2019-06-12

Behavioral Economics: Ethics and Integrative Thinking

Nina Mazar. (2019). "Behavioral Economics: Ethics and Integrative Thinking." In A. Samson (Ed.), The Behavioral Economics Guide 2019 (with an Introduction by Uri Gneezy) (pp. 2-15). Retrieved from https://www.behavioraleconomics.com.

https://hdl.handle.net/2144/39951

Behavioral Economics: Ethics and Integrative Thinking

Nina Mažar

nina@ninamazar.com

'Behavioral Economics', 'Behavioral Science', 'Behavioral Insights', 'Behavioral Change', 'Nudge', etc. People all around the world, namely, experts and novices, academics and practitioners, policymakers and decision-makers, private and public individuals, use these various terms, mostly interchangeably.

As my last year's predecessor and BU Questrom colleague Robert Metcalfe wrote in his editorial, it is an exciting time to be in our field. We have seen Nobel prizes awarded to some of our field's brightest, with several reputable academics making our work more palatable to the mainstream through bestselling books and advising and collaborating with for-profit, non-for-profit, multinational, and government organizations to use behavioral insights to tackle problems ranging from voter turnout, to tax compliance, to healthy eating, etc. In fact, increasingly more companies and organizations are looking to apply insights from behavioral economics in their daily work: Internally with their employees (e.g. to reduce biases [Banuri, Dercon, and Gauri, 2018; Morewedge et al., 2015] or to increase diversity [e.g. Chilazi, Asundi and Bohnet, 2019]) and externally in client and consumer-facing projects (e.g. to understand whether consumers are paying off credit card debt as theory would predict [Quispe-Torreblanca, Stewart, Gathergood, & Loewenstein, 2019] or how to reduce credit card delinquency [Mazar, Mochon, and Ariely, 2018]). This in turn spurs the demand for our expertise and the creation of Masters and various other certificate programs in behavioral economics (see graduate programs listed in this publication).

A great deal has been written about the remarkable progress that has been made over the last few years in making behavioral economics a household name. Nonetheless, with fame and recognition also comes responsibility, which in turn requires humility and the willingness to self-reflect critically, and to question and improve, in order to get better and to rise to one's full potential. This notion is true for humans, but it is also for a discipline.

And as we as a discipline reflect upon our achievements and failures, as well as our potential and challenges, I believe two of the most salient topics for our field are – and will revolve around increasingly more so – ethics (defined in a broad sense) and integrative thinking. While I'm presenting them separately as distinct topics, for me at least, they are intricately connected and therefore will feed into each other.

Ethics in a Narrow Sense: Manipulation

Various debates have discussed government and multinational organizations using behavioral insights, and whether this approach is ethical. Some critics have warned that behavioral economics interventions can interfere with people's autonomy and thus be coercive and manipulative (Bennhold, 2013; Blumenthal-Barby, 2012, p. 352; Hausman & Welch, 2010, pp. 128-130; Reiss, 2013, p. 294; Wilkinson, 2013a, 2013b); for example, several organizations object to the use of default opt-in policies for organ donation, because they believe it may take advantage of vulnerable individuals (see discussion in Baldwin, 2014).

However, their objections relate not only to interventions: A recent publication (Meyer et al., 2019) showed that people are more likely to object to an experiment designed to test the relative effectiveness of two different positive treatments, before deciding which one to implement universally, than to have one of the two treatments universally implemented without comparative testing. This finding is robust among novices and experts, across domains ranging from healthcare, to improving recruitment, to poverty reduction, as well as among respondents with varying degrees of educational attainment and scientific literacy. For example, in one scenario, “a hospital director[...] wants to reduce deadly and costly catheter-related hospital infections and thinks that providing doctors with a checklist of standard safety pre-cautions might help” (p. 2). One intervention consisted of having the standard safety precautions printed on the back of doctors’ hospital ID badges, whilst the other intervention involved displaying a poster with the standard safety precautions in hospital rooms.

