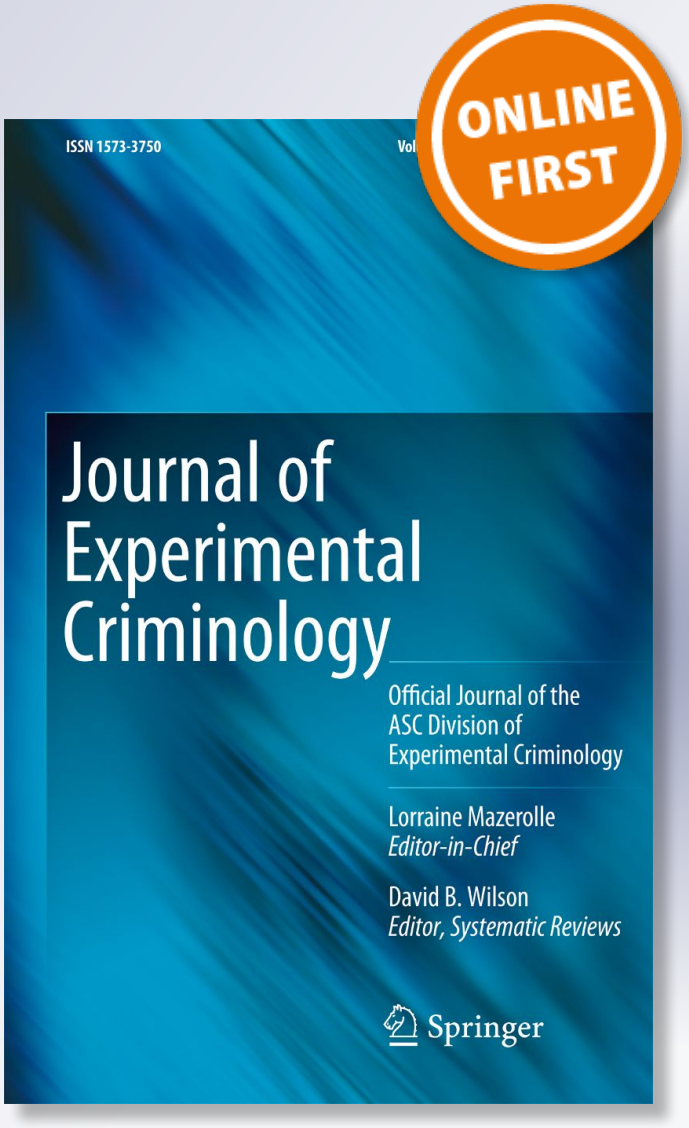
Experimental criminology: looking back and forward on the 20th anniversary of the Academy of Experimental Criminology

David P. Farrington, Friedrich Lösel, Anthony A. Braga, Lorraine Mazerolle, Adrian Raine, Lawrence W. Sherman & David Weisburd

Journal of Experimental Criminology

ISSN 1573-3750

J Exp Criminol
DOI 10.1007/s11292-019-09384-z



Springer

Journal of Experimental Criminology
https://doi.org/10.1007/s11292-019-09384-z

Experimental criminology: looking back and forward on the 20th anniversary of the Academy of Experimental Criminology

David P. Farrington[1] · Friedrich Lösel[2,3] · Anthony A. Braga[4] · Lorraine Mazerolle[5] · Adrian Raine[6] · Lawrence W. Sherman[2,7] · David Weisburd[8,9]

Published online: 04 December 2019

Abstract

Objectives This article summarizes key points made in a panel at the American Society of Criminology (ASC) meeting in Atlanta in November 2018, entitled "20th Anniversary of the Academy of Experimental Criminology (AEC): Looking Back and Forward," organized by Friedrich Lösel as the AEC president.
Method Seven (current and former) presidents of AEC contribute short papers about the past and future of experimental criminology, focusing on different and emerging areas of criminological experimentation, as well as identifying topics that require more attention in future, including field experiments and experimental neurocriminology.
Results This article informs readers about the history of AEC, its links with the *Journal of Experimental Criminology*, current issues, and potential future developments in experimental criminology. It also briefly deals with arguments that question whether experiments are the "gold standard," which were addressed by Daniel Nagin and Robert Sampson in another ASC session at Atlanta. Experimental panel members did not view randomized controlled trials (RCTs) as a ritualized and general "gold standard" for criminological research, because many important topics cannot be investigated in this type of design.
Conclusions This article is not intended to be a missionary statement for RCTs, but it does argue that experiments should be used whenever feasible, because they are most robust in ensuring internal validity as the basis for external validity and for generalizations that are necessary for effective practice and policy making.

