 2014; Thorndike et al. 2014; Thorndike et al. 2012)
Visibility enhancements (k = 25)		Healthier options more visible: e.g., eye-level shelf position, transparent containers, placed first on menus, placed near cash register (Bartholomew and Jowers 2006; Cohen et al. 2015; Dayan and Bar-Hillel 2011; Ensaff et al. 2015; Foster et al. 2014; Gamburzew et al. 2016; Geaney et al. 2016; Hanks et al. 2013; Kroese et al. 2016; Levy et al. 2012; Meyers and Stunkard 1980; Perry et al. 2004; Policastro et al. 2015)	Unhealthier options less visible (not eye-level shelf positions, middle of the menu), previous unhealthier consumption more visible: e.g., leftover chicken wings un-bussed (Bartholomew and Jowers 2006; Baskin et al. 2016; Dayan and Bar-Hillel 2011; Meyers and Stunkard 1980; Wansink and Payne 2007)
Hedonic enhancements (k = 7)		Vivid hedonic descriptions (e.g., "Dynamite beets", "Twisted citrus-glazed carrots") or attractive displays, photos, or containers (Cohen et al. 2015; Ensaff et al. 2015; Greene et al. 2017; Hanks et al. 2013; Morizet et al. 2012; Olstad et al. 2014; Perry et al. 2004; Turnwald et al. 2017; Wilson et al. 2017)	
Healthy eating calls (k = 42)		Written or oral injunction to choose healthier options: e.g., "Make a fresh choice" or "Have a tossed salad for lunch!" (Anzman-Frasca et al. 2018; Buscher et al. 2001; Cohen et al. 2015; Ensaff et al. 2015; Greene et al. 2017; Hanks et al. 2013; Hubbard et al. 2015; Mayer et al. 1986; Mazza et al. 2017; Mollen et al. 2013; Perry et al. 2004; Policastro et al. 2017; Schwartz 2007; Thomas et al. 2017; van Kleef et al. 2015)	Written or oral injunctions to change unhealthy choices: e.g., "Your meal doesn't look balanced" or "Would you like to take a half portion?" (Donnelly et al. 2018; Freedman 2011; Miller et al. 2016; Mollen et al. 2013; Schwartz et al. 2012)
Convenience enhancements (k = 65)		Healthier options are easier to select or consume: e.g., more convenient utensils, "grab and go" line, pre-sliced, pre-portioned, or pre-served food, healthy food as default or placed earlier in a cafeteria line when the tray is free (Adams et al. 2005; Buscher et al. 2001; Cohen et al. 2015; De Bondt et al. 2017; de Wijk et al. 2016; Elsbernd et al. 2016; Friis et al. 2017; Goto et al. 2013; Greene et al. 2017; Hanks et al. 2012; Lachat et al. 2009; Olstad et al. 2014; Redden et al. 2015; Rozin et al. 2011; Steenhuis et al. 2004; Tal and Wansink 2015; Wansink et al. 2016; Wansink and Hanks 2013; Wansink et al. 2013; Wilson et al. 2017)	Unhealthier options are less convenient to select or consume: e.g., making unhealthy food less accessible or harder to reach, less convenient serving utensils, or placed later in a cafeteria line when the tray is full and returning already-chosen healthier food requires backtracking and inconveniencing others (Hanks et al. 2012; Mishra et al. 2012; Rozin et al. 2011; Wansink and Hanks 2013)
Size enhancements (k = 17)		Larger plates for healthier options (DiSantis et al. 2013)	Smaller plates or portions for unhealthy options (Diliberti et al. 2004; DiSantis et al. 2013; Freedman and Brochado 2010; van Ittersum and Wansink 2013; Wansink and Kim 2005; Wansink and van Ittersum 2013; Wansink et al. 2006; Wansink et al. 2014)

Note: Articles cited in multiple categories either implemented different types of interventions or implemented interventions combining different types of nudges.

Such injunctions are affectively-oriented because, rather than informing people about the healthiness of the food options available, they seek to change people's eating goals, which are inherently affect laden (Shiv and Fedorikhin 1999). This is particularly the case when these injunctions are made verbally by the waiters or the "lunch ladies", which create strong affective responses (Herman et al. 2003; McFerran et al. 2010).

Behaviorally-oriented interventions. The third group consists of two types of interventions that aim to impact people's behaviors without necessarily influencing what they know or how they feel, often without people being aware of their existence. "Convenience enhancements" make it physically easier for people to select healthy options (e.g., by making them the default option or placing them in faster "grab & go" cafeteria lines) or to consume them (e.g., by pre-slicing fruits or pre-serving vegetables), or make it more cumbersome to select or consume unhealthy options (e.g., by placing them later in the cafeteria line when trays are already full or by providing less convenient serving utensils). The second type, which we call "size enhancements," modifies the size of the plate, bowl, or glass, or the size of pre-plated portions, either increasing the amount of healthy food they contain or, most commonly, reducing the amount of unhealthy food. Another difference is that visual attention is necessary for cognitively-oriented interventions to influence behaviors but not for behaviorally-oriented interventions. In fact, plate and portion size changes have stronger effects when people do not pay attention to them (van Ittersum and Wansink 2012; Zlatevska et al. 2014) and even influence food intake when people are eating in the dark (Scheibehenne et al. 2010). Unlike cognitively-oriented and affectively-oriented interventions which influence food choices through vision or audition, behaviorally-oriented interventions influence eating primarily through physical interactions, which leads to different food decisions (for a review, see Krishna 2012). For example, Hagen et al. (2016) showed that physical involvement in obtaining food (e.g., when people need to serve themselves rather than being served, or need to touch, unwrap, modify the food vs. if the food is pre-portioned and pre-plated) leads to healthier food choice because it increases attribution of responsibility for food consumption.

Hypotheses. For food choices, cognitive factors tend to be less predictive of choice than affective factors, which strongly influence even restrained eaters who eat according to cognitive rules (Macht 2008). When asked about what drives their food choices, Americans and Europeans place affective factors like taste well ahead of cognitive factors like nutrition or weight control (Glanz et al. 1998; Januszewska et al. 2011). Even interventions that successfully change beliefs about the health consequences of behaviors often fail to lead to meaningful behavioral changes (Carpenter 2010; Sniehotta et al. 2014).

Because eating is largely habitual and prone to self-regulation failures, affective factors tend to be less predictive of food choices than behavioral factors (Herman and Polivy 2008; Ouellette and Wood 1998). For example, directly changing the eating environment (e.g., avoiding exposure to tempting food) is a more successful self-control strategy than cognitive or affective strategies such as thinking about health and nutrition or relying on willpower (Duckworth et al. 2016; Wansink and Chandon 2014). Similar conclusions were reached in a study of sales elasticity for 74 mostly food brands (Srinivasan et al. 2010), which found that changes in distribution (a behaviorally-oriented intervention) had a larger impact than changes in advertising (a cognitive or affective intervention), and that affective changes in liking were more predictive of brand choice than cognitive changes in awareness. Further support comes from a recent review which found that "interventions that facilitate vaccination directly by leveraging, but not trying to change, what people think and feel are by far the most effective" whereas "few randomized trials have successfully changed what people think and feel about vaccines, and those few that succeeded were minimally effective in increasing uptake" (Brewer et al. 2017). We therefore expect the effectiveness of healthy eating interventions to increase as their focus switches from cognition to affect and to behavior.

2.2. The role of eating behavior type

As shown in Figure 1, we differentiate between different types of eating behaviors. Some studies measure actual food consumption, others only capture food selection (e.g., the purchase of food in a grocery store, cafeteria, or restaurant) without knowing whether the food was entirely consumed. One might expect larger effect sizes for selection than for consumption if some consumers, after being nudged to try a healthier food, are disappointed by its taste and only consume part of it. On the other hand, people usually have a stronger preference for *what* to eat relative to *how much* to eat (Wansink and Chandon 2014). Hence, we expect no differences between studies measuring selection and those measuring actual consumption. This hypothesis is consistent with the results of existing meta-analyses of specific types of healthy eating nudges (Holden et al. 2016; Hollands et al. 2015; Littlewood et al. 2016; Sinclair et al. 2014).

We compare studies that measure total eating (e.g., total calorie content of the food selected or consumed) or focus on the selection or consumption of healthy or unhealthy foods. We expect smaller effect sizes when the dependent variable is the total amount of food ordered or consumed for two reasons. The first is that people must eat—it is difficult, psychologically and physiologically, to sustain an imbalance between energy intake and energy expenditure. In contrast, people have more flexibility in choosing how to allocate their total calorie intake between healthy and unhealthy foods. Second, healthy foods have calories too, so replacing unhealthy food with healthier options—although clearly a form of healthier eating—does not necessarily mean a reduction in the total quantity of food ordered or consumed. This hypothesis is consistent with an existing meta-analysis of interpretive nutrition labels, which found that they were more effective in helping consumers in choosing healthier products than in changing total intake (Cecchini and Warin 2016).

Finally, we hypothesize that interventions aimed at reducing unhealthy eating have a stronger effect size than those aimed at promoting healthy eating. This prediction is based on the fact that more than two thirds of Americans are overweight or obese, and about half of the latter are actively trying to lose weight in any given year (Snook et al. 2017). Dieters should therefore be particularly receptive to interventions that reduce calorie intake, which is most effectively accomplished by reducing the consumption of unhealthy foods rather than by increasing healthy food consumption. Indeed, dieters and overweight consumers may be wary of increasing their intake of foods presented as "healthy," which often actually have a high energy density (Chernev 2011; Wansink and Chandon 2006). More generally, people often exhibit dynamically inconsistent preferences, choosing unhealthy food in the short term and regretting it later (Prelec and Loewenstein 1998; Wertenbroch 1998). Interventions that help resist the temptation of unhealthy food should therefore be particularly attractive to the many people who have long-term healthy eating goals and are aware that they need help resisting unhealthy foods. Our hypothesis is consistent with the results of two existing meta-analyses which found that the effectiveness of plate and portion size changes was higher for unhealthy foods compared to healthy foods (Hollands et al. 2015; Zlatevska et al. 2014).

2.3. The role of population characteristics

We distinguish between studies conducted in onsite eating settings (e.g., university or worksite cafeterias), offsite eating eateries (e.g., restaurants, cinemas, cafes), and grocery stores. We expect weaker effects in grocery stores compared to the other two settings. This is because it should be easier to respond to healthy eating nudges when choosing for oneself, from among a limited number of options, and for a single immediate consumption in a cafeteria or a restaurant, than when choosing for the entire family, from among a huge variety of tempting options, and for multiple consumption occasions in a grocery store. This hypothesis is consistent with research showing that uncertainty about future preferences (when buying for the entire family, for example) increases the variety of choices (Walsh 1995), thereby mitigating the effects of nudges. It is also consistent with the systematic review conducted by Seymour et al. (2004), which concluded that healthy eating nudges have weaker effects in grocery stores than in restaurants or in university or worksite cafeterias.

Prior research has established that adults are more interested in nutrition than children (Croll et al. 2001), and also more sensitive to portion size changes (Hollands et al. 2015; Zlatevska et al. 2014). We thus expect them to be more responsive to all types of interventions than children.

Finally, we expect to find higher effect sizes in studies conducted in the US than in other countries, for three reasons: the higher proportion of overweight people in the US, the larger size of portions there (Rozin et al. 2003), and Americans' higher interest in and knowledge of the health consequences of eating (Rozin et al. 1999), and their greater reliance on external than internal cues when making food decisions (Wansink et al. 2007).

2.4. The role of study characteristics

We include two study characteristics as control variables. The duration of the intervention varies from a single exposure and consumption to interventions implemented over many months. In field experiments, treatment effects usually decay over time as people revert to habitual behavior (Brandon et al. 2017). However, a long study duration can only capture the evolution (decay or strengthening) of the effects of an intervention over time if it entails repeated exposure by the same people. Although this may be the case for longitudinal studies in worksite cafeterias, for example, it may not be true for restaurants patronized by different customers over time. In the absence of information about the level of repetition, we only include study duration as a covariate. Note that not enough studies measure post-intervention effects to allow us to estimate carryover effects.

We distinguish between studies using a pre-post design without control, those using a single-difference (treatment vs. control) design, and those using a double-difference design (before vs. after in control and treated locations). Although designs with stronger levels of control should have a lower statistical bias in the estimation of the effect size, the type of design itself should not influence the size of the effect, hence we cannot formulate hypotheses about the effect of design on effect sizes and include this factor simply as a control variable.

3. Data collection

3.1. Inclusion criteria

Appendix A provides detailed information on the search strategy, including the SPICE (Setting, Population, Intervention, Comparison, Valuation) framework (Booth 2006) for the selection of keywords and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Moher et al. 2009) flow diagram showing the number of articles included and excluded. Briefly, we searched for relevant articles published in scholarly journals until January 1 2017 through keyword searches on Science Direct, PubMed, and Google Scholar. We also examined all the references from 11 meta-analyses (Arno and Thomas 2016; Broers et al. 2017; Cecchini and Warin 2016; Holden et al. 2016; Hollands et al. 2015; Littlewood et al. 2016; Long et al. 2015; Nikolaou et al. 2014; Robinson et al. 2014; Sinclair et al. 2014; Zlatevska et al. 2014) and 7 systematic reviews (Bucher et al. 2016; Hollands et al. 2013; Nornberg et al. 2016; Roy et al. 2015; Skov et al. 2013; Thapaa and Lyford 2014; Wilson et al. 2016). Both authors developed the protocol detailing the search and inclusion criteria, coding categories for predictors, and computation rules. The first author was trained to code all the studies and was responsible for extracting data. The second author checked the results, and disagreements were solved through discussion.

Next, we contacted all authors of the studies cited in the December 2017 draft and asked them whether they agreed with our categorization. We received new data and feedback that allowed us to fine tune our calculation of effect size and our categorization choices. Note that we did not include any retracted publications. Including three recently-retracted publications (Wansink et al. 2012a; Wansink et al. 2012b; Wansink et al. 2008) did not affect the results in any way. Moreover, in light of recent criticisms regarding some of the studies conducted by the Cornell Food and Brand Lab (Robinson 2017), we estimated the multivariate model including an additional binary variable controlling for the 51 observations originating from this lab. This variable was insignificant (p = .33).