What is particularly puzzling about their findings is that these conditions are very different from, for example, the setting that sparked the outcry over Facebook’s experiment with the emotional content of their users’ newsfeed (Goel, 2014). Unlike the Facebook study that was designed to test how both positive as well as negative emotions spread on social media (Kramer, Guillory, & Hancock, 2014), across all the domains that Meyer et al. (2019) tested, neither of the treatments were negative per se, and the goal was always very clearly to improve wellbeing. These latest findings pose an interesting ethical dilemma for behavioral economists: On the one hand, we believe in the gold standard of randomized control trials as being – most of the time – one of the most unbiased and ethical ways to improve wellbeing, and yet the population whose welfare we want to improve may have strong objections about this methodology. I believe it is our responsibility to understand from where these opposing views about experimentation originate, in order to initiate actively a discourse and to encourage actions (e.g. the establishment of independent ethics review boards akin to those for academic research; informed referendums) to overcoming this tension.

Ethics in a Broad Sense: Long-Term and Side-Effects

Long-Term Effects

Less time seems to be spent on talking about how we ensure we assess the full impact of behavioral economics interventions to make sure there are no undesired side-effects and problematic longer-term effects. After all, if we want our intervention ideas to be taken seriously, be applied in the real world, and scaled up, we need to be willing to take responsibility for their consequences.

For example, it can be relatively easy, particularly online, to run some quick experiments, such as having an honor code displayed at the top or the bottom of an automobile insurance audit form, analyze the data, and learn that the reported odometer mileages are higher and thus likely more honest when the honor code is at the top of the form (Shu, Mazar, Gino, Ariely, & Bazerman, 2012). But what were to happen at the next audit wave if the company went back to their regular form with the honor code at the bottom? Would the single-shot intervention effect persist (Frey & Rogers, 2014), or, more critically, would it backfire and lead the same people who reported more honestly to compensate (“Licensing”; e.g. Monin & Miller, 2001) for their “losses” at the subsequent audit (see Ghesla, Grieder, & Schmitz, 2019, for a laboratory study of a default effect intervention on subsequent donation decisions)?

Brookmeyer et al. (2016) examined the impact of enforcement email interventions on tax compliance in Costa Rica, not only immediately afterwards, but also a year later (without any additional communication), and they found: (1) An immediate increase in income tax filing rates and payment rates and (2) that the effect persisted in the medium term. In a different context, the behavioral economics team of the Australian Government (2019) recently reported on an SMS message intervention program encouraging consumers who consistently paid only the minimum of their monthly credit card debt, to repay more. Immediately after the experiment ended, as well as six months after the first SMS was sent, "those who had received a message had balances $249 lower on average than those who had not received a message ($10,374 versus $10,623). At twelve months, this result was even more pronounced, at $365 ($9,206 versus $9,571)" (p. 12/13). This result is astonishing, especially given that it could equally have been that the increases in repayments were offset later by reduced payments or increased expenditures.

Whilst examining the longer-term persistency of behavioral interventions is an important first step, another critically important temporal aspect of behavioral interventions is whether repeated exposure to the same intervention alters its effectiveness. That is, would the automobile insurance company in our experiment (Shu et al., 2012) find the same impact of the honor code at the top of the form if a client saw it again the next time the audit form was sent out, or would we find adaptation resulting in a diminished or even no-effect (perhaps even reversing)? I recently conducted an experiment with my co-authors Nicole Robitaille and Julian House (2018) with the Ontario Finance Ministry, where we tested the effectiveness of a repeated implementation intention-intervention on organizations' likelihood to file their overdue taxes. In addition to replicating our findings across two waves, we found no evidence of habituation to our intervention for companies exposed to our intervention over two consecutive years.

Does this mean this finding would continue to replicate in the years to come? Or would habituation only be a possibility if exposure to our intervention was more frequent than once a year? The response to such a question probably depends on various factors, one of which is to what extent do novelty or unfamiliarity contribute to an intervention's effectiveness, or to what extent are individuals motivated to not pay attention to an intervention? When it comes to questions of morality, for example, where individuals want to think of themselves as honest but are also tempted to transgress for some personal gain, being reminded of one's morals is useful (e.g. Mazar, Amir, & Ariely, 2008), but individuals will likely also try to not pay attention to these reminders so that they feel less bad about themselves if engaging in the transgression. Moreover, the less attention-grabbing an intervention or the easier it is to circumvent as one gains familiarity with it, the less effective it should become.

On the other hand, where undesirable behavior is due simply to forgetting, the more often a reminder, perhaps the better the possible outcome. For example, Lombard, Lombard, and Winnett (1995) tested the effectiveness of phone call prompts to increase walking, varying the frequency of these prompts. They found that participants who were prompted every seven days walked for significantly more weeks than the participants who were prompted once every 21 days – a statistically significant difference that was maintained over a three-month time frame post-intervention.