Keywords Experimentation · Quasi-experimentation · Control groups · Neurocriminology · Policing

✉ David P. Farrington
dpf1@cam.ac.uk

Extended author information available on the last page of the article

A brief history of the AEC (by Friedrich Lösel, Lawrence W. Sherman, and David P. Farrington)

The Academy of Experimental Criminology (AEC) was founded in 1998 to advance the use of randomized trials and systematic reviews in crime and justice research. The AEC also recognizes criminologists who have led and/or promoted randomized field experiments to advance evidence-informed public policy and practice. As a learned society, the AEC aims to focus attention on experimenters and create a platform to support and encourage experimental approaches to inform crime and justice policy and practice. The AEC is thus a dedicated forum devoted to advancing experimentation by supporting newcomers to the field, creating synergies among those already conducting experiments, helping to increase the uptake of experimental findings by policy makers, and generally facilitating discourse around the design, implementation, management, and outcomes from experimental research (see Weisburd et al. 2007).

The foundation of the AEC had various motives and sources. On the one hand, it was triggered by the experience of a deficit of experiments in criminology. Although there were encouraging developments of experimental research in countries like Great Britain in the 1960s and 1970s, influences on Government funding had reduced this promising strand of research (Farrington 2003b; Farrington and Welsh 2006). These scientific and/or political changes contradicted the fact that, in important political areas, such as offender rehabilitation, a lack of sound evidence had been a complaint since the 1970s (Lipton et al. 1975; Sechrest et al. 1979). Even decades later, reviews and registries on criminological interventions often state that there are too few replicated and long-term experimental evaluations (Lösel 2018; Mihalic and Elliott 2015).

On the other hand, there were also experiences and developments that supported experimental criminology in the USA in the 1960s through the 1990s. The first modern criminal justice policy tested with a randomized trial design in the US was the comparison of release on recognizance with money bail in Manhattan, the first project of the Vera Institute of Justice (Ares et al. 1963). This was followed by the early efforts of the Police Foundation to conduct quasi-experiments, such as the 15-beat Kansas City Preventive Patrol Experiment (Kelling et al. 1974). A National Academy of Sciences/National Research Council (1977) report then recommended that the Justice Department should use more large-sample randomized controlled field experiments to develop knowledge about deterrence, which led to many such experiments being funded and carried out. Among them was the Minneapolis Domestic Violence Experiment (Sherman and Berk 1984), which prompted a debate about randomized experiments with individual offenders and was followed by six replications (Sherman 1992). At several ASC sessions in the early 1990s, Lawrence Sherman, Peter Greenwood, Joan Petersilia, Joan McCord, David Weisburd, David Farrington, and others discussed the advantages and special challenges of RCTs.

The Maryland report on evidence-based crime prevention was another milestone (Sherman et al. 1997, 2002). It demonstrated that in many important areas of crime prevention there were some, but too few, replicated evaluations that used RCTs or sound quasi-experimental methods. Stimulated by this report, in 1997, the Jerry Lee Foundation decided to provide substantial investment in experimental criminology at the University of Maryland, including the appointment of David Farrington as a Research Professor and post-doctoral fellows David Wilson and Spencer De Li to work with Professors Doris MacKenzie and Denise Gottfredson.

The Jerry Lee Foundation was established in 1996 by Jerry Lee and his radio station business partner, David Kurtz, primarily to promote better education and public safety through greater use of research. During the Foundation's early scanning for opportunities to make a difference, Jerry Lee read about the Maryland Report in the New York Times. He immediately contacted the report's first author, Lawrence Sherman, and together they developed a two-decade collaboration to promote experimental criminology and evidence-based practice.

In 1998, Lawrence Sherman proposed to the growing group of experimentalists that they should organize an independent, international academy that would recognize achievements by electing fellows. As they agreed to this proposal, AEC was established, and in 1999, Lawrence Sherman was elected the first president of the AEC at the ASC meeting in Toronto. This early development shows what is often important in academia and policy fields: An intellectually sound mission needs a small group of initiators and also a sponsor like Jerry Lee, who was convinced that this was an important issue for research, policy, and practice. As of 2019, the AEC has elected 72 fellows and 13 honorary fellows, including Jerry Lee (https://expcrim.org/aec-fellows/). In close contact with AEC, the Campbell Crime and Justice Group was founded in Philadelphia in the spring of 2000 (Farrington and Petrosino 2001) and since then has commissioned and published systematic reviews on measures of crime prevention (https://campbellcollaboration.org/component/tags/tag/crime-and-justice.html).

In 2001, David Farrington became the second president of AEC, followed by Joan McCord, David Weisburd, Doris MacKenzie, Lorraine Mazerolle, Anthony Braga, Adrian Raine, Peter Greenwood, Friedrich Lösel, and Heather Strang. After Joan McCord's death in 2004, the AEC created an annual distinguished lecture in her name which is now published each year in the *Journal of Experimental Criminology*. In 2005, David Weisburd founded this journal, as the official journal of the Academy of Experimental Criminology.