To be included in the meta-analysis, the study had to test a nudge intervention consistent with our definition (e.g., not a price change nor a nutrition education campaign). Since our focus was on pure nudges, studies (or conditions in studies) combining nudges with changes in economic incentives or education efforts were not included. The intervention had to be tested in a field experiment in which participants made food decisions where participants aren't usually aware that their food choices are being monitored. We excluded studies conducted in a laboratory or online. This is important because previous reviews found marked differences between studies conducted in the field and those conducted in a laboratory or online (Long et al. 2015), and between studies conducted with aware or unaware participants (Holden et al. 2016). Finally, the dependent variable of the study had to provide an objective measure of food selection or consumption (either in weight or energy). We rejected studies relying on consumption intentions as well as field studies without a control condition or a pre-intervention baseline condition.

Overall, the meta-analysis includes 299 effect sizes derived from 96 studies published in 90 articles. The number of observations per study ranged from 36 to 100 million, with a median of 1,231. The total number of original observations was 133.6 million with two outlier articles: one with 100 million transactions (Bollinger et al. 2011) and one with 29 million transactions (Nikolova and Inman 2015). The other 88 articles represent more than 4.6 million original observations.

3.2. Effect size calculations

We calculated the effect sizes of studies with a binary outcome, such as the number of participants who chose a healthy option, by computing the log odds ratio or by obtaining it directly from the paper in the few cases when it was available. We computed the odds ratio as the odds of a healthy selection in the treatment group divided by the odds of a healthy selection in the control group. We then computed its standard error. We calculated the effect sizes of studies with a continuous dependent variable, such as unhealthy food intake, by computing the standardized mean difference, except in the few instances when the standardized mean difference, also known as Cohen's (1988) *d*, was already reported in the paper. We computed the *d* value as the mean difference in consumption between the treatment and control condition, divided by the pooled standard deviation. Given that we had two different effect-size metrics, we converted the log odds ratio into *d* using the formula proposed by Borenstein et al. (2009). After the conversion, the 157 effects sizes originally calculated as log odds ratio were not statistically different from the 142 effect sizes computed as *d* ($p = .43$). Hence, we report Cohen's *d* in the paper because it is the most common measure and allows direct comparisons with other meta-analyses.

The most common unreported data was the sample size per experimental condition (e.g., intervention vs. control). We contacted the authors to obtain this information but were not always successful. When only the total sample was reported, we divided the total number of observations by the number of conditions. When only the number of observations in the control group or in the intervention group was reported, we used the same number for the other group. Whenever several assumptions were possible, we conservatively chose the assumption that yielded the smaller effect size or the largest standard error.

When results were reported separately for each food in the same study (e.g., Ensaff et al. 2015), we calculated separate effect sizes per food and accounted for their dependence in the statistical analysis. We also computed separate effect sizes for the few studies (e.g., Schwartz 2007) that measured both food selection (e.g., putting a food item on a cafeteria tray) and consumption (e.g., how much of it was consumed). When a study had a two-phase intervention (e.g., one intervention during the first phase and then another intervention during a later phase, Thorndike et al. 2012), we computed separate effect sizes for each phase and compared both phases to the baseline period. When a study tested multiple interventions separately (e.g., Mollen et al. 2013), we computed separate effect sizes for each intervention. Because only two studies reported results separately for men and women (Baskin et al. 2016; Wansink and Payne 2007), we could not examine the role of gender in the meta-analysis and calculated the average effect size for these two studies.

3.3. Coding

We categorized the intervention into one of the seven types discussed. To check the validity of the categorization, we emailed all the authors and adjusted the categorization in the few cases when they disagreed with our initial categorization. Field experiments that implemented multiple interventions at once were treated separately. There are not enough studies testing each possible combination of cognitively-oriented, affectively-oriented, and behaviorally-oriented interventions (for example, only one study mixed a cognitively-oriented and a behaviorally-oriented intervention), so we had to rely on an ad hoc coding of “mixed interventions” based on their frequency in our sample. The first type of mixed intervention consists of studies mixing cognitively-oriented interventions with affectively-oriented and/or behaviorally-oriented interventions (e.g., descriptive nutrition labeling and hedonic enhancements). We named them "mixed: cognitive present." The second type of mixed interventions consists of studies combining affectively-oriented and behaviorally-oriented interventions. We named them "mixed: cognitive absent." We therefore have five categories for nudge interventions: three for pure cognitively-oriented, affectively-oriented, and behaviorally-oriented interventions and two for mixed interventions (mixed: cognitive present and mixed: cognitive absent).

When the dependent variable of the study was the total amount of food selected or consumed, it was categorized as "total eating." When it was the selection or consumption of fruits, vegetables and water, or foods color-coded green in the study, we categorized it as "healthy eating." We categorized the selection or consumption of calorie-dense and nutrient-poor foods such as desserts or sodas and those color-coded red in studies as "unhealthy eating." We created a fourth category ("mixed eating") for foods that could not be categorized as healthy or unhealthy or which were color-coded yellow by the researcher (rather than green or red). Because our goal is to examine healthy eating, we reverse coded the effect sizes for unhealthy eating and for total eating.

We coded population characteristics according to where the study was conducted (school or workplace onsite cafeterias; offsite restaurants, cinemas, or cafes; or grocery stores). We coded whether the participants were children or adults. Finally, we distinguished between studies conducted in the United States and those conducted in other countries (Belgium, Canada, France, Ireland, Israel, Japan, Netherlands, and the United Kingdom). The number of studies in these other countries was too low to enable a more refined level of analysis.

Regarding study characteristics, we measured the duration of the treatment as the number of weeks of the intervention period. For example, if a study with a pre-post design measured food choices in the 4 weeks prior to the intervention and in the 2 weeks during which the intervention was implemented, study duration is coded as 2 weeks. Last, we distinguished between "double-difference" designs (which assigned participants to two independent control and treatment conditions, with observations before and after the intervention), "single-difference treatment-control" designs (which assigned respondents to two independent control and treatment conditions), and "single-difference pre-post" designs (which used a pre-post study design without a control group, comparing observations before and after the intervention). Note that all are quasi-experiments because the randomization was not done at the participant level but at the level of the store, restaurant, cafeteria, or at best, cafeteria line.

4. Analyses and results

As indicated in Table 1, the 11 existing meta-analyses used a standard two-level meta-analytical model (Borenstein et al. 2009). In contrast, we used a three-level model (Cheung 2014), which accounts for the fact that some observations come from the same field experiment (e.g., studies testing two types of interventions or measuring their impact on healthy and unhealthy foods separately). We estimated a mixed-effects three-level meta-analytic model with the "metafor" R package provided in Viechtbauer (2010), via maximum likelihood.

4.1. Average meta-analytical effect: Intercept-only model

Let y_{ij} be the i^{th} effect size in the j^{th} study. The equations from the three levels are:

$$y_{ij} = \lambda_{ij} + e_{ij} \tag{1}$$

$$\lambda_{ij} = \kappa_j + u_{(2)ij} \tag{2}$$

$$\kappa_j = d_0 + u_{(3)j} \tag{3}$$

where λ_{ij} is the true effect size and $\mathrm{Var}(e_{ij}) = v_{ij}$ is the known sampling variance in the i^{th} effect size in the j^{th} study, κ_j is the average effect size in the j^{th} study, and $\mathrm{Var}(u_{(2)ij}) = \tau^2_{(2)}$ captures the heterogeneity in effect sizes between different eating behaviors (e.g., selection or consumption, healthy or unhealthy food) within the same study, when more than one outcome was measured. d_0 is the meta-analytic effect size estimated across all studies, and $\mathrm{Var}(\mathrm{u}_{(3)\mathrm{j}}) = \tau^2_{(3)}$ captures the heterogeneity between studies after controlling for the presence of multiple observations at level 2. The three equations can be combined as follows:

$$y_{ij} = d_0 + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{4}$$

We assessed the magnitude of effect size heterogeneity through the I^2 index (Higgins and Thompson 2002). We also report the decomposition of heterogeneity within-studies $I^2_{(2)}$ and between-studies $I^2_{(3)}$ as derived in Cheung (2014). Heterogeneity is considered to be low if the I^2 index is below 25%, medium if it is between 25% and 75%, and high if it is above 75% (Higgins and Thompson 2002).

The standard two-level model yields a statistically significant average effect size (d = .22, z = 13.45, p < .001) with a very large amount of heterogeneity (I^2 = 99.9%). The proposed three-level model fits the data significantly better than the two-level model ($\chi^2(1)$ = 100.5, p < .001) and yields a slightly larger estimate of the average effect size (d = .27, z = 9.53, p < .001). This effect size is considered small as per Cohen's (1988) definition. The three-level random-effects model shows that the total heterogeneity is lower within studies ($I^2_{(2)}$ = 25.7%) than between studies ($I^2_{(3)}$ = 74.2%). Additional analyses reported in detail in Appendix B (p-curve, trim and fill, sensitivity analyses) suggest minimal publication bias (Rothstein et al. 2006).

4.2. Influence of predictors: univariate vs. multivariate model

As shown in Table 1, the 11 existing meta-analyses on healthy eating nudges use univariate meta-analyses (i.e., they separately test the impact of each predictor/outcome). Univariate analyses exclude control variables and increase the possibility that significant differences are due to confounds. Multivariate models help to provide estimates with better statistical properties, as well as reduce the risk of bias such that a significant result in univariate analyses may not hold using the multivariate model (Jackson et al. 2011). We performed both univariate and multivariate analyses and confirm that the latter lead to a higher model fit as well as more conservative average effect sizes (Figure 2, Appendix C).

Univariate models. We estimated one univariate meta-regression for each predictor *x*. These univariate analyses provide benchmark values which can be compared to the estimates obtained in the full multivariate model. When the predictor is categorical (for intervention type, for example), the univariate model in Equation 5 estimates *S* coefficients β_s corresponding to each level of the categorical predictor, without any covariate. The third and fourth column of Figure 2 show, respectively, the mean and standard errors of the β_s coefficients, which capture the effect size for each level of the categorical variables as estimated in a univariate regression.

$$y_{ij} = \sum_{1}^{S} \beta_s x_{ij} + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{5}$$

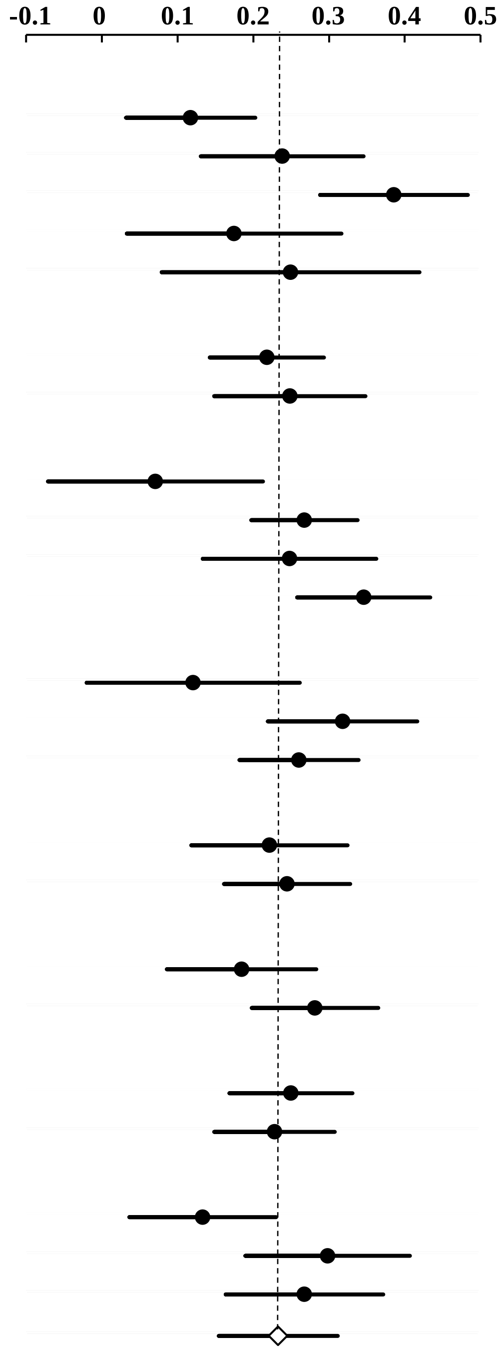
For study duration, which is a continuous variable measured in weeks, the univariate model estimated one intercept and one parameter, as shown in Equation 6. To provide a point estimate for short and long study durations, Figure 2 shows the model's intercept estimated at the first quartile (1 week) and third quartile (15 weeks) of the distribution of duration.

$$y_{ij} = d_0 + \beta_{Duration} Duration_{ij} + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{6}$$

Figure 2: Effect sizes in the univariate and full multivariate models

		Univariate		Multivariate	
	k	*d*	*se*	*d*	*se*
Intervention type					
Cognitively-oriented	116	.12*	.04	.12*	.04
Affectively-oriented	49	.30*	.05	.24*	.06
Behaviorally-oriented	82	.49*	.04	.39*	.05
Mixed: cognitive present	43	.17*	.08	.17*	.08
Mixed: cognitive absent	9	.31*	.09	.25*	.09
Eating behavior type					
Selection	251	.25*	.03	.22*	.04
Consumption	48	.34*	.05	.25*	.05
Eating behavior measure					
Total eating	23	.03	.06	.07	.08
Healthy eating	180	.27*	.03	.27*	.04
Mixed eating	17	.26*	.05	.25*	.06
Unhealthy eating	79	.38*	.04	.35*	.05
Population setting					
Grocery stores	38	.12	.08	.13	.07
Offsite eateries	76	.26*	.05	.32*	.05
Onsite cafeterias	185	.30*	.04	.26*	.04
Population age					
Children	93	.33*	.05	.22*	.05
Adults	206	.24*	.03	.25*	.04
Population country					
Other countries	77	.20*	.06	.19*	.05
US	222	.29*	.03	.28*	.04
Study duration					
Short (1 week)	299	.33*	.03	.25*	.04
Long (15 weeks)	299	.25*	.03	.23*	.04
Study design					
Single pre-post	176	.22*	.04	.13*	.05
Single treatment-control	70	.45*	.06	.30*	.06
Double-difference	53	.20*	.06	.27*	.05
Average effect	**299**	**.27***	**.03**	**.23***	**.04**

Forrest plot (mean and 95% confidence interval from the multivariate model)



* $p < .05$.