Side-Effects

Finally, one also needs to consider side-effects on other types of behaviors and/or other population types. For example, Tiefenbeck et al. (2013) demonstrated that those who received weekly feedback on their domestic water consumption lowered their water use but at the same time increased their electricity consumption. In another example, Damgaard and Gravert (2018) found that frequent reminders in a fundraising context, in an attempt to increase donations, also caused the hidden cost of canceled subscriptions by people who were annoyed about being approached in such a way. And yet in another instance, Anderson and Robinson

(2017) reported that those who had unjustly high beliefs in their financial literacy were affected negatively by a nudge that encouraged people to take care of their retirement savings. There are plenty more examples in the existing literature, and rather than summarizing them herein, I would like to refer the interested reader to the welfare section of my predecessor's editorial (Metcalfe, 2018, pp. 14-16) and to a recent contribution by Thunström (2019) on behavioraleconomics.com.

The main takeaway here should be that we need to be sensitive to an intervention's potentially differing effects on different behaviors and population types, and its efficiency in comparison to other, for example, more traditional financial incentives: Only then can we fully assess the importance and impact of our intervention. One nice illustration of such a comprehensive approach is a current working paper by Butera, Metcalfe, Morrison and Taubinsky (2019), in which the authors first develop and then empirically test (in the context of YMCA attendance) a framework for quantifying the social efficiency of social recognition programs. In a nutshell, they find that the crux of the issue is the shape of the underlying social recognition utility function, which determines the generation of deadweight loss.

Summary

Impact, net benefit, or welfare effect analyses are important steps we need to embrace humbly if we wish to make the case that the use of behavioral insights is important and meaningful. This requires not only taking into account the time dimension (in terms of both frequency of exposure as well as duration of effects) and immediate costs of implementation, but also broadening our view on the relevant dependent variables and considering actively – even if it may be impossible to be comprehensive and foresee every possible effect – heterogeneity and other behaviors and attitudes that could be affected.

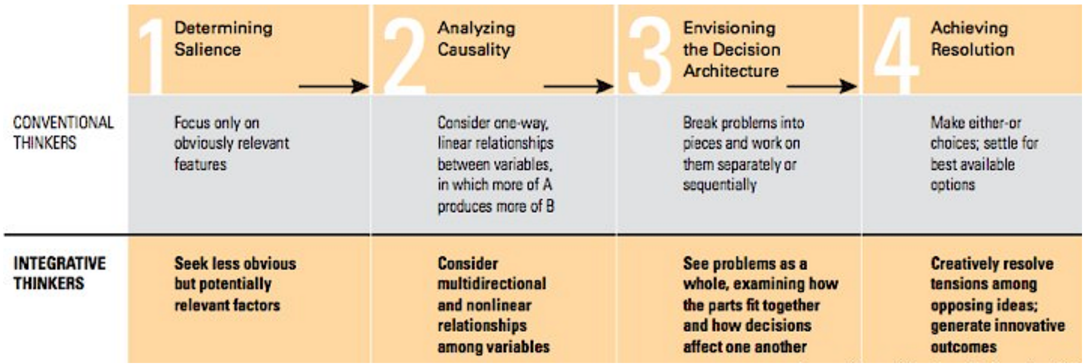
This of course creates at least two challenges for us behavioral economists: (1) For how long a period should we track the effects of our interventions before we feel comfortable recommending one of them for universal implementation and (2) How can we ensure we consider all other behaviors and group types in the population when examining side effects? These questions may be viewed as unfair to some; after all, and in particular for question (1), shouldn't organizations willing to apply behavioral insights know that people and contexts are constantly changing and thus be equally willing to reassess regularly the effectiveness of their interventions and update the design of their policies, programs, or actions? To what extent, therefore, should organizations share some of the responsibilities outlined above?

Rather than trying to answer these contentious questions, the solution may lie in us taking the responsibility to be more transparent about what we know and what alternatives are available; after all, it's not a secret: People don't like uncertainty (Ellsberg, 1961) and believe that experts should know the one best course of action (Rozenblit & Kalt, 2002). This notion, however, clashes with the fundamentals of a behavioral design process that believes in experimentation, randomization, and iteration, and places the burden on us behavioral scientists to set expectations correctly, challenge status quo thinking, and apply a rigorous, comprehensive process with high standards to our problem-solving.