Since 1999, the AEC has held regular sessions at ASC conferences. At the meeting in 2007, Friedrich Lösel proposed to include younger scholars in the work of the AEC and offered to sponsor a young scholars' award for 5 years, from his Stockholm Prize money. This idea was supported by the AEC fellows, and later, the AEC took a formal move to encourage young scholars to participate in the field, via the structure of a Division of Experimental Criminology (DEC) of the ASC.

AEC presidents Sherman, Farrington, Weisburd, and MacKenzie agreed that a petition should be filed with the ASC Board. Doris MacKenzie circulated the petition as the founding Chair of the ASC Division, which led to the establishment of the DEC by the ASC Board. At its first meeting, the DEC elected Lawrence Sherman to serve the first 2-year term as Chair. The DEC created its own awards, including the Jerry Lee Lifetime Achievement Award in Experimental Criminology and the Award for an Outstanding Experimental Field Trial. The new DEC agreed to work in close cooperation with the AEC, building on its record, taking over duties, and adopting the *Journal of Experimental Criminology* as the official journal of the DEC and distributed to all members of the DEC as part of their divisional dues.

Because of the tireless work of its editors in chief, David Weisburd and Lorraine Mazerolle, and also David Wilson as the editor for systematic reviews, the journal became quickly visible and in 2018 had an impact factor of 3.9, which is high in criminology and criminal justice. Of course, bibliometric data may change and should not be interpreted too simply as a proof of achievement, but within the context of the

other above-mentioned facts, it shows that the AEC has had a big impact on criminology and good reason to celebrate its 20th anniversary. However, looking back is not enough, and therefore, the presenters at the 2018 ASC session were also asked to address challenges and future developments in their field of experimental research. These are presented in the following parts of this article, starting with a review of the last 15 years of contributions from the *Journal of Experimental Criminology*.

Growing the impact of experimentation in crime and justice: an analysis of the *Journal of Experimental Criminology* publications (by Lorraine Mazerolle)

The establishment of the *Journal of Experimental Criminology* (JEC) in 2005 represented an important milestone for the AEC. Published quarterly by Springer, the JEC is a peer-reviewed journal that publishes high-quality systematic reviews and experimental and quasi-experimental research as a way to develop and inform evidence-based crime and justice policy. The journal is also committed to the advancement of the science of systematic reviews and experimental methods in criminology and criminal justice, publishing empirical papers on experimental and quasi-experimental studies, systematic reviews on substantive criminal justice problems, and methodological papers on experimentation and systematic reviews.

Springer has now (up to the end of 2018) published 57 issues of the JEC since its inception in 2005. With an average of 6.8 articles and contributions per issue over 14 years, the JEC is now a well-established journal in the crime and justice landscape. This section of our article offers some analysis of the 386 original papers, research notes, review papers, short reports, introductions, and other types of communications published in the JEC over time, focusing on some of the ways that the JEC has contributed to growing the impact and influence of experimentation in crime and justice.

One way to assess trends in the influence of a journal is to examine the impact factor. The impact factor is a measure of the number of times a contribution (any type of contribution) in the journal is cited during the succeeding 2 years and the succeeding 5 years. The JEC is listed in the Criminology and Penology category of 61 journals. The 2-year 2017 impact factor for JEC is now 3.912, making the journal the number 3 ranked journal in the Criminology and Penology category. As Fig. 1 shows, 2010 was the entry year of JEC with an impact factor, starting with a healthy impact factor of 2.12. Since that time, the journal's impact factor was initially quite stable for several years and then escalated from 2014, reflecting the strength of the articles published in 2012 and 2013 under the Editorship of the founder, David Weisburd.

To further understand the drivers of the increase in the impact factor over time, we analyzed the number and sum of downloads for the top ten papers for 2015, 2016, and 2017. As Fig. 2 shows, the sum of the top ten publications each year (for 2015, 2016, and 2017) has increased from 10,224 in 2015 to 30,087 in 2017. In 2015, the number one paper was downloaded 1747 times, but in 2017, the number one paper was downloaded 7122 times.

Another way to understand what is driving the growing influence of the JEC is to assess what types of topics are generating the growth in downloads. To begin this analysis, we examined the population of all 386 journal contributions from 2005 to

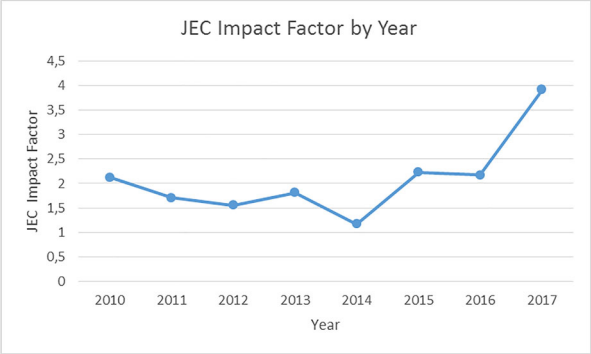


Fig. 1 JEC impact factor by year

2018. Of these, we included 314 research papers (including short reports) and excluded 72 entries that were introductions, errata, editorial notes, corrections, letters, book reviews, and other types of non-research communications. We then analyzed the 314 research papers and generated six categories using Leximancer software. Leximancer is a text analytics software tool developed at the University of Queensland that can be used to analyze the content of textual documents. Conceptual "bubbles" are created from the analysis that represent the conceptual structure of a corpus of text (see https://info.leximancer.com).