Univariate analyses suggested that the effectiveness of healthy eating nudges varies by intervention type (R^2 = 32%, $\chi^2(4) = 39$, $p < .001$). Figure 2 shows that the estimated effect sizes in the univariate analysis of intervention type vary between d = .12 for cognitively-oriented interventions and d = .51 for behaviorally-oriented interventions, which are all statistically different from zero. Univariate analyses found no difference between selection and consumption (R^2 = 5%, $\chi^2(1) = 3.53$, p = .06) but significant effect depending on the behavior measured (total eating, healthy eating, mixed eating, or unhealthy eating: R^2 = 18%, $\chi^2(3) = 28$, p < .001). They found no differences between adults and children (R^2 = 3%, $\chi^2(1) = 2.09$, p = .15); between grocery stores, offsite eateries, or onsite cafeterias (R^2 = 5%, $\chi^2(2) = 3.97$, p = .14), or between studies conducted in the US and outside the US (R^2 = 2%, $\chi^2(1) = 1.93$, p = .16). Finally, they found a significant effect of duration (R^2 = 12%, $\chi^2(1) = 10.29$, p < .01) and study design (R^2 = 12%, $\chi^2(2) = 13$, p < .01).

Multivariate model. We estimated a full model with all the predictors entered simultaneously as shown in equation 7, where s corresponds to the categories for each predictor k. The multivariate model explained 49% of the variance, a significant improvement over the intercept-only model ($\chi^2(15) = 78$, p < .001). It is also a significant improvement over the best univariate model, the one with intervention type ($\chi^2(7) = 40$, p < .001). This suggests that our overall conceptual framework captured a substantial variation in the effect sizes, much more than any separate univariate model.

$$y_{ij} = d_0 + \sum_{s}^{S-1} \sum_{k}^{K} \beta_{ks} x_{kij} + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{7}$$

The fifth and sixth columns in Figure 2 show the multivariate effect sizes estimated for each level of a given predictor, when all the other predictors are at their mean value. Overall, the multivariate model yielded effect sizes that are 9.4% smaller than those of the univariate models.

After controlling for all covariates, the average effect size computed across all 299 observations shrinks slightly, from d =.27 to d = .23, but remains significantly different from zero (z = 5.83, p < .001). Other effect sizes show stronger reductions. The reduction is particularly strong for the largest effect sizes, like the estimate for behaviorally-oriented interventions (which shrinks from .49 to .39). In the next section, we examine whether these smaller differences are still statistically significant.

4.3. Planned contrasts

We estimated the full multivariate model (equation 7) using ANOVA coding (e.g., consumption = ½, selection = -½) so that the coefficients of the categorical predictors represent a contrast with the reference category (see Table 4). Effect sizes vary significantly between the three types of interventions. Note that we chose to report two-tailed *p*-values throughout the paper to remain conservative, but that a one-tailed test would also be appropriate given that our hypothesis is about the ordering of cognitively-oriented, affectively-oriented, and behaviorally-oriented interventions. As hypothesized, cognitively-oriented interventions are significantly less effective than affectively-oriented interventions (β = -.12, z = -2.17, p = .03) or behaviorally-oriented interventions (β = -.27, z = -4.73, p < .001). As expected, affectively-oriented interventions are less effective than behaviorally-oriented interventions (β = -.15, z = -2.61, p < .01). Finally, mixed interventions are not more effective than pure cognitively-oriented interventions, whether they include a cognitively-oriented intervention or not (respectively, β = .05, z = .75, p = .45 and β = .13, z = 1.49, p = .14). As expected, effect sizes are similar for food selection and actual consumption (β = .03, z = .73, p = .47). However, effect sizes are significantly lower for total eating compared with healthy eating (β = -.20, z = -2.81, p < .01) or unhealthy eating (β = -.28, z = -3.88, p < .001). As hypothesized as well, effect sizes are significantly higher for unhealthy eating than for healthy eating (β = .08, z = 2.39, p = .02). As expected, and in contrast to what the univariate analyses suggested, effect sizes are significantly lower for grocery stores compared to offsite eateries (β = -.20, z = -2.55, p = .01) or onsite eateries (β = -.14, z = -2.05, p = .04). There are no differences between onsite and offsite eateries

Table 4: Parameter estimates of the multivariate model

	β	***se***	***Z***
Intercept	.23***	.04	5.83
Intervention type			
Cognitively-oriented	(ref)		
Affectively-oriented	.12*	.06	2.17
Behaviorally-oriented	.27***	.06	4.74
Mixed: cognitive present	.05	.08	.75
Mixed: cognitive absent	.13	.09	1.49
Eating behavior type			
Selection	(ref)		
Consumption	.03	.04	.73
Eating behavior measure			
Total eating	-.20**	.08	-2.81
Healthy eating	(ref)		
Mixed eating	-.02	.05	-.41
Unhealthy eating	.08*	.03	2.39
Population setting			
Grocery stores	(ref)		
Offsite eateries	.20*	.08	2.55
Onsite cafeterias	.14*	.07	2.05
Population age			
Children	(ref)		
Adults	.02	.05	.44
Population country			
Other countries	(ref)		
US	.10*	.05	2.03
Study duration			
Intervention length (week)	-.002	.001	-1.22
Study design			
Single-difference pre-post	(ref)		
Single-difference treatment-control	.17*	.05	3.25
Double-difference	.13*	.06	2.26
K (observations)	299		
N (studies)	96		
R^2	.49		
LR test vs. intercept-only model	78***		

***$p < .001$, **$p < .01$, *$p < .05$.
Note: Each coefficient is interpreted as the difference with the reference category, denoted as "(ref)."

(β = -.06, z = -1.08, p = .28). As expected, and contrary to the univariate results, effect sizes are significantly higher in the US than in other countries (β = .10, z = 2.03, p = .04). Contrary to our hypothesis, there is no difference between children and adults (β =.02, z = .45, p = .66). Effect sizes are unrelated to study duration (β = -.002, z = -1.22, p = .22), contrary to the univariate results. Last, effect sizes are significantly lower in pre-post studies than in treatment-control studies (β = -.17, z = -3.25, p < .01), and than in double-difference studies (β = -.13, z = -2.26, p = .02). There is no difference between studies using a treatment-control and a double-difference design (β = -.03, z = -.46, p = .65).

5. Discussion

It is easy to understand the growing enthusiasm for healthy eating nudges in academic and policy circles. They promise to improve people's diet at a fraction of the cost of economic incentives or education programs, without imposing new taxes or constraints on businesses or consumers. But do they really deliver on this promise? Existing reviews and meta-analyses only examined a subset of interventions, and often included studies conducted in laboratory or online settings. More importantly, existing meta-analyses relied on univariate comparisons between two or three groups of studies and failed to control for important differences in eating behaviors, population and studies, or for the fact that some studies yielded multiple effect sizes.

5.1. Do healthy nudges work, and to what extent?

Our analysis of 299 effect sizes derived from 90 articles and 96 field experiments shows that the average effect size of healthy eating nudges is d = .23, 95% CI [.16; .31]). This estimate is considered "small" (Cohen 1988) and is lower than what would have been obtained without controlling for the characteristics of the eating behaviors, population, and studies. To get a more intuitive grasp of what this means, we computed the daily energy equivalent that one would expect from such an effect size using the method described in Hollands et al. (2015). Since d is the standardized mean difference, a d of .23 means that, on average, healthy eating nudges increase healthy eating by .23 standard deviations. Assuming that the standard deviation in daily energy intake is 537 kcal[1] for an adult (Hollands et al. 2015), the average effect size of .23 translates into a .23*537 = 124 kcal change in daily energy intake (-7.2% of the 1,727 kcal average energy intake). Given that a teaspoon of sugar contains 16 kcal, this is equivalent to about 8 fewer teaspoons of sugar per day (see Table 5).

Table 5: Expected daily energy equivalents by intervention type

	Effect sizes	**Daily equivalents** [a]		
	Cohen's *d* (SMD)	Energy intake change (kcal)	Energy intake change (%)	Teaspoons sugar [b]
Cognitively-oriented interventions	.12	-64	-3.7%	-4.0
Affectively-oriented interventions	.24	-129	-7.5%	-8.1
Behaviorally-oriented interventions	.39	-209	-12.1%	-13.1
Overall meta-analytical effect	.23	-124	-7.2%	-7.7

Notes: [a] The daily equivalents are computed using the mean and standard deviation in daily energy intake of 1,727 ± 537 kcal reported in Hollands et al. (2015). [b] One teaspoon of sugar contains 16 kcal.

5.2. Which type of healthy eating nudge works best?

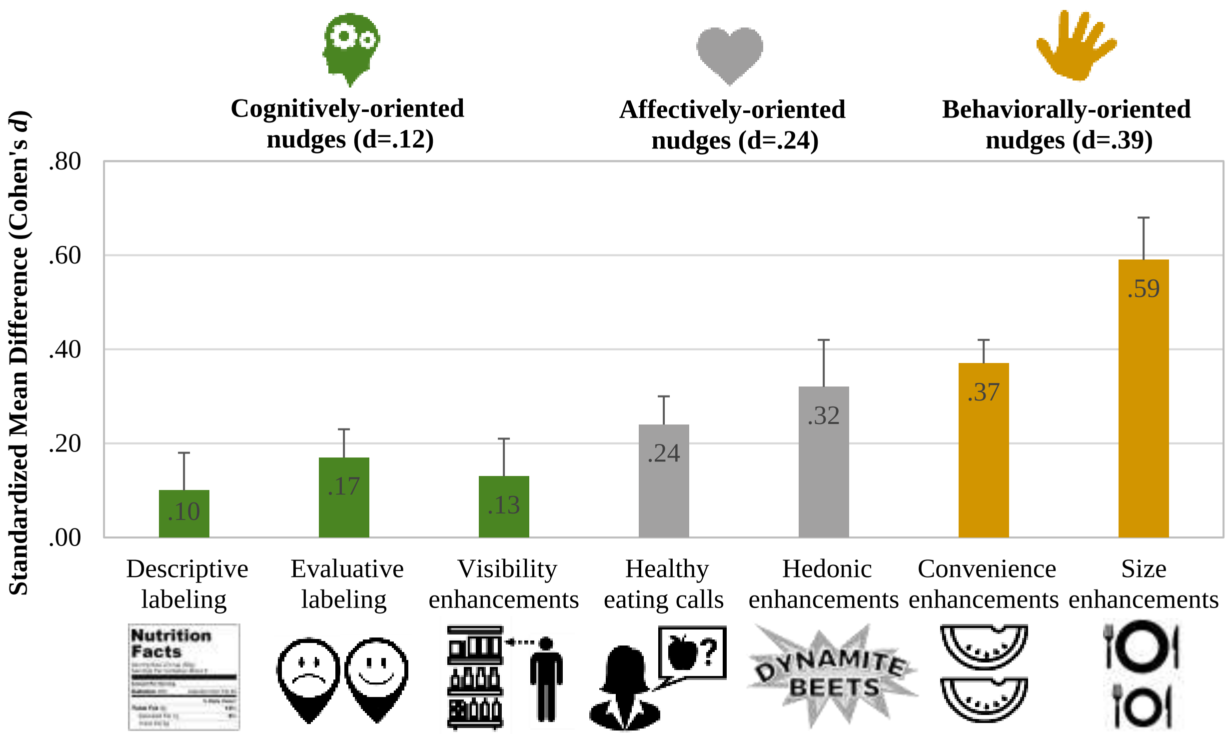
Table 5 provides the daily equivalents of the average effect sizes of cognitively-oriented, affectively-oriented, and behaviorally-oriented interventions. It shows that effect sizes increase by 100% between cognitively-oriented and affectively-oriented interventions (reducing daily energy intake from 64 kcal to 129 kcal). Even more remarkable, moving from a cognitively-oriented to a behaviorally-oriented intervention is estimated to increases effect sizes by a factor of 3.2 (reducing daily energy intake from 64 to 209 kcal per day).

There are also important differences between each type of cognitively-oriented, affectively-oriented, and behaviorally-oriented nudges. As detailed in Figure 3 and Appendix C, we estimated another meta-regression which, instead of estimating five effect sizes (for the three pure types and the two mixed types), estimated a separate effect size for each of the seven subcategories, for the two mixed types of intervention, and for a tenth subcategory consisting of studies combining multiple cognitively-oriented interventions (e.g., evaluative nutrition labeling

[1] We use the official measure "kcal" for Calories, such that 1 kcal = 1 Calorie

and visibility enhancements). There were no studies combining the two types of affectively-oriented interventions or the two types of behaviorally-oriented interventions.

Figure 3: Effect sizes by nudge type



Note: Error bar represent standard error.

Multivariate estimates for each of the ten interventions types are compared to the univariate results in Appendix C. We find that smaller effect sizes are slightly higher in the multivariate analyses while larger effect sizes are lower in the multivariate analyses. For example, the univariate average effect size for size enhancements shrinks by 17% (from *d* = .71 to *d* = .59). Note that this univariate estimate (*d* = .71) is very similar to the value (*d* = .76) reported by Holden et al. (2016) for studies with unaware participants (e.g., excluding laboratory or online studies). In addition to overestimating effect sizes, the univariate analysis incorrectly ranks some of the interventions, suggesting, for example, that visibility enhancements are more effective than evaluative labeling, when they are not.