Integrative Thinking (and Doing!)

The former dean at my previous institution and "the world's #1 management thinker" (Thinkers 50, 2017: a biannual ranking of the most influential global business thinkers), Roger Martin, defines 'integrative thinking' in his book *The Opposable Mind* (Martin, 2009) as "the ability to face constructively the tension of opposing ideas and, instead of choosing one at the expense of the other, generate a creative resolution of the tension in the form of a new idea that contains elements of the opposing ideas but is superior to each" (p. 27).

According to Martin, "integrative thinking is at once a mindset, a methodology and a pedagogy for problem solving".[1] In essence, it is the way successful leaders think. Figure 1 displays how an integrative thinker in comparison to a conventional thinker works through the four stages of a problem-solving process. Fundamental to the practice of integrative thinking is the recognition that others may interpret a same situation very differently, and an openness to consider the value inherent in these differing views. This empowers integrative thinkers to solve richer, more complex problems in innovative new ways.



	1 Determining Salience	2 Analyzing Causality	3 Envisioning the Decision Architecture	4 Achieving Resolution
CONVENTIONAL THINKERS	Focus only on obviously relevant features	Consider one-way, linear relationships between variables, in which more of A produces more of B	Break problems into pieces and work on them separately or sequentially	Make either-or choices; settle for best available options
INTEGRATIVE THINKERS	**Seek less obvious but potentially relevant factors**	**Consider multidirectional and nonlinear relationships among variables**	**See problems as a whole, examining how the parts fit together and how decisions affect one another**	**Creatively resolve tensions among opposing ideas; generate innovative outcomes**

Figure 1: Conventional thinkers vs integrative thinkers. *Source*: Martin (2007), How Successful Leaders Think, *Harvard Business Review*.

Integrative thinking as a mindset is also useful for us behavioral economists seeking to increase the reach of our discipline. It is easy to fall into the trap of thinking in terms of *either-or* decisions, for example either a nudge or a financial incentive to create long-term behavior change (for a comprehensive overview of the latter, see Uri Gneezy's introduction to this year's Guide), or either a behavioral approach or a design thinking approach. However, the truth of the matter is that combining approaches, perspectives, and views may oftentimes allow not only for more impactful effects, but also the ability to solve more complex problems. In my view, only through the willingness to engage in integrative thinking and doing can we realize the full potential of behavioral economics.

[1] See http://www.rotmanithink.ca/what-is-integrative-thinking

Behavioral Economics or Traditional Economics?

Just recently, George Loewenstein and Nick Chater (2017) wrote an insightful paper "Putting Nudges in Perspectives", in which they focused in particular on the real tension between traditional and behavioral economics, as summarized in Table 1:

Rationale for intervention	Type of intervention		
	Traditional economic (e.g. taxes and subsidies)	Hybrid policies (e.g. carefully 'framed' taxes and subsidies)	Behavioral
Traditional economic (e.g. externalities, asymmetric information)	A (pure economic theory)	B	C
Hybrids (e.g. company optimally responding to consumer biases)	D	E	F
Behavioral economic (e.g. internalities, bounded rationality)	G	H	I (pure behavioral economics)

Table 1: A taxonomy of policy interventions. *Source*: Loewenstein and Chater (2017).

In their paper, Loewenstein and Chater (2017) made the intuitive but not necessarily self-evident point that the rationale for an intervention (market-level, structural factors such as externalities, misaligned incentives, and information asymmetries vs. individual-level factors such as internalities and bounded rationality) and the type of intervention (traditional economic factors such as taxes, subsidies, or regulations vs. behavioral elements such as framing) do not need to be of the same kind. Typically, when people think of behavioral economics, they think of problems grounded in humans' bounded rationality that are solved with insights from behavioral science (cell I in Table 1). But as Loewenstein and Chater (2017) point out, "economic problems may have behavioural solutions, and [...] behavioural problems may have economic solutions. Moreover, understanding many of society's problems and formulating policy solutions will involve hybrids between traditional and behavioural economics, rather than pure applications of either" (p. 48). Their paper is filled with examples and references for each of the cells in Table 1, ranging from smoking, obesity, and retirement savings, to some of society's most challenging and complex problems of our time, such as climate change, income and wealth inequality, and changes in the nature of employment.