The six categories were as follows: research design and methodologies (including concepts related to advancing the science of experimentation including experimental, control group, significance testing, and innovation in evaluation methods), policing interventions (including concepts related to officers, public/community engagement, procedural justice, police legitimacy, and hotspots), substance abuse (including concepts related to substance use, treatment, prevention, behavior), criminal justice (excluding police but including concepts related to offenders, court, justice, recidivism, treatment/programs in justice settings), crime prevention and community interventions (including concepts related to family, schools, community, and non-offenders such as high-risk juveniles who have not yet committed a crime), and systematic reviews and systematic review updates. Fig. 3, below graphs, shows the total number of downloads from 2005 to 2018 across the six categories for the 314 papers included in the analysis.

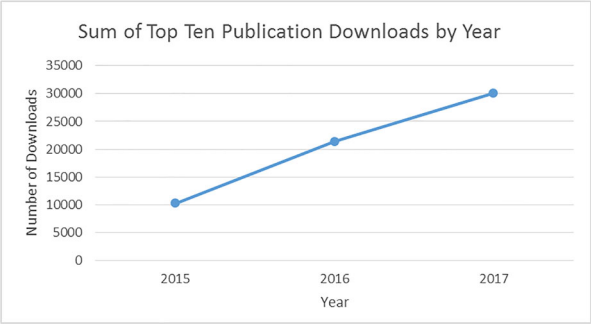


Fig. 2 Sum of top ten publication downloads by year

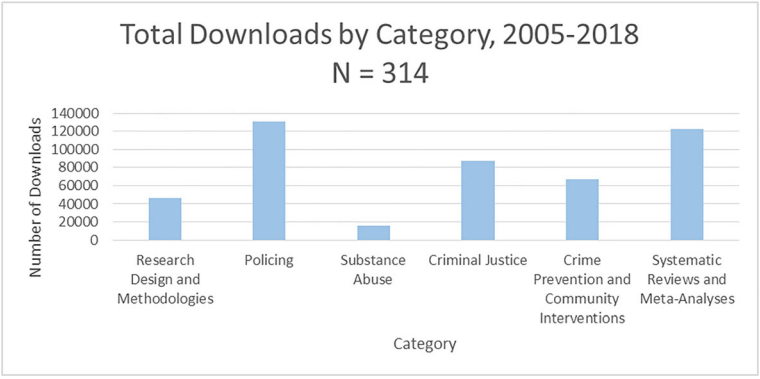


Fig. 3 Total downloads by category, 2005–2018 ($N = 314$)

As Fig. 3 shows, policing and systematic review papers are driving the number of downloaded JEC papers. The 55 policing papers published in the JEC from 2005 to 2018 have been downloaded 130,660 times (average of 2375.6 downloads per paper); the 37 systematic review papers (and updates) were downloaded 121,999 times (average of 3297.3 downloads per paper); the 60 criminal justice papers were downloaded 86,815 times (average of 1446.9 downloads per paper); the 50 crime prevention/community-based interventions were downloaded 66,715 times (average of 1334.3 downloads per paper); the 89 research design and methods papers were downloaded 46,106 times (average of 518.0 downloads per paper); and the 23 substance abuse papers were downloaded 16,248 times (average of 706.4 downloads per paper).

Overall, these results reveal both the strengths and vulnerabilities of the JEC. On the one hand, it is clear that the demand for policing and systematic review papers is at the core of what has driven the success of the journal. This result is likely a reflection on the editorship of the journal (both editors in chief, David Weisburd and Lorraine Mazerolle, are policing scholars and the editor for systematic reviews is David Wilson, one of the world's leading systematic review scholars). On the other hand, it offers some growing opportunities for the journal, focusing on attracting papers to JEC from field experiments testing why people offend (see David Farrington's section below) to experiments that bridge into other domains such as public health, behavioral economics, and neurocriminology (see Adrian Raine's section below).