5.3. Which other factors influence the effectiveness of healthy eating nudges?

By explaining 49% of the variance among effect sizes, our study shows that some of the characteristics of the eating behavior, population, and study significantly impact the effectiveness of healthy eating nudges. In Table 6, we summarize the results of the hypothesis tests and show when they replicate, fail to replicate, or extend the results of existing meta-analyses. First, we find that interventions more easily reduce unhealthy eating than improve healthy eating or decrease total eating. In other words, it is easier to make people eat less chocolate cake than to make them eat more vegetables, and the most difficult is to make them simply eat less. In fact, the estimated effect size for total eating ($d = .07$, $z = .98$, $p = .32$, see Figure 2) is not statistically different from zero. This finding is consistent with what we know about the difficulty—perhaps even pointlessness—of hypocaloric diets.

Table 6: Summary of hypothesis test results

Hypothesis (effect sizes)	Validated	Results replicate previous meta-analyses?
Intervention type Cognitive < Affective < Behavioral	Yes	New result
Outcome behavior Selection < Consumption	No	Replication: Holden et al. (2016); Hollands et al. (2015); Littlewood et al. (2016); Sinclair et al. (2014)
Outcome measure Healthy eating < Unhealthy eating	Yes	Replication: Hollands et al. (2015); Zlatevska et al. (2014)
Outcome measure Total eating < other eating measures	Yes	Replication: Cecchini and Warin (2016)
Population: age Children < Adults	No	Does not replicate: Hollands et al. (2015); Zlatevska et al. (2014)
Population setting Grocery stores < Cafeterias or restaurants	Yes	New result
Population: country Other countries < USA	Yes	New result

Our finding of a 30% stronger effect size for reducing unhealthy eating than for increasing healthy eating is consistent with prior research on self-control (Prelec and Loewenstein 1998; Wertenbroch 1998). Dynamically inconsistent preferences and self-control lapses can explain why people would particularly welcome interventions that reduce unhealthy eating and help them stick to their long-term goals and avoid regret (Schwartz et al. 2014).

On the other hand, we replicate prior findings of similar effect sizes for food selection rather than actual consumption. This is an important result because it suggests that researchers or practitioners may not need to measure actual consumption to test the impact of their interventions, which is usually considerably more onerous to measure than just the number of consumers picking healthier options.

We find that effect sizes are unaffected by the duration of the study. As Figure 2 shows, our model predicts that increasing the duration of the study from 1 week to 15 weeks would reduce effect size by only 12% (from d = .26 to d = .23). As noted earlier, study duration captures the length of the intervention, but not necessarily the difference between short- and long-term effects since some settings (e.g., restaurants) may have mostly first-time customers even when the intervention is tested over a relatively long period. To examine this issue, we explored whether duration interacted with study location (restaurant vs. cafeteria vs. grocery stores) and found that it did not (See Appendix D). As noted earlier as well, the sample of studies did not allow us to measure potential carry-over effects once the intervention was stopped.

We also find that effect sizes increase with the level of control in the design of the study. On average, effect sizes are 131% larger in studies with a control group compared with those with a simple pre-post design without a control group. However, treatment-control groups could be subject to a selection bias. This suggests that researchers should use stronger controls as much as possible. It also provides a way to correct the effect sizes found in pre-post studies and to forecast what they might have been in a more controlled setting.

Turning to population characteristics, effect sizes are 146% (or 100%) smaller on average among grocery shoppers than among restaurant (or cafeteria) eaters. This is consistent with our hypothesis and with the literature, although more research is needed to determine if it is because of the differences between choosing for immediate or future consumption, because of different levels of competition, or because different goals are salient when grocery shopping vs. eating.

Also consistent with our hypothesis, effect sizes are 47% larger in studies conducted in the US than in other countries. This may be because Americans focus less on the experience and more on the health effects of eating (Rozin et al. 1999) or because they rely more strongly on external eating cues than internal ones (Wansink et al. 2007). It could also be caused by the higher proportion of overweight people in the US and the larger size of portions (Rozin et al. 2003).

On the other hand, the difference in effect sizes between adults and children is not statistically significant. Still, compared to the univariate analyses which suggest larger effects for children than for adults, our results are in the direction (smaller effects for children) that we hypothesized, consistent with the literature. To determine conclusively whether children and adults respond differently to healthy eating nudges, more research is needed, especially on cognitively-oriented interventions which have, so far, been tested primarily with adults.

Our analysis allows us to predict the effect size to expect when conducting a field experiment with any combination of predictors, including the most typical and the most effective combination. Table 7 summarizes the typical and best scenarios, as well as the contribution of the different predictors. It shows that researchers choosing the most typical level of each predictor (studying the effects of a cognitively-oriented intervention on the healthy food selection of US adult cafeteria eaters for a pre-post 15-week study) could expect an effect size of only $d = .12$, 95% CI [.03, .21]). In contrast, researchers choosing the best combination of predictors (studying the effects of a behavioral intervention on the unhealthy food consumption of adult restaurant eaters for a single-difference 1-week study) could expect an effect size four and a half times larger ($d = .74$, 95% CI [.60, .88]). Computing the daily energy equivalents, we get a reduction by 64 kcal for the typical nudge study, and a reduction by 397 kcal for the best one.

Table 7: Expected effectiveness increase between typical and best nudge study

Predictor	Typical scenario	Best scenario	Increase (d)	Increase (contribution)
Intervention type	Cognitively-oriented	Behaviorally-oriented	.27	43%
Eating behavior type	Selection	Consumption	.03	5%
Eating behavior measure	Healthy	Unhealthy	.08	13%
Study Duration	15 weeks	1 week	.02	4%
Study Design	Pre-post	Single-difference	.17	27%
Population: Country	US	US		
Population: Location	Onsite cafeterias	Onsite cafeterias		
Population: Age	Adults	Adults		
Effect Size (d)	.12	.74	.62	100%

5.5. Limitations and directions for future research

We categorized visibility enhancements as cognitively-oriented nudges because they seek to draw people's attention to healthier options. Because consumers rarely look at all the food options available, enhancing the visibility of healthy options or reducing it for unhealthy options changes people's knowledge of the healthiness of the options that are available to them. Some visibility enhancements are purely cognitively-oriented. For example, placing healthier foods in a visible place on the menu or near the cash register rather than earlier in the cafeteria line makes them easier to see but not easier to order or grab. Similarly, leaving leftover chicken wings on the table of all-you-can-eat restaurants draws attention to the amount eaten but does not make it less convenient to eat more. Other visibility enhancements however, such as placing healthier foods at eye level on a supermarket shelf, make these foods easier to see but also easier to reach, and have therefore a behavioral component. To examine this issue, we estimated a model in which the 25 visibility observations were categorized as behaviorally-oriented rather than cognitively-oriented. This reduces the effect size of behaviorally-oriented interventions (d = .32, z = 6.63, p < .001) while slightly increasing the estimate for cognitively-oriented interventions (d = .14, z = 2.82, p < .01). However, the difference in effect sizes between the two interventions remains statistically significant (Δ = .18, z = 3.13, p < .01). Moreover, this alternative categorization fits the data less well (the R^2 diminishes from .49 to .44, $\chi^2(1) = 11.42$, p < .001).

Our findings offer insights into where more research is needed and where it is not. Table 3 shows the number of observations by intervention type and target eating behavior (healthy or unhealthy). From this it is immediately apparent that no field experiment has tested the effectiveness of displaying unattractive product descriptions or photos of unhealthy foods, focusing on negative hedonic aspects. Similarly, and with the exception of (Donnelly et al. 2018), little attention has been given to the use of dissuasive photos and warnings used on cigarette packs in some countries (Kees et al. 2006). Although degrading other brands may be difficult because of trademark laws, it has shown promise in laboratory studies (Hollands et al. 2011) – retailers or restaurants could test this strategy with their own unhealthy products.

It would also seem important to run more studies increasing portion, plate, or glass size for healthy foods and beverages, rather than for unhealthy ones. Further studies are needed to examine the effectiveness of hedonic enhancements, for which we only have 9 effect sizes. Precedence should be given to testing interventions in grocery stores and outside the US, and for unhealthy foods. These issues should have priority over other well-researched topics, such as studying the effects of cognitively-oriented interventions on healthy eating in cafeterias using a pre-post design.

Beyond filling out the underpopulated cells of the framework, we also encourage author to follow a strategy to increase the precision of knowledge in the field using a procedure that reduces collinearity among design variables (Farley et al. 1998). Similarly, three research areas appear particularly fruitful. The first is to study interaction effects. Lack of data in our sample makes it impossible to estimate interactions effects between each intervention type and the other predictors. In Appendix D we report the results of a simplified model including interactions but using linear coding for intervention type and eating behavior (from total eating to healthy eating). This preliminary analysis suggests that shifting from cognitively-oriented to behaviorally-oriented interventions is particularly impactful on consumption (vs. selection), for adults (vs. children), and in the US (vs. in other countries). These results qualify the lack of main effect for eating behavior and for participant age reported in the main results. Additional field experiments orthogonally manipulating intervention type and population or study characteristics would be necessary to confirm these results.

Second, it is important to directly compare nudges and economic incentives and see if they can complement each other. To provide initial insights, we compared our results to those of Afshin et al. (2017), who estimated the impact of a 10% price cut on the selection of fruit and vegetables using 22 effect sizes from 15 studies. After collecting the sample sizes from the original 15 studies, we were able to convert these 10% estimates into standardized mean differences. We found a d of .27 (se = .07, z = 3.95, p < .001) for the effects of a 10% price reduction on healthy eating selection, which is equal to the mean effect size that we found for healthy eating (vs. unhealthy, mixed, or total eating), across all nudges (d = .27) without control variables (as done in their meta-analysis). This suggests that nudges are equivalent to a 10% permanent price reduction in this particular context. This preliminary analysis should encourage future research comparing nudge and economic interventions both in terms of their effects on healthy eating and their cost.

Finally, the prevalence and severity of non-communicable diseases is strongly associated with socioeconomic and cultural factors such as income, education, gender, ethnicity, and culture (Bartley 2017). Surprisingly, this data was almost never available in the studies that we analyzed. Future research should therefore measure socioeconomic data, as well as biomarkers such as body mass or diabetes, and traits such as cognitive restraint or impulsivity that strongly influence food choices and health (Ma et al. 2013; Sutin et al. 2011). Such information should be provided to better characterize the respondent population, but it would be even better to report results separately for different population types. This should be done systematically even in the absence of significant differences. When prior research or theory predicts an effect, finding none can be informative.

More broadly, future research should expand the dependent variables beyond purchase and consumption. To encourage the adoption of healthy eating nudges in commercial operations, it is important to measure their impact on the consumer's experience, satisfaction, and perception of value, as well as on the company's top and bottom lines. Even interventions that lead to a reduction in consumption can be good for business if they attract new consumers who value their ability to nudge them away from unhealthy choices that they will later regret.

Similarly, one of the core tenets of nudges is that they improve consumer welfare as judged by consumers themselves (Sunstein 2018a). It would be important to know whether people, upon learning that they have been subject to an intervention, would agree that it led them to make better decisions compared to the status quo ante, and also to interventions such as taxes and other economic incentives. This is important because, while they preserve freedom of choice, the interventions analyzed here are nevertheless paternalistic. Finding that consumers welcome these interventions, as some studies suggest they do (Loewenstein et al. 2015), and that they are compatible with commercial goals, would go a long way to encourage their adoption.

5.5. Toward a "living" meta-analysis

One of the biggest challenges of studying healthy eating nudges is the exponential increase in the studies carried out. This upsurge makes meta-analyses even more valuable but also means that they rapidly become out of date. Compounding the problem, research on healthy eating nudges is published in a wide variety of scientific publications in marketing, nutrition, psychology, and health sciences, which are indexed in different databases and not always available to all researchers. To mitigate these problems and correct possible categorization errors, the spreadsheet containing the raw data will be made available online (post publication). In addition, we have created a simple survey (available at http://tinyurl.com/healthy-eating-nudge) to allow researchers to correct and update the database by entering information about their study. We hope that this "living" meta-analysis will encourage the consolidation and diffusion of knowledge and contribute to making the science more open.

6. References*

* Adams, M.A., R.L. Pelletier, M.M. Zive, J.F. Sallis. 2005. Salad bars and fruit and vegetable consumption in elementary schools: A plate waste study. *J Am Diet Assoc*. **105**(11) 1789-1792.

Afshin, A., J.L. Peñalvo, L. Del Gobbo, J. Silva, M. Michaelson, M. O'Flaherty, S. Capewell, D. Spiegelman, G. Danaei, D. Mozaffarian. 2017. The prospective impact of food pricing on improving dietary consumption: A systematic review and meta-analysis. *PLoS One*. **12**(3) e0172277.

* Anzman-Frasca, S., A.C. Braun, S. Ehrenberg, L.H. Epstein, A. Gampp, L.A. Leone, A. Singh, S. Tauriello. 2018. Effects of a randomized intervention promoting healthy children's meals on children's ordering and dietary intake in a quick-service restaurant. *Physiol Behav*. **192** 109-117.

Arno, A., S. Thomas. 2016. The efficacy of nudge theory strategies in influencing adult dietary behaviour: A systematic review and meta-analysis. *BMC Public Health*. **16**(1) 676.

* Auchincloss, A.H., G.G. Mallya, B.L. Leonberg, A. Ricchezza, K. Glanz, D.F. Schwarz. 2013. Customer responses to mandatory menu labeling at full-service restaurants. *Am J Prev Med*. **45**(6) 710-719.

Barry, T.E., D.J. Howard. 1990. A review and critique of the hierarchy of effects in advertising. *Int J Advert*. **9**(2) 121-135.

* Bartholomew, J.B., E.M. Jowers. 2006. Increasing frequency of lower-fat entrees offered at school lunch: An environmental change strategy to increase healthful selections. *J Acad Nutr Diet*. **106**(2) 248-252.

Bartley, M. 2017. *Health inequality : An introduction to concepts, theories and methods*. Polity, Cambridge, UK ; Malden, MA, USA.