In essence, hybrid approaches allow one to go beyond solving the so-called "last mile" problems (Soman, 2015). Going after low-hanging fruit and fixing a 'broken' policy such as retirement savings by changing defaults, introducing automatic sign-ups, re-framing messages, or sending SMS reminders is a good, important, and easy way to help people become better off. However, we should not shy away from a broader, more flexible mindset that allows for behavioral economics to contribute to the "first mile" (see Soman's, 2015, p. 11: the strategy formulation and the program, product, or policy development stage; Loewenstein & Chater, 2017). As Ruth Schmidt (2019) wrote in her recent contribution to the *Behavioral Scientist*, in which she calls for the need for a systems view: "[...] when we fail to consider whether the system within which interventions take place is sound and equitable, and how correcting for one kind of change might leave out other important considerations or lead to unintended consequences, we risk optimizing behavior that perpetuates or amplifies broader systemic or cultural biases".

Behavioral Economics or What?

In order to give behavioral economics the chance to reach its potential fully, it also requires the willingness to engage in integrative thinking with fields and methodologies beyond standard economics. The social sciences more generally (in particular, not only psychology but also, for example, sociology and anthropology) as well as design thinking and AI are a few that easily come to mind.

Social Sciences

The social sciences focus more generally on the study of human behavior on various levels. The goal is to examine society and understand, for example, how and why humans and their minds work (psychology), how and why humans build and maintain social relationships and institutions (sociology), and how and why they interact and develop as a culture (anthropology). Behavioral economics as a discipline is grounded primarily in psychological insights (see e.g. Kahneman, 2011), thereby allowing us to change individual behavior successfully. However, we reach our limits when it comes to trying to change the behavior of entire groups of individuals, such as in the example of organizations, where additional factors such as hierarchy, power, responsibility, accountability, group identity, network effects, and many more play a role. In fact, even when it comes to decisions that are made in dyads, such as various intra-household financial decisions, we limit the effectiveness of our interventions if we target individual decision making only. Similarly, if we want to try to help address bigger societal issues like corruption, we need to understand the culture within which corruption is operating. For example, how did it develop, why is it persisting, what are people's views about corruption, to what extent is it generally assumed and accepted, etc.? In any of these examples, clearly it would be beneficial to familiarize ourselves with the insights taken from, and study methodologies used by, a broader set of social sciences.

Design Thinking

Design thinking draws on engineering and design methods to create innovation in product development and business processes. In comparison to behavioral economics, design thinking is comfortable with a much more holistic system view and, most importantly, with messiness and ambiguity, ethnographic exploration, and qualitative data. Messiness and ambiguity refer to the fact that design thinking is often used in settings where both the problem and the solution are unknown at the outset of the problem-solving process. Thus, even when the general direction of the problem may be clear, considerable time and effort are spent on defining and shaping the problem (Baeck & Gremett, 2012; Brown, 2008; Brown, 2009).

The reliance on ethnographic exploration, on the other hand, stands for a human values-centered approach that stresses open-mindedness and a willingness to approach a problem without any preconceived notions. "Assume you don't know." This type of beginner's mindset is lost so easily in a behavioral economics project when we focus on the low-hanging fruit, recycling previously tested interventions without really questioning whether this is the right context for it, or entertaining other possible underlying issues that may be served better in a different way. For example, when thinking about how to protect low-income teenage schoolgirls in South Africa, who are three times more likely to be HIV-positive than boys in their age group, one could come up with all kinds of message interventions to try to persuade teenagers to not have sex, have less sex, or use protection. Yet, through extensive exploration, a team around ideas42 (Datta et al., 2015) found that the main culprits were 1) more teenage girls getting involved with older men and (2) misperceptions about the link between age and HIV risk. In particular, they found that teenage schoolgirls often – and incorrectly – identified older men as safer sexual partners, focusing on their financial situation and maturity as signs of responsibility and discounting the actual risk created by the higher number of potential partners older men will have had. As a consequence, Datta et al. (2015) decided to develop and test a computer-based "HIV risk game" that increased treatment participants' understanding of the relationship between HIV risk and age by 1.65 times in comparison to control, and which resulted in significantly longer information retention. The key to this shift in the problem definition came from the group's willingness and interest to understand South African teenagers within their cultural context, including the consideration of language, symbols, rituals, and other shared meanings that populate their world.