Realistic field experiments on stealing and dishonesty (by David P. Farrington)

Criminology should be an experimental science, with statements backed up by evidence, quantitative data, systematic observation, valid and reliable empirical measures of underlying theoretical constructs, controlled experiments, falsifiable theories, testing causal hypotheses, and replication of empirical results. There have been many randomized field experiments in criminology designed to study policing, early prevention, corrections, courts, or community treatment (see Farrington 1983, 2013), but very few designed to test hypotheses about explanations of offending. I would like to encourage criminologists to conduct realistic field experiments with offending as the dependent

variable. I carried out a program of field experiments on stealing between 1975 and 1978 (Farrington 1979; Farrington and Kidd 1977; Farrington and Knight 1979, 1980a, b), but it seems that very few realistic field experiments on offending have been conducted by criminologists since then (see, e.g., Gabor and Barker 1989; Gabor et al. 1986).

Field experiments rather than laboratory experiments are needed in order to maximize external validity as well as internal validity (Farrington 1980). Actual offending behavior needs to be studied rather than verbal statements about the likelihood of offending, because words and deeds may be different. There is a long history of research, often using hypothetical scenarios, in which people have been asked whether they think they would commit crimes. However, one of the first investigations of the validity of such verbal statements about offending in relation to real-life offending was conducted by Farrington et al. (1980). They asked youths about their stealing in hypothetical situations and also gave them an opportunity to steal (in a coin-sorting task). Surprisingly, they found that the youths who actually stole were not significantly more likely to say that they would steal in a hypothetical situation. However, this comparison was based on small numbers (25 youths).

I concluded that realistic field experiments on stealing and dishonesty were needed. In the first experiment by Farrington and Kidd (1977), the experimenter walked past a member of the public in the street, pretended to pick up a coin, and then ran after the person, offering the coin and asking whether he or she had dropped it. The person then had the opportunity to claim the coin dishonestly. We found that people were more likely to steal from a female than from a male experimenter. We hypothesized that this was because the subjective cost of dishonesty was greater with the female experimenter, because members of the public viewed her more favorably. For example, with the female experimenter, one participant said, "I'd take it if I thought it was mine, but why don't you have it?" In contrast, with the male experimenter, participants said "I often drop money" or "I must have a hole in my pocket."

In later experiments, Farrington and Knight (1979, 1980b) left stamped, addressed, apparently lost, unsealed letters on the street, each containing a handwritten note and also (except for control conditions) a sum of money. The experimenter, who was blind to the condition of each letter, observed the personal characteristics and behavior of each person who picked up the letter. Each person could honestly mail the letter and money to the intended recipient or could dishonestly steal the money. We found that behavior after picking up the letter predicted stealing. Almost all of the participants were observed to take out the note and read it. Those who then walked along holding the letter were likely to return it, whereas those who put the letter in a pocket or handbag were likely to steal it. This suggested that the decision to steal was made immediately. The prevalence of stealing varied remarkably, from about 20 to 80% in different conditions (depending on the age, gender, and affluence of the victim). This suggested that, depending on the experimental conditions, almost everyone would steal or almost no one would steal.

Our experiments were inspired by subjective expected utility theories (Farrington 1979; Farrington and Knight 1980a). We found that stealing increased with the amount of money that could be stolen, decreased when the apparent victim was an impoverished old lady (high cost) compared with an affluent young man (low cost), and decreased when the probability of detection was greater (with a postal order compared with cash). We also found that younger people were more likely to steal

than older ones, although in most cases (except when there was a large amount of money), there were few gender differences in the likelihood of stealing.

While few realistic field experiments on stealing or dishonesty seem to have been carried out by criminologists in recent years, several have been conducted by behavioral economists. Unfortunately, few criminologists read economic journals such as *Experimental Economics* or the *Journal of Behavioral and Experimental Economics*. Hugo Gomes, Ivy Defoe, and I are reviewing recent experiments on stealing and dishonesty in the hope of bringing them to the attention of criminologists and encouraging them to conduct realistic field experiments. Here are two examples.

Alem et al. (2018) carried out an experiment in Tanzania in 2013–2014 in which individuals received an apparently misdirected money transfer and then immediately afterwards received a text message asking them to return the money. There were three experimental conditions: the message (1) was neutral (the control condition), (2) allowed them to keep 25% of the money as a reward (the kindness condition), or (3) informed them that the money was intended for an orphanage (the guilt condition). They found that 24% returned the money in the control condition, compared with 43% in the kindness condition and 37% in the guilt condition.

One year before this experiment, the same individuals had been surveyed and asked what they would do in a similar hypothetical situation of receiving misdirected money. The researchers found that 35% of those who said that they would return the money actually returned it, compared with 30% of those who said that they would keep the money who actually returned it. Therefore, in agreement with Farrington et al. (1980), Alem et al. (2018) concluded that there was very little relationship between words and deeds in the study of stealing.