* Baskin, E., M. Gorlin, Z. Chance, N. Novemsky, R. Dhar, K. Huskey, M. Hatzis. 2016. Proximity of snacks to beverages increases food consumption in the workplace: A field study. *Appetite*. **103** 244-248.

BIT. 2014. *East: Four simple ways to apply behavioural insights*. Behavioural Insight Team, London.

Bloom, D., E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, L. Bloom, S. Fathima, A. Feigl, T. Gaziano, A. Hamandi, M. Mowafi. 2012. *The global economic burden of noncommunicable diseases*. Program on the Global Demography of Aging.

* Bollinger, B., P. Leslie, A. Sorensen. 2011. Calorie posting in chain restaurants. *Am Econ J Econ Policy*. **3**(1) 91-128.

Booth, A. 2006. Clear and present questions: Formulating questions for evidence based practice. *Library Hi Tech*. **24**(3) 355-368.

Borenstein, M., L.V. Hedges, J.P.T. Higgins, H.R. Rothstein. 2009. *Introduction to meta-analysis*. John Wiley & Sons, Chichester, U.K.

Brandon, A., P.J. Ferraro, J.A. List, R.D. Metcalfe, M.K. Price, F. Rundhammer. 2017. Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments. *NBER Working Paper Series*. **No. 23277**.

Breckler, S.J. 1984. Empirical validation of affect, behavior, and cognition as distinct components of attitude. *J Pers Soc Psychol*. **47**(6) 1191-1205.

* Article included in the meta-analysis.

Brewer, N.T., G.B. Chapman, A.J. Rothman, J. Leask, A. Kempe. 2017. Increasing vaccination: Putting psychological science into action. *Psychol Sci Public Interest*. **18**(3) 149-207.
* Brissette, I., A. Lowenfels, C. Noble, D. Spicer. 2013. Predictors of total calories purchased at fast-food restaurants: Restaurant characteristics, calorie awareness, and use of calorie information. *J Nutr Educ Behav*. **45**(5) 404-411.
Broers, V.J., C. De Breucker, S. Van den Broucke, O. Luminet. 2017. A systematic review and meta-analysis of the effectiveness of nudging to increase fruit and vegetable choice. *Eur J Public Health*. **27**(5) 912-920.
Bucher, T., C. Collins, M.E. Rollo, T.A. McCaffrey, N. De Vlieger, D. Van der Bend, H. Truby, F.J. Perez-Cueto. 2016. Nudging consumers towards healthier choices: A systematic review of positional influences on food choice. *Br J Nutr*. **115**(12) 2252-2263.
* Buscher, L.A., K.A. Martin, S. Crocker. 2001. Point-of-purchase messages framed in terms of cost, convenience, taste, and energy improve healthful snack selection in a college foodservice setting. *J Acad Nutr Diet*. **101**(8) 909-913.
Carpenter, C.J. 2010. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. *Health Communication*. **25**(8) 661-669.
* Cawley, J., M.J. Sweeney, J. Sobal, D.R. Just, H.M. Kaiser, W.D. Schulze, E. Wethington, B. Wansink. 2015. The impact of a supermarket nutrition rating system on purchases of nutritious and less nutritious foods. *Public Health Nutr*. **18**(01) 8-14.
Cecchini, M., L. Warin. 2016. Impact of food labelling systems on food choices and eating behaviours: A systematic review and meta-analysis of randomized studies. *Obesity Reviews*. **17**(3) 201-210.
Chance, Z., M. Gorlin, R. Dhar. 2014. Why choosing healthy foods is hard, and how to help: Presenting the 4ps framework for behavior change. *Customer Needs and Solutions*. **1**(4) 253-262.
Chandon, P., J.W. Hutchinson, E.T. Bradlow, S.H. Young. 2009. Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *J Marketing*. **73**(6) 1-17.
Chernev, A. 2011. The dieter's paradox. *J Consum Psychol*. **21**(2) 178-183.
Cheung, M.W.L. 2014. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol Methods*. **19**(2) 211-229.
* Chu, Y.H., E.A. Frongillo, S.J. Jones, G.L. Kaye. 2009. Improving patrons' meal selections through the use of point-of-selection nutrition labels. *Am J Public Health*. **99**(11) 2001-2005.
Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.
* Cohen, J.F., S.A. Richardson, S.A. Cluggish, E. Parker, P.J. Catalano, E.B. Rimm. 2015. Effects of choice architecture and chef-enhanced meals on the selection and consumption of healthier school foods: A randomized clinical trial. *JAMA Pediatr*. **169**(5) 431-437.
* Crockett, R.A., S.A. Jebb, M. Hankins, T.M. Marteau. 2014. The impact of nutritional labels and socioeconomic status on energy intake. An experimental field study. *Appetite*. **81** 12-19.
Croll, J.K., D. Neumark-Sztainer, M. Story. 2001. Healthy eating: What does it mean to adolescents? *J Nutr Educ*. **33**(4) 193-198.
* Dayan, E., M. Bar-Hillel. 2011. Nudge to nobesity ii: Menu positions influence food orders. *Judgm Decis Mak*. **6**(4) 333.
* De Bondt, C., A. Van Kerckhove, M. Geuens. 2017. 'My lips are sealed' - the impact of package resealability on the consumption of tempting foods. *Appetite*. **117** 143-151.
* de Wijk, R.A., A.J. Maaskant, I.A. Polet, N.T. Holthuysen, E. van Kleef, M.H. Vingerhoeds. 2016. An in-store experiment on the effect of accessibility on sales of wholegrain and white bread in supermarkets. *PLoS One*. **11**(3) e0151915.
* Diliberti, N., P.L. Bordi, M.T. Conklin, L.S. Roe, B.J. Rolls. 2004. Increased portion size leads to increased energy intake in a restaurant meal. *Obesity*. **12**(3) 562-568.
* DiSantis, K.I., L.L. Birch, A. Davey, E.L. Serrano, J. Zhang, Y. Bruton, J.O. Fisher. 2013. Plate size and children's appetite: Effects of larger dishware on self-served portions and intake. *Pediatrics*. **131**(5) e1451-e1458.
Dolan, P., M. Hallsworth, D. Halpern, D. King, R. Metcalfe, I. Vlaev. 2012. Influencing behaviour: The mindspace way. *J Econ Psychol*. **33**(1) 264-277.
* Donnelly, G.E., L.Y. Zatz, D. Svirsky, L.K. John. 2018. The effect of graphic warnings on sugary-drink purchasing. *Psychol Sci*.
* Downs, J.S., J. Wisdom, B. Wansink, G. Loewenstein. 2013. Supplementing menu labeling with calorie recommendations to test for facilitation effects. *Am J Public Health*. **103**(9) 1604-1609.
* Dubbert, P.M., W.G. Johnson, D.G. Schlundt, N.W. Montague. 1984. The influence of caloric information on cafeteria food choices. *J Appl Behav Anal*. **17**(1) 85-92.

Duckworth, A.L., T.S. Gendler, J.J. Gross. 2016. Situational strategies for self-control. *Perspectives on Psychological Science*. **11**(1) 35-55.
* Dumanovsky, T., C.Y. Huang, C.A. Nonas, T.D. Matte, M.T. Bassett, L.D. Silver. 2011. Changes in energy content of lunchtime purchases from fast food restaurants after introduction of calorie labelling: Cross sectional customer surveys. *BMJ*. **343**.
Duval, S., R. Tweedie. 2000. Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. **56**(2) 455-463.
* Elbel, B., J. Gyamfi, R. Kersh. 2011. Child and adolescent fast-food choice and the influence of calorie labeling: A natural experiment. *Int J Obes*. **35**(4) 493-500.
* Elbel, B., R. Kersh, V.L. Brescoll, L.B. Dixon. 2009. Calorie labeling and food choices: A first look at the effects on low-income people in new york city. *Health Aff*. **28**(6) w1110-1121.
* Elbel, B., T. Mijanovich, L.B. Dixon, C. Abrams, B. Weitzman, R. Kersh, A.H. Auchincloss, G. Ogedegbe. 2013. Calorie labeling, fast food purchasing and restaurant visits. *Obesity*. **21**(11) 2172-2179.
* Ellison, B., J.L. Lusk, D. Davis. 2013. Looking at the label and beyond: The effects of calorie labels, health consciousness, and demographics on caloric intake in restaurants. *Int J Behav Nutr Phys Act*. **10**(1) 21.
* Elsbernd, S., M. Reicks, T. Mann, J. Redden, E. Mykerezi, Z. Vickers. 2016. Serving vegetables first: A strategy to increase vegetable consumption in elementary school cafeterias. *Appetite*. **96** 111-115.
* Ensaff, H., M. Homer, P. Sahota, D. Braybrook, S. Coan, H. McLeod. 2015. Food choice architecture: An intervention in a secondary school and its impact on students' plant-based food choices. *Nutrients*. **7**(6) 4426-4437.
Farley, J.U., D.R. Lehmann, L.H. Mann. 1998. Designing the next study for maximum impact. *J Marketing Res*. **35**(4) 496-501.
Fernandes, A.C., R.C. Oliveira, R.P.C. Proença, C.C. Curioni, V.M. Rodrigues, G.M.R. Fiates. 2016. Influence of menu labeling on food choices in real-life settings: A systematic review. *Nutr Rev*. **74**(8) 534-548.
* Finkelstein, E.A., K.L. Strombotne, N.L. Chan, J. Krieger. 2011. Mandatory menu labeling in one fast-food chain in king county, washington. *Am J Prev Med*. **40**(2) 122-127.
* Foster, G.D., A. Karpyn, A.C. Wojtanowski, E. Davis, S. Weiss, C. Brensinger, A. Tierney, W. Guo, J. Brown, C. Spross, D. Leuchten, P.J. Burns, K. Glanz. 2014. Placement and promotion strategies to increase sales of healthier products in supermarkets in low-income, ethnically diverse neighborhoods: A randomized controlled trial. *Am J Clin Nutr*. **99**(6) 1359-1368.
* Freedman, M.R. 2011. Point-of-selection nutrition information influences choice of portion size in an all-you-can-eat university dining hall. *Journal of Foodservice Business Research*. **14**(1) 86-98.
* Freedman, M.R., C. Brochado. 2010. Reducing portion size reduces food intake and plate waste. *Obesity*. **18**(9) 1864-1866.
* Friis, R., L.R. Skov, A. Olsen, K.M. Appleton, L. Saulais, C. Dinnella, H. Hartwell, L. Depezay, E. Monteleone, A. Giboreau, F.J.A. Perez-Cueto. 2017. Comparison of three nudge interventions (priming, default option, and perceived variety) to promote vegetable consumption in a self-service buffet setting. *PLoS One*. **12**(5) e0176028.
* Gaigi, H., S. Raffin, M. Maillot, L. Adrover, B. Ruffieux, N. Darmon. 2015. Expérimentation d'un fléchage nutritionnel dans deux supermarchés à marseille «le choix vita+». *Cahiers de Nutrition et de Diététique*. **50**(1) 16-24.
* Gamburzew, A., N. Darcel, R. Gazan, C. Dubois, M. Maillot, D. Tomé, S. Raffin, N. Darmon. 2016. In-store marketing of inexpensive foods with good nutritional quality in disadvantaged neighborhoods: Increased awareness, understanding, and purchasing. *Int J Behav Nutr Phys Act*. **13**(1) 104.
* Geaney, F., C. Kelly, J.S. Di Marrazzo, J.M. Harrington, A.P. Fitzgerald, B.A. Greiner, I.J. Perry. 2016. The effect of complex workplace dietary interventions on employees' dietary intakes, nutrition knowledge and health status: A cluster controlled trial. *Prev Med*. **89** 76-83.
Glanz, K., M. Basil, E. Maibach, J. Goldberg, D.A.N. Snyder. 1998. Why americans eat what they do: Taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *J Am Diet Assoc*. **98**(10) 1118-1126.
* Goto, K., A. Waite, C. Wolff, K. Chan, M. Giovanni. 2013. Do environmental interventions impact elementary school students' lunchtime milk selection? *Appl Econ Perspect Policy*. **35**(2) 360-376.
* Greene, K.N., G. Gabrielyan, D.R. Just, B. Wansink. 2017. Fruit-promoting smarter lunchrooms interventions: Results from a cluster rct. *Am J Prev Med*. **52**(4) 451-458.
Hagen, L., A. Krishna, B. McFerran. 2016. Rejecting responsibility: Low physical involvement in obtaining food promotes unhealthy eating. *J Marketing Res*. **54**(4) 589-604.
* Hanks, A.S., D.R. Just, L.E. Smith, B. Wansink. 2012. Healthy convenience: Nudging students toward healthier choices in the lunchroom. *J Public Health*. **34**(3) 370-376.