Finally, at its core, design thinking is about ongoing decision-making: "[...] you never really get to stop questioning the assumptions that you're making and that are underlying what it is that you're creating – those fundamental premises. Things change, you change, the market changes. And you have to have as part of your practice an ongoing process of revisiting those fundamental assumptions" (Altringer, 2017).

Artificial Intelligence (AI)

"Artificial Intelligence (AI), defined as a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 15), can be characterized by intentionality, intelligence, and adaptation (for a detailed report see West & Allen, 2018). Today, AI increasingly influences our experiences and interactions in almost any walk of life. Intelligent algorithms can be found in areas ranging from autonomous vehicles, to portfolio choices, to healthcare, to virtual home assistants such as Google's Alexa, to social media bots that shape the selection of news we see, and criminal sentencing (Dressel & Farid, 2018). Given the ubiquity of AI in our daily lives, it is to some extent surprising that we haven't seen much of a connection with behavioral economics, albeit there is indeed a lot of value in that connection. In a way, we can think of Behavioral Economics 1.0 as testing the effectiveness of interventions to either increase a desired behavior or to reduce an undesired behavior, and then universally implement the best option. Behavioral Economics 1+X.0, on the other hand, uses AI algorithms with high prediction accuracy to suggest case-by-case, personalized interventions for individuals in situations where the best option may vary across people (for an interesting read in the context of bail decisions see Kleinberg et al., 2017).

Another possible connection lies in the use of AI in the search for new 'behavioral'-type variables that affect decisions (Camerer, 2019; see also Risdon, 2017, on the "Behavior Genome"). Particularly, the comparison of human and AI decisions, and the identification of commonalities and discrepancies, can be informative and help us uncover new insights about less obvious, underlying psychological mechanisms and variables that affect human behavior. Thus, AI could help make the explorative phase of a behavioral economics project more effective, allowing for the development of innovative new ideas.

Finally, given the ubiquity of AI, behavioral economics can play an important role in helping to study human-AI interactions, to understand how it is perceived and used by individuals and institutions to reduce, exacerbate, or exploit human limits, and to learn how to use AI most productively (Camerer, 2019). This agenda ranges from the study of "algorithm aversion", its drivers (e.g. perceptions of uniqueness neglect), and ways to overcome it (e.g. introducing feelings of control or personalization; Dietvorst, Simmons, & Massey, 2016; Longoni, Bonezzi, & Morewedge, 2019), to understanding perceived and actual AI bias, for example, because of imperfections in training data or how data were collected (Kleinberg, Mullainathan, & Raghavan, 2017), fairness (e.g. through deliberately blinding a detention decision algorithm in terms of race and ethnicity; Kleinberg, Mullainathan, & Raghavan, 2017), and morality (see individuals' preferences and perceptions of conflicts of interest-decisions made by autonomous vehicles; Awad et al., 2018), the latter of which brings us back, full circle, to the issue of ethics and responsibility.

Conclusion

It is indeed an exciting time to be a behavioral economist. We are at a crossroads at which we can decide to either be satisfied with what we have achieved or challenge our discipline to examine more actively the impact and welfare effects of our interventions and broaden our influence and reach beyond the last mile to win a (more influential) seat at the policy and/or c-suit table. If we decide to go for the latter, we will need to consider the ethics of our actions and embrace integrative thinking, in order to demonstrate both accountability and responsibility.

The Author

Nina Mažar is Professor of Marketing and Co-Director of the Susilo Institute for Ethics in the Global Economy at Questrom School of Business at Boston University. She is also the co-founder of BEworks.

Nina, who is currently the president of the Society for Judgment and Decision Making, was named one of "The 40 Most Outstanding B-School Profs Under 40 In The World". With her focus on behavioral economics she investigates the implication of human biases on welfare, development, and policy. Her research topics range from donations to financial decision making to morality. She uses various approaches from large scale field experiments, to controlled, incentive compatible behavioral lab studies to neuroscience methods. Popular accounts of her work have appeared among others on NPR, BBC, in the New York Times, Wired, Harvard Business Review, and various NYTimes Bestsellers as well as in the documentary feature film "The Honest Truth about Dishonesty."

Nina serves as behavioral economics advisor on boards of various government and non-for-profit and profit organizations (e.g., Irrational Labs in San Francisco, CA). She was previously a post-doctoral associate and lecturer in marketing at MIT Sloan School of Management and the MIT Media Lab (Dan Ariely's eRationality Group). She holds a PhD in Marketing from the University of Mainz in Germany. Her website is www.ninamazar.com.