Franzen and Pointner (2013) carried out a laboratory experiment with students in Cologne (Germany) and Bern (Switzerland) in 2009–2012. The students received vouchers for 10 Euros (in Germany) or 20 Francs (in Switzerland) and were told that they could divide the money between themselves and another person in any way that they wanted. There was an experimental intervention in Switzerland, based on whether the students did or did not see a photograph of the (alleged) other person, but this did not influence the amount of money given to the other person. Some weeks or months after the experiment, students in both countries received an apparently misdirected letter containing 10 Euros or 20 Francs and had the opportunity to keep the money dishonestly or return it honestly. In both countries, the amount of money given to the other person in the laboratory predicted the probability of returning the money honestly. The researchers concluded that behavior in the laboratory had external validity in relation to real behavior in the field.

I believe that criminologists should attempt to carry out more experiments to investigate theories of offending, using a realistic measure of offending as the dependent variable. The most feasible dependent variables are probably stealing and vandalism; it is hard to imagine conducting an experiment with real violence as the dependent variable, although verbal aggression might possibly be studied. Experiments on traffic offenses such as worn tires (Buikhuisen 1974) and turning against a red light (Sigelman and Sigelman 1976) are also feasible.

There is now much more interest by criminologists in conducting experiments, as evidenced by the foundation of the Academy of Experimental Criminology and by the establishment of the *Journal of Experimental Criminology*. However, there is surely a

need for more realistic field experiments designed to test theories of offending, so that we can finally get serious about the idea that criminology is an experimental science.

Experimental neurocriminology (by Adrian Raine)

A key aim of this section is to highlight new developments in the field of experimental criminology. One such development is the fledgling area of *experimental neurocriminology*. Neurocriminology is the application of neuroscience principles and techniques to understand the causes and treatment of crime—incorporating knowledge on human physiological processes for understanding a wide range of antisocial behaviors throughout the life course. This research is, however, predominantly correlational. Experimental neurocriminology, in contrast, highlights experimental and quasi-experimental research aimed at understanding how causal physiological processes shape crime and violence and how social and psychological processes impact biological risk factors for crime—a sub-area known as social neurocriminology (Choy et al. 2015).

One example of experimental neurocriminology comes from a recent randomized controlled trial (RCT) aimed at reducing criminal intent. Choy et al. (2018) randomized adults to receive either electrical stimulation to the prefrontal cortex using transcranial direct current stimulation, or sham stimulation. Participants were then given social vignettes involving a confrontational aggressive encounter and questioned on their likelihood of perpetrating a violent crime. Prefrontal upregulation of the prefrontal cortex cuts in half the intention to commit physical and sexual violence.

Experimental neurocriminology embraces health perspectives on antisocial and violent behavior. One way of manipulating brain structure and function is by supplementing diets with omega-3, a long-chain fatty acid that is critical for brain structure and function. One RCT in Mauritius randomized children into either receiving omega-3 in a fruit juice drink each day for 6 months or receiving the same juice drink without omega-3 supplementation (Raine et al. 2015). Omega-3 supplementation resulted in reduced antisocial and aggressive behavior 6 months after the treatment had ended. Similar long-term effects of omega-3 supplementation have been found in RCTs of children in Philadelphia (Raine et al. 2016) and Singapore (Raine et al. 2018). One meta-analysis has confirmed the efficacy of omega-3 in reducing aggression (Gajos and Beaver 2016). While studies on incarcerated offenders are sparse, the two studies conducted to date provide support for the efficacy of omega-3 (Gesch et al. 2002; Zaalberg et al. 2010).

Experimental neurocriminology also seeks to understand which biological risk variables *moderate* treatment outcomes. For example, one experimental study provided better nutrition, more physical exercise, more daytime sleep, and cognitive stimulation to children for 2 years starting at age 3 (Raine et al. 2003). This early enrichment enhanced brain functioning at age 11 as measured by the EEG, maturing the brains of those in the experimental group by 1.1 years (Raine et al. 2001). This brain-enhanced enriched group went on to show a 36% reduction in criminal offending at age 23. While significant reductions in conduct disorder at age 17 were also found, this main effect was moderated by nutritional status before the start of the intervention. Specifically, malnourished children who experienced the early enrichment showed a 53% reduction

in age 17 conduct disorder compared with malnourished control children. In contrast, the enrichment did not reduce conduct disorder in children who were not malnourished at baseline. The moderating effect of nutritional status on the enrichment–conduct disorder relationship suggests, but does not prove, that the nutritional enhancement, a biological risk factor for adolescent antisocial behavior (Liu et al. 2004), was a key factor in reducing teenage antisocial behavior.

Yet another field of enquiry embraced by experimental neurocriminology concerns how interventions alter biological risk factors for crime. One review of 11 treatment studies that incorporated biological risk factors showed that treatment programs for antisocial behavior tended to normalize neurobiological risk factors such as low cortisol and that to some extent changes in these biological risk factors were associated with behavioral improvement (Cornet et al. 2015). The value of such studies lies in the identification of mechanisms of action in treatments which could help to develop improved interventions and ultimately enhanced treatment efficacy for antisocial and violent behavior. This social neurocriminology perspective helps to sharpen mainstream criminological theories by clarifying *how* social interventions can reduce crime (Choy et al. 2015).