* Hanks, A.S., D.R. Just, B. Wansink. 2013. Smarter lunchrooms can address new school lunchroom guidelines and childhood obesity. *J Pediatr*. **162**(4) 867-869.
Hanssens, D.M., K.H. Pauwels, S. Srinivasan, M. Vanhuele, G. Yildirim. 2014. Consumer attitude metrics for guiding marketing mix decisions. *Marketing Sci*. **33**(4) 534-550.
Herman, C.P., J. Polivy. 2008. External cues in the control of food intake in humans: The sensory-normative distinction. *Physiol Behav*. **94**(5) 722-728.
Herman, C.P., D.A. Roth, J. Polivy. 2003. Effects of the presence of others on food intake: A normative interpretation. *Psychol Bull*. **129**(6) 873-886.
Higgins, J., S.G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med*. **21**(11) 1539-1558.
Hilgard, E.R. 1980. The trilogy of mind: Cognition, affection, and conation. *J Hist Behav Sci*. **16**(2) 107-117.
* Hoefkens, C., C. Lachat, P. Kolsteren, J. Van Camp, W. Verbeke. 2011. Posting point-of-purchase nutrition information in university canteens does not influence meal choice and nutrient intake. *Am J Clin Nutr*. **94**(2) 562-570.
Holden, S.S., N. Zlatevska, C. Dubelaar. 2016. Whether smaller plates reduce consumption depends on who's serving and who's looking: A meta-analysis. *Journal of the Association for Consumer Research*. **1**(1) 134-146.
Hollands, G., G. Bignardi, M. Johnston, M. Kelly, D. Ogilvie, M. Petticrew, A. Prestwich, I. Shemilt, S. Sutton, T. Marteau. 2017. The tippme intervention typology for changing environments to change behaviour. *Nature Human Behaviour*.
Hollands, G.J., A. Prestwich, T.M. Marteau. 2011. Using aversive images to enhance healthy food choices and implicit attitudes: An experimental test of evaluative conditioning. *Health Psychol*. **30**(2) 195.
Hollands, G.J., I. Shemilt, T.M. Marteau, S.A. Jebb, M.P. Kelly, R. Nakamura, M. Suhrcke, D. Ogilvie. 2013. Altering micro-environments to change population health behaviour: Towards an evidence base for choice architecture interventions. *BMC Public Health*. **13**(1) 1218.
Hollands, G.J., I. Shemilt, T.M. Marteau, S.A. Jebb, H.B. Lewis, Y. Wei, J.P. Higgins, D. Ogilvie. 2015. Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. *Cochrane Database Syst Rev*. **9** CD011045.
* Hubbard, K.L., L.G. Bandini, S.C. Folta, B. Wansink, M. Eliasziw, A. Must. 2015. Impact of a smarter lunchroom intervention on food selection and consumption among adolescents and young adults with intellectual and developmental disabilities in a residential school setting. *Public Health Nutr*. **18**(2) 361-371.
Jackson, D., R. Riley, I.R. White. 2011. Multivariate meta-analysis: Potential and promise. *Stat Med*. **30**(20) 2481-2498.
Januszewska, R., Z. Pieniak, W. Verbeke. 2011. Food choice questionnaire revisited in four countries. Does it still measure the same? *Appetite*. **57**(1) 94-98.
Kees, J., S. Burton, J.C. Andrews, J. Kozup. 2006. Tests of graphic visuals and cigarette package warning combinations: Implications for the framework convention on tobacco control. *J Public Policy Mark*. **25**(2) 212-223.
* Kiesel, K., S.B. Villas-Boas. 2013. Can information costs affect consumer choice? Nutritional labels in a supermarket experiment. *Int J Ind Organ*. **31**(2) 153-163.
Kraak, V.I., T. Englund, S. Misyak, E.L. Serrano. 2017. A novel marketing mix and choice architecture framework to nudge restaurant customers toward healthy food environments to reduce obesity in the united states. *Obesity Reviews*. **18**(8) 852-868.
* Krieger, J.W., N.L. Chan, B.E. Saelens, M.L. Ta, D. Solet, D.W. Fleming. 2013. Menu labeling regulations and calories purchased at chain restaurants. *Am J Prev Med*. **44**(6) 595-604.
Krishna, A. 2012. An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *J Consum Psychol*. **22**(3) 332-351.
* Kroese, F.M., D.R. Marchiori, D.T. de Ridder. 2016. Nudging healthy food choices: A field experiment at the train station. *J Public Health*. **38**(2) e133-e137.
* Lachat, C.K., R. Verstraeten, B. De Meulenaer, J. Menten, L.F. Huybregts, J. Van Camp, D. Roberfroid, P.W. Kolsteren. 2009. Availability of free fruits and vegetables at canteen lunch improves lunch and daily nutritional profiles: A randomised controlled trial. *Br J Nutr*. **102**(07) 1030-1037.
* Levin, S. 1996. Pilot study of a cafeteria program relying primarily on symbols to promote healthy choices. *J Nutr Educ*. **28**(5) 282-285.
* Levy, D.E., J. Riis, L.M. Sonnenberg, S.J. Barraclough, A.N. Thorndike. 2012. Food choices of minority and low-income employees: A cafeteria intervention. *Am J Prev Med*. **43**(3) 240-248.
Littlewood, J.A., S. Lourenço, C.L. Iversen, G.L. Hansen. 2016. Menu labelling is effective in reducing energy ordered and consumed: A systematic review and meta-analysis of recent studies. *Public Health Nutr*. **19**(12) 2106-2121.

Loewenstein, G., C. Bryce, D. Hagmann, S. Rajpal. 2015. Warning: You are about to be nudged. *Behavioral Science & Policy*. **1**(1) 35-42.
Long, M.W., D.K. Tobias, A.L. Cradock, H. Batchelder, S.L. Gortmaker. 2015. Systematic review and meta-analysis of the impact of restaurant menu calorie labeling. *Am J Public Health*. **105**(5) e11-e24.
Ly, K., N. Mazar, M. Zhao, D. Soman. 2013. *A practitioner's guide to nudging*. Research Report Series, Behavioural Economics in Action, University of Toronto.
Ma, Y., K.L. Ailawadi, D. Grewal. 2013. Soda versus cereal and sugar versus fat: Drivers of healthful food intake and the impact of diabetes diagnosis. *J Marketing*. **77**(3) 101-120.
Macht, M. 2008. How emotions affect eating: A five-way model. *Appetite*. **50**(1) 1-11.
* Mayer, J.A., J.M. Heins, J.M. Vogel, D.C. Morrison, L.D. Lankester, A.L. Jacobs. 1986. Promoting low-fat entree choices in a public cafeteria. *J Appl Behav Anal*. **19**(4) 397-402.
* Mazza, M.C., L. Dynan, R.M. Siegel, A.L. Tucker. 2017. Nudging healthier choices in a hospital cafeteria: Results from a field study. *Health promotion practice* 1-10.
McFerran, B., D.W. Dahl, G.J. Fitzsimons, A.C. Morales. 2010. Might an overweight waitress make you eat more? How the body type of others is sufficient to alter our food consumption. *J Consum Psychol*. **20**(2) 146-151.
* Meyers, A.W., A.J. Stunkard. 1980. Food accessibility and food choice: A test of schachter's externality hypothesis. *Arch Gen Psychiatry*. **37**(10) 1133-1135.
* Miller, G.F., S. Gupta, J.D. Kropp, K.A. Grogan, A. Mathews. 2016. The effects of pre-ordering and behavioral nudges on national school lunch program participants' food item selection. *J Econ Psychol*. **55** 4-16.
* Mishra, A., H. Mishra, T.M. Masters. 2012. The influence of bite size on quantity of food consumed: A field study. *Journal of Consumer Research*. **38**(5) 791-795.
Moher, D., A. Liberati, J. Tetzlaff, D.G. Altman, P. Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLoS Med*. **6**(7) e1000097.
* Mollen, S., R.N. Rimal, R.A. Ruiter, G. Kok. 2013. Healthy and unhealthy social norms and food selection. Findings from a field-experiment. *Appetite*. **65** 83-89.
* Morizet, D., L. Depezay, P. Combris, D. Picard, A. Giboreau. 2012. Effect of labeling on new vegetable dish acceptance in preadolescent children. *Appetite*. **59**(2) 399-402.
Murimi, M.W., M. Kanyi, T. Mupfudze, M.R. Amin, T. Mbogori, K. Aldubayan. 2017. Factors influencing efficacy of nutrition education interventions: A systematic review. *J Nutr Educ Behav*. **49**(2) 142-165.e141.
Nikolaou, C.K., C.R. Hankey, M.E.J. Lean. 2014. Calorie-labelling: Does it impact on calorie purchase in catering outlets and the views of young adults? *Int J Obes*. **39** 542.
* Nikolova, H.D., J.I. Inman. 2015. Healthy choice: The effect of simplified pos nutritional information on consumer food choice behavior. *J Marketing Res*. **7**(December) 817–835.
Nornberg, T.R., L. Houlby, L.R. Skov, F.J. Perez-Cueto. 2016. Choice architecture interventions for increased vegetable intake and behaviour change in a school setting: A systematic review. *Perspectives in public health*. **136**(3) 132-142.
* Ogawa, Y., N. Tanabe, A. Honda, T. Azuma, N. Seki, T. Suzuki, H. Suzuki. 2011. Point-of-purchase health information encourages customers to purchase vegetables: Objective analysis by using a point-of-sales system. *Environ Health Prev Med*. **16**(4) 239-246.
Oliver, R.L. 1999. Whence consumer loyalty? *J Marketing*. **63** 33-44.
* Olstad, D.L., L.A. Goonewardene, L.J. McCargar, K.D. Raine. 2014. Choosing healthier foods in recreational sports settings: A mixed methods investigation of the impact of nudging and an economic incentive. *Int J Behav Nutr Phys Act*. **11**(1) 6.
* Olstad, D.L., J. Vermeer, L.J. McCargar, R.J. Prowse, K.D. Raine. 2015. Using traffic light labels to improve food selection in recreation and sport facility eating environments. *Appetite*. **91** 329-335.
Ouellette, J.A., W. Wood. 1998. Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychol Bull*. **124**(1) 54-74.
* Perry, C.L., D.B. Bishop, G.L. Taylor, M. Davis, M. Story, C. Gray, S.C. Bishop, R.A.W. Mays, L.A. Lytle, L. Harnack. 2004. A randomized school trial of environmental strategies to encourage fruit and vegetable consumption among children. *Health Educ Behav*. **31**(1) 65-76.
* Policastro, P., T. Palm, J. Schwartz, G. Chapman. 2017. Targeted calorie message promotes healthy beverage consumption better than charity incentive. *Obesity*. **25**(8) 1428-1434.
* Policastro, P., Z. Smith, G. Chapman. 2015. Put the healthy item first: Order of ingredient listing influences consumer selection. *J Health Psychol*. **22**(7) 853-863.
Prelec, D., G. Loewenstein. 1998. The red and the black: Mental accounting of savings and debt. *Marketing Sci*. **17**(1) 4-28.
* Pulos, E., K. Leng. 2010. Evaluation of a voluntary menu-labeling program in full-service restaurants. *Am J Public Health*. **100**(6) 1035-1039.

* Redden, J.P., T. Mann, Z. Vickers, E. Mykerezi, M. Reicks, S. Elsbernd. 2015. Serving first in isolation increases vegetable intake among elementary schoolchildren. *PLoS One*. **10**(4) e0121283.
* Reicks, M., J.P. Redden, T. Mann, E. Mykerezi, Z. Vickers. 2012. Photographs in lunch tray compartments and vegetable consumption among children in elementary school cafeterias. *JAMA*. **307**(8) 784-785.
* Roberto, C.A., P.D. Larsen, H. Agnew, J. Baik, K.D. Brownell. 2010. Evaluating the impact of menu labeling on food choices and intake. *Am J Public Health*. **100**(2) 312-318.

Robinson, E. 2017. The science behind smarter lunchrooms. *PeerJ Preprints*. **5** e3137v3131.

Robinson, E., S. Nolan, C. Tudur-Smith, E.J. Boyland, J.A. Harrold, C.A. Hardman, J.C. Halford. 2014. Will smaller plates lead to smaller waists? A systematic review and meta-analysis of the effect that experimental manipulation of dishware size has on energy consumption. *Obes Rev*. **15**(10) 812-821.

Rothstein, H.R., A.J. Sutton, M. Borenstein. 2006. *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Roy, R., B. Kelly, A. Rangan, M. Allman-Farinelli. 2015. Food environment interventions to improve the dietary behavior of young adults in tertiary education settings: A systematic literature review. *J Acad Nutr Diet*. **115**(10) 1647-1681.e1641.

Rozin, P., C. Fischler, S. Imada, A. Sarubin, A. Wrzesniewski. 1999. Attitudes to food and the role of food in life in the u.S.A., japan, flemish belgium and france: Possible implications for the diet-health debate. *Appetite*. **33**(2) 163-180.

Rozin, P., K. Kabnick, E. Pete, C. Fischler, C. Shields. 2003. The ecology of eating: Smaller portion sizes in france than in the united states help explain the french paradox. *Psychol Sci*. **14**(5) 450-454.

* Rozin, P., S. Scott, M. Dingley, J.K. Urbanek, H. Jiang, M. Kaltenbach. 2011. Nudge to nobesity i: Minor changes in accessibility decrease food intake. *Judgm Decis Mak*. **6**(4) 323-332.

Scheibehenne, B., P.M. Todd, B. Wansink. 2010. Dining in the dark. The importance of visual cues for food consumption and satiety. *Appetite*. **55**(3) 710-713.

Schwartz, J., D. Mochon, L. Wyper, J. Maroba, D. Patel, D. Ariely. 2014. Healthier by precommitment. *Psychol Sci*.

* Schwartz, J., J. Riis, B. Elbel, D. Ariely. 2012. Inviting consumers to downsize fast-food portions significantly reduces calorie consumption. *Health Aff*. **31**(2) 399-407.
* Schwartz, M.B. 2007. The influence of a verbal prompt on school lunch fruit consumption: A pilot study. *Int J Behav Nutr Phys Act*. **4**(1) 6.

Seymour, J.D., A. Lazarus Yaroch, M. Serdula, H.M. Blanck, L.K. Khan. 2004. Impact of nutrition environmental interventions on point-of-purchase behavior in adults: A review. *Prev Med*. **39**(2) 108-136.

* Shah, A.M., J.R. Bettman, P.A. Ubel, P.A. Keller, J.A. Edell. 2014. Surcharges plus unhealthy labels reduce demand for unhealthy menu items. *J Marketing Res*. **51**(6) 773-789.

Shiv, B., A. Fedorikhin. 1999. Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *J Cons Res*. **26**(December) 278-292.

Simonsohn, U., L.D. Nelson, J.P. Simmons. 2014a. P-curve and effect size: Correcting for publication bias using only significant results. *Perspect Psychol Sci*. **9**(6) 666-681.

Simonsohn, U., L.D. Nelson, J.P. Simmons. 2014b. P-curve: A key to the file-drawer. *J Exp Psychol Gen*. **143**(2) 534.

Sinclair, S.E., M. Cooper, E.D. Mansfield. 2014. The influence of menu labeling on calories selected or consumed: A systematic review and meta-analysis. *J Acad Nutr Diet*. **114**(9) 1375-1388.e1315.