References

Baeck, A., & Gremett, P. (2012). Design Thinking: Expanding UX Methods Beyond Designers. In H. Degen & X. Yuan, UX Best Practices: How

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I (2018). The moral machine experiment, *Nature, 563,* 59-64.

Altringer, B. (2017). AI and ethical design. *Open Transcripts.* Retrieved from http://opentranscripts.org/transcript/ai-ethical-design/.

Anderson, A., & Robinson, D. T. (2019). Who feels the nudge? Knowledge, self-awareness and retirement savings decisions. *Swedish House of Finance Research Paper No. 17-15.* Retrieved from http://dx.doi.org/10.2139/ssrn.3021962.

Baeck, A., & Gremett, P. (2012). Design thinking: Expanding UX methods beyond designers. In H. Degen & X. Yuan (Eds.), *UX best practices: How to achieve more impact with user experience* (pp. 229-250). New York, NY: Osborne.

Baldwin, R. (2014) From regulation to behaviour change: Giving nudge the third degree. *The Modern Law Review, 77*(6), 831-857.

Brown, T. (2008). Design Thinking. *Harvard Business Review, 86*(6), 84-92.

Brown, T. (2009). *Change by design: How design thinking transforms organizations and inspires innovation.* New York, NY: Harper Business.

Behavioral Economics Team of the Australian Government (2019). *Credit when it's due: Timely reminders help consumers reduce their credit card debt.* Retrieved from https://behaviouraleconomics.pmc.gov.au/sites/default/files/projects/credit-cards-final-report-accessible.pdf.

Bennhold, K. (2013, December 7). Britain's Ministry of Nudges. *The New York Times.* Retrieved from https://www.nytimes.com/2013/12/08/business/international/britains-ministry-of-nudges.html.

Blumenthal-Barby, J. S. (2012) Between reason and coercion: Ethically permissible influence in health care and health policy contexts. *Kennedy Institute of Ethics Journal, 22*(4), 345-366.

Brockmeyer, A., Hernandez, H., Antonio, M., Kettle, S., Smith, S. D. (2016). *Casting the tax net wider: Experimental evidence from Costa Rica.* Policy Research Working Paper No. WPS 7850. Washington, D.C.: World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/660661475679714942/Casting-the-tax-net-wider-experimental-evidence-from-Costa-Rica.

Butera, L., Metcalfe, R., Morrison, W., & Taubinsky, D. (2019). *The deadweight loss of social recognition.* NBER Working Paper No. 25637. Retrieved from https://www.nber.org/papers/w25637.

Camerer, C. F. (2019). Artificial intelligence and behavioral economics. In A. Agarwal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 587-608). University of Chicago Press.

Chilazi, S., Asundi, A., & Bohnet, I. (2019, March 12). Venture capitalists are using the wrong tools to improve gender diversity. *Behavioral Scientist.* Retrieved from https://behavioralscientist.org/venture-capitalists-are-using-the-wrong-tools-to-improve-gender-diversity/.

Damgaard, M., & Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics, 157,* 15-26.

Datta, S., Burns, J., Maughan-Brown, B., Darling, M., & Eyal, K. (2015). Risking it all for love? Reseatting beliefs about HIV risk among low-income South African teens. *Journal of Economic Behavior & Organizations, 118,* 184-198.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155-1170.

Dressel, J., & Farid, H. (2018). *The accuracy, fairness, and limits of predicting recidivism.* Open Dartmouth: Faculty Open Access Articles. https://digitalcommons.dartmouth.edu/facoa/1342.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics, 75*(4), 643-669.

Frey, E., & Rogers, T. (2014). Persistence: How treatment effects persist after interventions stop. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 172–179.

Ghesla, C., Grieder, M., & Schmitz, J. (2019), Nudge for good? Choice defaults and spillover effects. *Frontiers in Psychology, 10,* 178.

Gneezy, U. (2019). Incentives and behavior change. In A. Samson (Ed.), *The behavioral economics guide 2018 (with an introduction by Uri Gneezy)* (pp. NN-NN). Retrieved from http://www.behavioraleconomics.com.

Goel, V. (2014, June 29). Facebook tinkers with users' emotions in news feed experiment, stirring outcry. *The New York Times*. Retrieved from https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html.