Experimental neurocriminology also seeks to understand how biological interventions alter psychosocial processes in order to bring about change. In the RCT of Choy et al. (2018), which documented how prefrontal upregulation reduced criminal intent, it was found that this neurobiological intervention enhanced participants' moral sense. More importantly however, enhanced morality partially mediated the prefrontal stimulation–reduced criminal intent effect, accounting for 31% of the treatment effect. This exemplifies how a biological manipulation (tDCS) can influence social judgment (morality) in a way to reduce criminal intent.

Yet another question addressed by experimental neurocriminology concerns what biological characteristics of the offender predict success or failure in treatment studies. One review of 10 such studies (Cornet et al. 2014) found that offenders with low physiological arousal—a well-replicated risk factor for crime—were *less* likely to improve after treatment. In contrast, those with high levels of arousal were *more* likely to benefit. This knowledge offers the potential in practice to assign individuals to treatment programs based on their biological profile and to develop alternative treatments for those with low arousal.

Experimental neurocriminology examines how psychological interventions can alter neurobiological risk factors for crime. One RCT found that, while the experience of incarceration impairs neurocognitive functioning, a cognitive behavioral (CBT)/mindfulness intervention buffers against this neurocognitive decline (Umbach et al. 2018). The study's importance lies in showing not only that imprisonment further strengthens a well-documented neurocognitive risk factor for crime but also that psychological interventions, which are known from other experimental studies to enhance brain functioning (mindfulness), can protect against such a decline, with the potential to reduce future recidivism.

While the fledgling field of experimental neurocriminology has not yet taken flight, it does represent a notable future development. In 2019, the ASC meeting will witness the first panel devoted to this area, and in 2020, a special edition on experimental neurocriminology is planned for this journal. Its future promise lies in providing the discipline of criminology with new insights into the mechanisms of action by which

psychosocial interventions bring about behavioral change, as well as a new vista on benign neurobiological interventions to reduce antisocial and criminal behavior.

The specific conditions in control groups: an issue that needs more attention in experiments (by Friedrich Lösel)

As I held a chair of psychology over many years, the mission of the Academy of Experimental Criminology is a matter of course for me. Since the time of Wilhelm Wundt, the founder of academic psychology in the nineteenth century, psychologists were always engaged in experimental research. The German Society of Psychology was even founded as the Society of Experimental Psychology in 1904 and chose its current name in 1929. Although I have carried out laboratory experiments (Lösel and Bliesener 2003), I was more interested in field experiments, for example in a matched pairs evaluation of an alternative to remand prisons for young offenders (Lösel and Pomplun 1997) or a long-term group-wise randomized study of developmental prevention (Lösel et al. 2013). Field experiments often contain more threats to internal validity than laboratory experiments, but they may have greater external validity.

Campbell's (1969) seminal article on reforms as experiments inspired me to strive for the best possible design under any given circumstances. Unfortunately, his differentiated view seems to be forgotten in black and white controversies about experiments in criminology. He described how to assess a treatment effect in experiments and sound quasi-experiments under various conditions (for details, see Shadish et al. 2002). Within this context, we developed a differentiated assessment of threats to validity in criminological treatment studies (Lösel and Köferl 1989). Although large RCTs with no systematic attrition are optimal for causal inference, I have repeatedly applied quasi-experimental designs when I was asked to evaluate a program that had been implemented years before (which does not allow an RCT). Should one then say "no" and leave a program without any systematic evaluation? Cronbach et al. (1980) emphasized that at least some controlled evaluation is better than none.

My own and many other evaluations mainly focus on the program or treatment content. If the outcome is positive, the authors conclude that the program works. If there is a nonsignificant or even negative outcome, they discuss more or less convincing reasons why the program did not work under these circumstances. However, researchers rarely ask to what extent the result could have been due to the conditions in the control or comparison group. For example, a young colleague carried out an evaluation of the social therapeutic prison in Erlangen (Bavaria). He did not find a significant treatment effect in comparison with a well-matched group of offenders who were released from regular prisons with no treatment. At nearly the same time, another study reported desirable effects of social therapy in a prison at Berlin. This led us to reanalyze and compare both sets of data (Lösel and Egg 1997), and we found that the rates of any recidivism and also of serious recidivism were the same in the treatments in Erlangen and Berlin. However, the control group in the Erlangen study had lower recidivism rates than the control group in Berlin. Obviously, it was not the treatment condition, but the somewhat better control group in Bavarian prisons, that led to a nonsignificant treatment effect.