Skov, L.R., S. Lourenco, G.L. Hansen, B.E. Mikkelsen, C. Schofield. 2013. Choice architecture as a means to change eating behaviour in self-service settings: A systematic review. *Obes Rev*. **14**(3) 187-196.

Sniehotta, F.F., J. Presseau, V. Araújo-Soares. 2014. Time to retire the theory of planned behaviour. *Health Psychol Rev*. **8**(1) 1-7.

Snook, K.R., A.R. Hansen, C.H. Duke, K.C. Finch, A.A. Hackney, J. Zhang. 2017. Change in percentages of adults with overweight or obesity trying to lose weight, 1988-2014. *JAMA*. **317**(9) 971-973.

Srinivasan, S., M. Vanhuele, K. Pauwels. 2010. Mind-set metrics in market response models: An integrative approach. *J Marketing Res*. **47**(4) 672-684.

* Steenhuis, I., P. van Assema, G. van Breukelen, K. Glanz, G. Kok, H. de Vries. 2004. The impact of educational and environmental interventions in dutch worksite cafeterias. *Health Promot Int*. **19**(3) 335-343.

Sunstein, C.R. 2018a. “Better off, as judged by themselves”: A comment on evaluating nudges. *International Review of Economics*. **65**(1) 1-8.

Sunstein, C.R. 2018b. Misconceptions about nudges. *Journal of Behavioral Economics for Policy*. **2**(1) 61-67.

Sutin, A.R., L. Ferrucci, A.B. Zonderman, A. Terracciano. 2011. Personality and obesity across the adult life span. *J Pers Soc Psychol*. **101**(3) 579.

* Tal, A., B. Wansink. 2015. An apple a day brings more apples your way: Healthy samples prime healthier choices. *Psychol Mark*. **32**(5) 575-584.

* Tandon, P.S., C. Zhou, N.L. Chan, P. Lozano, S.C. Couch, K. Glanz, J. Krieger, B.E. Saelens. 2011. The impact of menu labeling on fast-food purchases for children and parents. *Am J Prev Med*. **41**(4) 434-438.
Thaler, R.H., C.R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books, New York.
Thapaa, J.R., C.P. Lyford. 2014. Behavioral economics in the school lunchroom: Can it affect food supplier decisions? A systematic review. *Int Food Agribus Man*. **17**(A) 187-208.
* Thomas, J.M., A. Ursell, E.L. Robinson, P. Aveyard, S.A. Jebb, C.P. Herman, S. Higgs. 2017. Using a descriptive social norm to increase vegetable selection in workplace restaurant settings. *Health Psychol*. **36**(11) 1026.
* Thorndike, A.N., J. Riis, L.M. Sonnenberg, D.E. Levy. 2014. Traffic-light labels and choice architecture: Promoting healthy food choices. *Am J Prev Med*. **46**(2) 143-149.
* Thorndike, A.N., L. Sonnenberg, J. Riis, S. Barraclough, D.E. Levy. 2012. A 2-phase labeling and choice architecture intervention to improve healthy food and beverage choices. *Am J Public Health*. **102**(3) 527-533.
* Turnwald, B.P., D.Z. Boles, A.J. Crum. 2017. Association between indulgent descriptions and vegetable consumption: Twisted carrots and dynamite beets. *JAMA Internal Medicine*. **177**(8) 1216-1218.
van Ittersum, K., B. Wansink. 2012. Plate size and color suggestibility: The delboeuf illusion's bias on serving and eating behavior. *J Cons Res*. **39**(2) 215-228.
* van Ittersum, K., B. Wansink. 2013. Extraverted children are more biased by bowl sizes than introverts. *PLoS One*. **8**(10) e78224.
* van Kleef, E., O. van der Broek, H. van Trijp. 2015. Exploiting the spur of the moment to enhance healthy consumption: Verbal prompting to increase fruit choices in a self-service restaurant. *Appl Psychol Health Well Being*. **7**(2) 149-166.
* Vanderlee, L., D. Hammond. 2014. Does nutrition information on menus impact food choice? Comparisons across two hospital cafeterias. *Public Health Nutr*. **17**(06) 1393-1402.
* Vasiljevic, M., E. Cartwright, M. Pilling, M.-M. Lee, G. Bignardi, R. Pechey, G.J. Hollands, S.A. Jebb, T.M. Marteau. 2018. Impact of calorie labelling in worksite cafeterias: A stepped wedge randomised controlled pilot trial. *Int J Behav Nutr Phys Act*. **15**(1) 41.
Viechtbauer, W. 2010. Conducting meta-analyses in r with the metafor package. *J Stat Softw*. **36**(3) 1-48.
Viechtbauer, W., M.W.L. Cheung. 2010. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. **1**(2) 112-125.
Wake, M. 2018. The failure of anti-obesity programmes in schools. *BMJ*. **360**.
Walsh, J.W. 1995. Flexibility in consumer purchasing for uncertain future tastes. *Marketing Sci*. **14**(2) 148-165.
Wansink, B. 2015. Change their choice! Changing behavior using the can approach and activism research. *Psychol Mark*. **32**(5) 486-500.
* Wansink, B., H. Bhana, M. Qureshi, J.W. Cadenhead. 2016. Using choice architecture to create healthy food interventions in food pantries. *J Nutr Educ Behav*. **48**(7) S35-S36.
Wansink, B., P. Chandon. 2006. Can 'low-fat' nutrition labels lead to obesity? *J Marketing Res*. **43**(4) 605-617.
Wansink, B., P. Chandon. 2014. Slim by design: Redirecting the accidental drivers of mindless overeating. *J Consum Psychol*. **24**(3) 413-431.
* Wansink, B., A.S. Hanks. 2013. Slim by design: Serving healthy foods first in buffet lines improves overall meal selection. *PLoS One*. **8**(10) e77055.
* Wansink, B., D.R. Just, A.S. Hanks, L.E. Smith. 2013. Pre-sliced fruit in school cafeterias: Children's selection and intake. *Am J Prev Med*. **44**(5) 477-480.
Wansink, B., D.R. Just, C.R. Payne. 2012a. Can branding improve school lunches? *Arch Pediatr Adolesc Med*. **166**(10) 967-968.
Wansink, B., D.R. Just, C.R. Payne, M.Z. Klinger. 2012b. Attractive names sustain increased vegetable intake in schools. *Prev Med*. **55**(4) 330-332.
* Wansink, B., J. Kim. 2005. Bad popcorn in big buckets: Portion size can influence intake as much as taste. *J Nutr Educ Behav*. **37**(5) 242-245.
Wansink, B., C. Payne, C. Werle. 2008. Consequences of belonging to the "clean plate club". *Arch Pediatr Adolesc Med*. **162**(10) 994-995.
* Wansink, B., C.R. Payne. 2007. Counting bones: Environmental cues that decrease food intake. *Percept Mot Skills*. **104**(1) 273-276.
Wansink, B., C.R. Payne, P. Chandon. 2007. Internal and external cues of meal cessation: The french paradox redux? *Obesity*. **15**(12) 2920-2924.
* Wansink, B., K. van Ittersum. 2013. Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *J Exp Psychol Appl*. **19**(4) 320-332.
* Wansink, B., K. van Ittersum, J.E. Painter. 2006. Ice cream illusions: Bowls, spoons, and self-served portion sizes. *Am J Prev Med*. **31**(3) 240-243.

* Wansink, B., K. van Ittersum, C.R. Payne. 2014. Larger bowl size increases the amount of cereal children request, consume, and waste. *J Pediatr*. **164**(2) 323-326.
* Webb, K.L., L.S. Solomon, J. Sanders, C. Akiyama, P.B. Crawford. 2011. Menu labeling responsive to consumer concerns and shows promise for changing patron purchases. *Journal of Hunger and Environmental Nutrition*. **6**(2) 166-178.
Wertenbroch, K. 1998. Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Sci*. **17**(4) 317-337.
Wilson, A.L., E. Buckley, J.D. Buckley, S. Bogomolova. 2016. Nudging healthier food and beverage choices through salience and priming. Evidence from a systematic review. *Food Qual Prefer*. **51** 47-64.
* Wilson, N.L., D.R. Just, J. Swigert, B. Wansink. 2017. Food pantry selection solutions: A randomized controlled trial in client-choice food pantries to nudge clients to targeted foods. *J Public Health*. **39**(2) 366-372.
Zlatevska, N., C. Dubelaar, S.S. Holden. 2014. Sizing up the effect of portion size on consumption: A meta-analytic review. *J Marketing*. **78**(May) 140-154.

Online Appendix to

"Which Healthy Eating Nudges Work Best?

A Meta-Analysis of Field Experiments"

This online appendix contains the following sections:

Appendix A. Search strategy and flow chart

Appendix B. Publication bias

Appendix C. Subcategory analyses

Appendix D. Interaction analyses

Appendix A: Search strategy, keyword selection, and flow diagram

This meta-analysis focuses on articles describing nudge interventions, without restriction on the population. Specific search terms were developed in accordance with the SPICE (Setting, Population, Intervention, Comparison, Evaluation) framework (Booth 2006) (see Table A1). We mainly searched for interventions that involved nudging, choice architecture, or behavioral economics. We considered all articles published in the English language that reported a nudge intervention in a field setting. As shown in Table A2, key terms within the SPICE framework were combined using the Boolean operators "AND" and "OR."

Table A1: Application of the SPICE framework for keyword selection

SPICE	Keywords
Setting	Food; eat*; fruit; vegetable; drink; beverage; diet; nutriti*; (un)healthy; calorie
Population	None assigned, interested in all populations
Intervention	Nudg*; choice architect*; behavioral economics; behavioral intervention
Comparator	Field study; field experiment
Evaluation	Selection; consumption; sales; choice

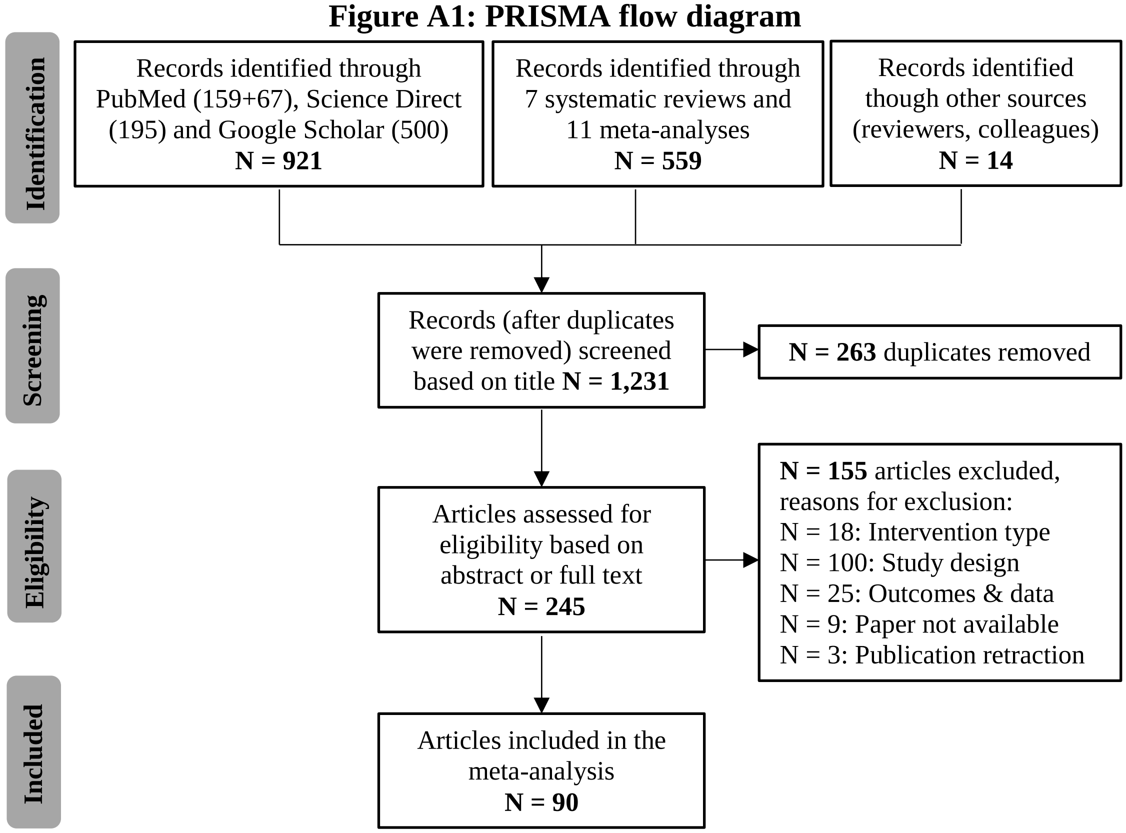
Table A2: Specific keywords used in database search using Boolean operators

Database	Keywords
Science Direct	("food" OR "eat" OR "fruit" OR "vegetable" OR "drink" OR "beverage" OR "diet" OR "nutriti" OR "calorie") AND ("nudg" OR "choice architect" OR "behavioral economics" OR "behavioral intervention") AND ("field study" OR "field experiment") AND NOT ("lab study" OR "lab experiment") AND ("selection" OR "consumption" OR "sales" OR "choice")
PubMed	("food" OR "eat*" OR "fruit" OR "vegetable" OR "drink" OR "beverage" OR "diet" OR "nutriti*" OR "calorie") AND ("nudg*" OR ("choice" AND "architect*"))
Google Scholar	"Nudge behavioral intervention food selection consumption choice"

Figure A1 reports the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Moher et al. 2009) flow diagram. The search strategy was first applied to three electronic databases: PubMed, Science Direct, and Google Scholar. The search was initially conducted in January 2016 and updated on October 2017. In June 2018, we updated the search by searching only through PubMed (67 references) and our living meta-analysis questionnaire (?? References).

In the database search (921 references), we focused intervention-based articles as well as review-based articles. In fact, we found 7 systematic reviews and 11 meta-analyses on the topic. We included in our identification base all 559 references cited in these review-based articles. Last, we also included 14 other references identified through other sources. After removing duplicates and references based on titles, we evaluated 245 articles. Of these, 155 articles were excluded for various reasons (see Figure 1), and 90 articles were included in the meta-analysis.