Hausman, D. M. & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy, 18*(1), 123-136.

House, J., Robitaille, N., & Mazar, N. (2018). *Effectiveness of repeated implementation intention-interventions on organizations' likelihood to file their overdue taxes.* Rotman School of Management Working Paper No. 2997993. Retrieved from https://ssrn.com/abstract=2997993.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri in my hand, who's the fairest in the land? On the interpretations, illustrations and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). *Human decisions and machine predictions*. NBER Working Paper No. 23180. Retrieved from https://www.nber.org/papers/w23180.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *Proceedings of Innovations in Theoretical Computer Science (ITCS)*. Retrieved from https://scholar.harvard.edu/files/sendhil/files/1609.05807.pdf.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788-8790.

Loewenstein, G., & Chater, N. (2017). Putting nudges in perspective. *Behavioural Public Policy*, *1*(1), 26-53.

Lombard, D. N., Lombard, T. N., & Winett, R. A. (1995). Walking to meet health guidelines: The effect of prompting frequency and prompt structure. *Health Psychology, 14*(2), 164–170.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research.* https://doi.org/10.1093/jcr/ucz013.

Martin, R. L. (2007, June). How successful leaders think. *Harvard Business Review*. Retrieved from https://hbr.org/2007/06/how-successful-leaders-think.

Martin, R. L. (2009). *The opposable mind: How successful leaders win through integrative thinking*. Boston, MA: Harvard Business School Press.

Mazar, N., Mochon, D., & Ariely, D. (2018). If you are going to pay within the next 24 hours, press 1: Automatic planning prompt reduces credit card delinquency, *Journal of Consumer Psychology*, *28*(3), 466-476.

Metcalfe, R. (2018). Behavioral economics: Under the microscope. In A. Samson (Ed.), *The behavioral economics guide 2018 (with an introduction by Robert Cialdini)* (pp. 2-22). Retrieved from http://www.behavioraleconomics.com.

Meyer, M. N., Heck, P. R., Holtzman, G. S., Anderson, S. M., Cai, W., Watts, D. J., & Chabris, C. F. (2019). Objecting to experiments that compare two unobjectionable policies or treatments. *Proceedings of the National Academy of Sciences*. Retrieved from https://www.pnas.org/content/early/2019/05/08/1820701116.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81,* 33-43.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C., Korris, J., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 129-140*.*

Quispe-Torreblanca, E., Stewart, N., Gathergood, J., & Loewenstein, G. (2019). The red, the black, and the plastic: Paying down credit card debt for hotels not sofas. *Management Science*, in press.

Reiss, J. (2013). *Philosophy of economics: A contemporary introduction*. New York, NY: Routledge.

Risdon, C. (2017, October 25). Scaling nudges with machine learning. *Behavioral Scientist.* Retrieved from https://behavioralscientist.org/scaling-nudges-machine-learning/.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26,* 521-562.

Schmidt, R. (2019, April 23). Broadening the nature of behavioral design. *Behavioral Scientist.* Retrieved from https://behavioralscientist.org/broadening-the-nature-of-behavioral-design/.

Banuri, S., Dercon, S., & Gauri, V. (2018). Biased policy professionals. *World Bank Economic Review.* Retrieved from https://ora.ox.ac.uk/objects/uuid:36e916cf-24bf-4131-bd8c-c1b-4919f72b4.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*(38), 15197-15200.

Soman, D. (2015): *The last mile: Creating social and economic value from behavioral insights*. Toronto, Canada: Rotman-UTP Publishing.

Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy, 57,* 160-171.

Thinkers 50. (2017). *Roger Martin.* Retrieved from https://thinkers50.com/biographies/roger-martin/.

Thunström, L. (2019, February 19). Good for some, bad for others: The welfare effects of nudges. *Behavioraleconomics.com*. Retrieved from https://www.behavioraleconomics.com/good-for-some-bad-for-others-the-welfare-effects-of-nudges/.

West, D. M., & Allen, J. R. (2018, April 24). How artificial intelligence is changing the world. *Brookings.* Retrieved from https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/.

Wilkinson, T. M. (2013a). Nudging and manipulation. *Political Studies, 61*(2), 341-355.

Wilkinson, T. M. (2013b). Thinking harder about nudges. *Journal of Medical Ethics, 39*(8), 486.