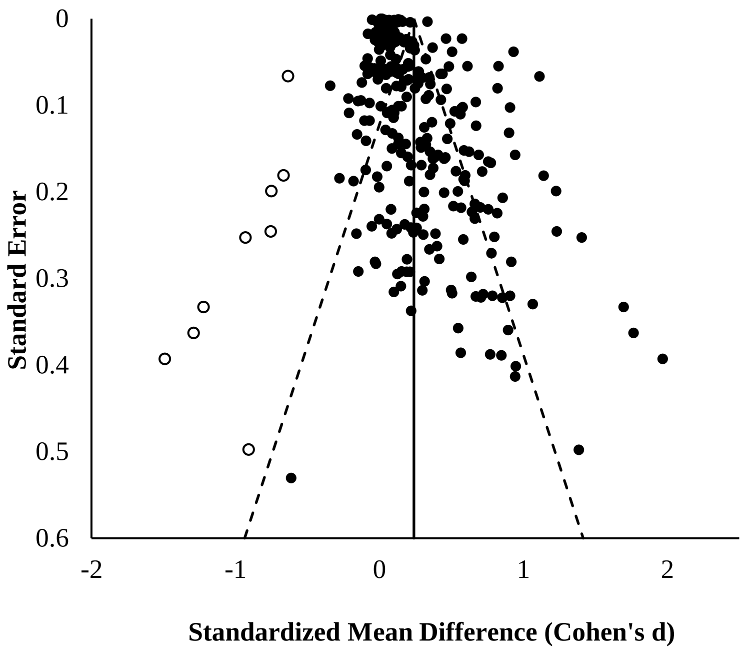
Figure A1: PRISMA flow diagram



Appendix B: Publication bias

In Figure A2, the funnel plot displays each observation as a function of the effect size (standardized mean difference or Cohen's *d*) and the standard error. Several observations appeared outside of the funnel on the right-hand side, suggesting a potential publication bias.

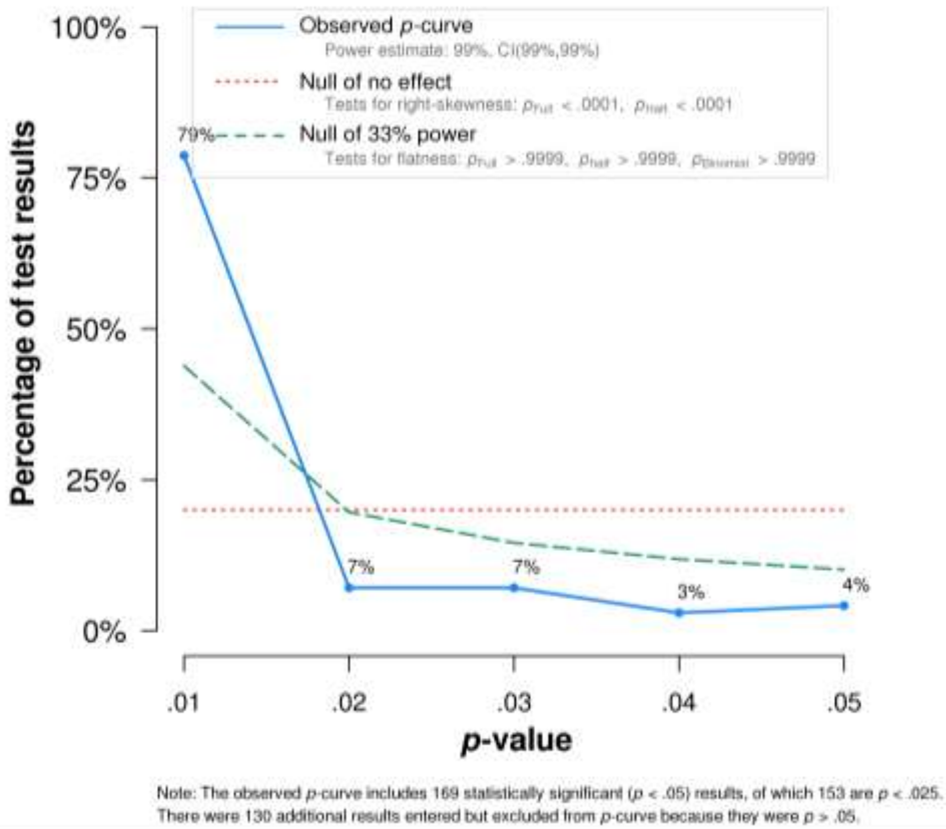
Figure A2: Funnel plot



First, using Duval and Tweedie's (2000) trim and fill method (not available for a three-level analysis), we included 9 missing studies (represented by the white dots in Figure 2), resulting in a slightly lower estimated effect size (d = .20, se = 0.2, p < .001) than in the unadjusted two-level analysis (d = .22, se = .02, p < .001). However, both estimated effect sizes are positive, considered small, and significantly different from zero.

Second, following Viechtbauer and Cheung (2010), we performed sensitivity analyses by removing 34 (30) observations for which the leverage (Cook's distance) was more than twice the average leverage (Cook's distance) in the overall sample. In both cases, the adjusted effect size is only slightly lower (respectively d = .27, se = .04, p < .001; and d = .23, se = .03, p < .001) compared to the original three-level analysis (d = .27, se = .03, p < .001).

Figure A3: *P*-curve



Third, we also produced the *p*-curve (Simonsohn et al. 2014b) for all effect sizes included in our sample. As shown in Figure A3, the *p*-curve is strongly right-skewed ($p < .001$) and indicates the presence of evidential value. Note that 130 out of 299 effect sizes (43%) were not significant. A closer examination of non-significant effects shows that only 16 out of 90 articles (18%) published only non-significant results, 31 papers (34%) published significant and non-significant effect sizes (e.g., the effect size is significant for fruits but not significant for vegetables) and the rest (48%) published only significant results. Using the procedure presented in Simonsohn et al. (2014a), we also estimated the corrected *p*-curve effect size on the sample of 169 significant effect sizes. The corrected effect size estimated through bootstrap ($d = .19$, $se = .03$, $p < .001$) was lower than the "naïve" estimate on the 169 significant effect sizes without the *p*-curve correction ($d = .35$, $se = .02$, $p < .001$), and lower than the "earnest" (regular, intercept-only) overall average effect size including all observations ($d = .22$, $se = .03$, $p < .001$). Last, the *p*-

curve effect size is similar to the average effect size from the overall three-level model including all observations and covariates (d = .23, se = .04, p < .001). All four estimates are positive, considered of small magnitude, and significantly different from zero.

Appendix C: Subcategory analyses

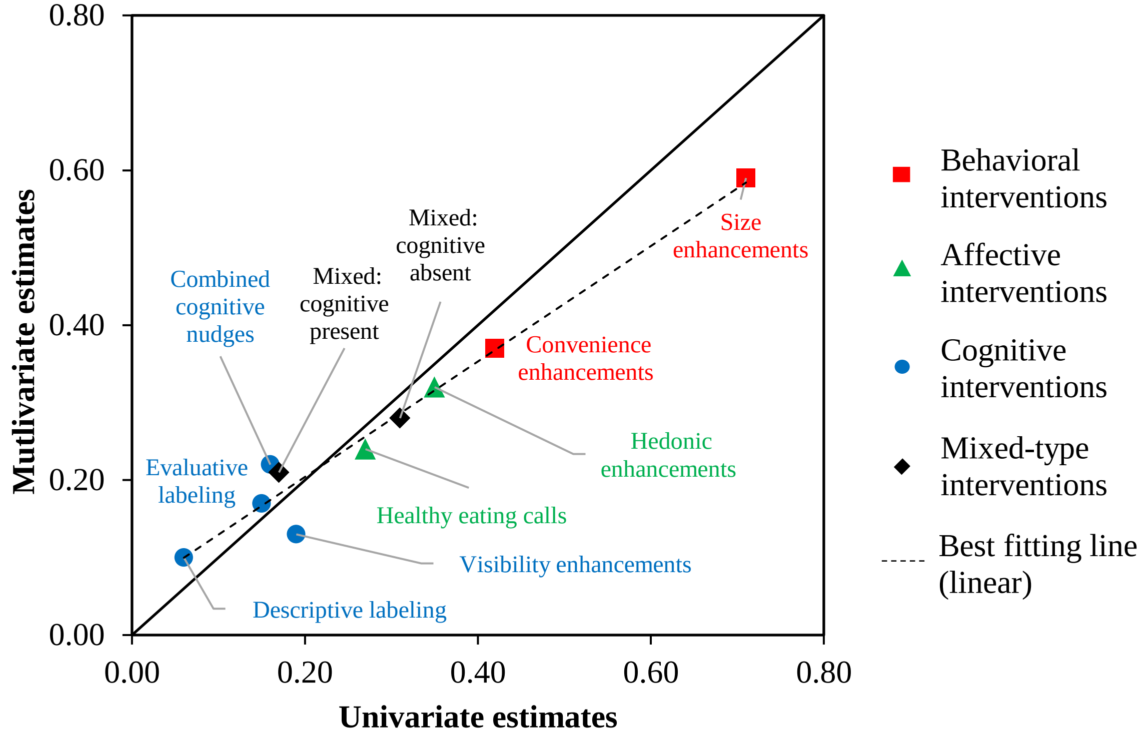
In Table A3 and Figure A4, we report the estimates and standard errors for the multivariate model using the detailed categorization for intervention type (partly plotted in Figure 3 in the main document).

Table A3: Estimates for detailed intervention types

		Univariate		**Multivariate**	
	k	***d***	***se***	***d***	***se***
Cognitive interventions					
Descriptive labeling	34	.06	.05	.10	.07
Evaluative labeling	43	.15*	.05	.17*	.06
Visibility enhancements	25	.19*	.08	.13	.08
Mixed: Cognitive only	14	.16*	.07	.22*	.09
Affective interventions					
Healthy eating calls	42	.27*	.05	.24*	.06
Hedonic enhancements	7	.35*	.10	.32*	.10
Behavioral interventions					
Convenience enhancements	65	.42*	.05	.37*	.05
Size enhancements	17	.71*	.09	.59*	.09
Mixed interventions					
Mixed: Cognitive present	43	.17*	.07	.21*	.08
Mixed: Cognitive absent	9	.31*	.08	.28*	.09

$^{*}p$ < .05

Figure A4: Effect sizes of intervention types estimated by univariate and multivariate models



Appendix D: Interaction analyses

The multivariate model shown in equation 7 and Table 4 does not lend itself to interaction analyses because it already requires estimating 16 coefficients. Adding interactions would require estimating more than 30 coefficients, which would quickly exhaust the available degrees of freedom, especially for some combinations with few data. Hence, we added interactions to a simplified model with only one coefficient for each of the 8 types of predictors. Essentially, we recoded or grouped categories together to provide a single (linear or binary) coefficient per predictor.

Simplified model. First, we used linear coding for intervention type (-1 = cognitive, 0 = affective, and 1 = behavioral). We included the mixed interventions into the cognition-affect-behavior categorization according to the first stage that they sought to influence. Hence, mixed interventions including at least one cognitive intervention ("mixed: cognitive present") were included among cognitive interventions. Mixed interventions excluding a cognitive one ("mixed:

cognitive absent") were included among affective ones (no study used a mixed intervention of behavioral nudges). We also used a linear coding for outcome measure (-1 = total eating, 0 = healthy eating and mixed eating, and 1 = unhealthy eating), while healthy and mixed eating were grouped together because of a small difference in effect sizes (see Table 2). Third, location was ANOVA coded (-½ for grocery stores vs. ½ for onsite or offsite cafeterias), while offsite and onsite locations were grouped together because of a small difference in effect sizes (see Table 2). Fourth, study design was ANOVA coded (-½ for pre-post and ½ for single and double difference), while single and double difference were grouped together because of a small difference in effect sizes (see Table 2). The rest of the variables were unchanged; that is, outcome behavior (selection vs. consumption), age (children vs. adults), and country (other vs. US) were ANOVA coded, while duration was measured in weeks and mean centered.

Table A4: Interactions analyses

	Model 1		Model 2	
Predictor	$\boldsymbol{\beta}$	**se**	$\boldsymbol{\beta}$	**se**
Constant	$.22^{***}$	.04	$.21^{***}$	.04
Intervention type (cognitive to affective to behavioral)	$.13^{***}$	.03	$.11^{**}$	.04
Outcome behavior (selection vs. consumption)	.03	.04	.05	.04
Outcome measure (total to healthy/mixed to unhealthy)	$.11^{***}$	.03	$.08^{*}$	.03
Population setting (cafeterias vs. grocery stores)	$.14^{*}$	.07	.10	.06
Population age (adults vs. children)	.01	.05	.04	.04
Population country (US vs. other)	.09	.05	$.11^{*}$	.04
Study duration (duration in weeks, mean centered)	-.002	.001	-.005	.003
Study design (single & double difference vs. pre-post)	$.15^{***}$	.04	$.13^{**}$	.04
Outcome behavior × intervention type			$.14^{**}$	.04
Outcome measure × intervention type			-.02	.03
Population setting × intervention type			-.05	.07
Population age × intervention type			$.15^{***}$	.05
Population country × intervention type			$.14^{**}$	.05
Study duration × intervention type			-.003	.003
Study design × intervention type			.033	.04
K (observations)	299		299	
N (studies)	96		96	
R^2	48%		58%	
LR test vs. intercept-only model	74^{***}		97^{***}	

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

Results. We first estimated a model with only main-effects (Model 1). Its R^2 decreased only slightly compared to the model with the categorical coding reported in the main text (from .46 to .44), and none of the hypotheses tests changed. To examine when each type of intervention is most effective, we included an interaction term between intervention type and all the other predictors. Table A4 shows that the effects of intervention type are stronger 1) for consumption than for selection, 2) for adults than for children, and 3) in the US than in other countries.

Because repeated visits are more likely to occur in work cafeteria than restaurants, we also ran a model including an interaction between study location and study duration. The interaction term was not significant (p = .58).