I also experienced the necessity of a closer look at control or comparison conditions in an RCT on training programs for prison officers. We had developed a structured

cognitive behavioral program to improve the interaction skills of prison officers. Four randomly selected groups received a training on the structured program, whereas four others participated in more group-dynamic approaches that were popular at this time. In the short-term and follow-up assessments, we found various positive effects of the CBT program (Lösel and Wittmann 1989). However, when we took a closer look at the content of the video-recorded program delivery, we observed that the CBT groups were homogeneous in basic educational dimensions such as the group climate, structure, and stimulation. Although the group-dynamic courses significantly differed on average from the CBT courses, they were also more heterogeneous. One course had very similar educational characteristics to the CBT groups and it also showed similarly desirable effects. Obviously, a black and white concept of treatment versus comparison conditions was too simple.

My experience of the importance of the control group content is not unique. For example, in developmental prevention, there are contradictory results on multisystemic therapy (MST) from independent researchers. Ogden and Amlund Hagen (2006) found a desirable effect of MST in Norway, but Sundell et al. (2008) reported no desirable results in Sweden. Canadian and Dutch studies also showed different results (Asscher et al. 2014; Leschied and Cunningham 2002). These differences may be due to the typical variation in findings on interventions for juveniles (Lipsey 2018; Lösel 2018). However, one should also ask to what extent the different findings are due to different control group conditions. Sundell et al. (2008) assumed that the nonsignificant MST effect in Sweden may have been caused by the very intensive family support system in this country that makes it difficult to detect an effect of an additional program. More generally, Andrée Löfholm et al. (2013) analyzed evaluations of MST in relation to the treatment-as-usual (TAU) conditions in control groups. Their findings showed that there was greater variation in underlying risks in the control groups than in the treatment groups. They concluded that the content of TAU may have been more important for the evaluation outcomes than the content of the MST treatment.

In the field of offender treatment, one could expect larger effects in more recent studies because programs have improved. However, this seems not to be the case (Schmucker and Lösel 2015; Tong and Farrington 2006). This may be due to more rigorous evaluation designs in more recent years, which often lead to smaller effects, but it is also possible that the conditions in the control groups improved over time because knowledge about "what works" diffused beyond specific programs. Such system changes are difficult to evaluate. However, it is a simple rule that one can obtain stronger intervention effects when a control group has rather unfavorable "normal" or TAU conditions.

The importance of control group conditions is not a specific issue of criminology. For example, Karlsson and Bergmark (2015) asked "compared to what?" for Cochrane and Campbell reviews of psychosocial treatment of substance use disorders. They found widely differing control conditions in the respective studies that were a major source of differential results. These and other examples suggest that more attention to the type and content of control group conditions is highly important. It is obvious that criminology needs to pay more attention to this often neglected issue. Recent articles on criminological prevention and treatment research suggest that we should not focus too much on brand names and model programs, but apply evidence-based characteristics of generic interventions in practice (Lipsey 2018). Lipsey emphasized four dimensions: the content

of the program, the quality of service delivery, treatment intensity, and the participants' risk level. Lösel (2018) proposed seven particularly promising principles: a multimodal concept, sound theoretical foundation, good quality of program delivery, staff competence, a favorable context, medium- to high-risk participants, and proper monitoring.

The AEC has substantially contributed to the knowledge about core features of successful interventions. This brief article suggests that we should also pay more attention to the specific conditions in the control group of criminological evaluations.

Randomized experiments and public policy: comments and cautions (by David Weisburd)

When I began my career, experimental design was something you learned about, but I mostly learned to think it was not possible in the real world. My "trial by fire" occurred in the late 1980s when Lawrence Sherman and I planned the Minneapolis Hot Spots Patrol Experiment (Sherman and Weisburd 1995). We wanted to test traditional thinking about police patrol. Scholars at the time stated without much equivocation that the police could not deter crime and certainly not through police patrol (Bayley 1994; Gottfredson and Hirschi 1990). David Bayley said on the first page of his important book *Police for the Future*, published in the early 1990s, that the idea that the police could prevent crime was a "myth" (1994, p. 3). Sherman and I set out to prove that this was wrong. We believed that random police patrol across large areas could not deliver enough dosage for a reasonable expectation of deterrence. But if the police focused resources on high activity places or hotspots, we could reasonably expect deterrence and a crime prevention impact.

Our problem was that we needed to design a study that would be persuasive enough to challenge conventional thinking. Another observational study of a police intervention that used matched groups of places or multivariate statistical methods to control for confounding was not likely to alter the strong prevailing "nothing works" narrative in policing (Weisburd and Braga 2006). However, a well-designed and implemented randomized field trial would in our view provide compelling evidence if its results showed that the police could prevent crime. This view of the potential impact of randomized experiments was bolstered by Sherman's successful randomized experiment on domestic violence cases in policing (Sherman and Berk 1984), which gained considerable public and academic visibility. The domestic violence experiment provided not only an example of the influence of experimental trials but also evidence that large-scale experimental trials could be implemented successfully in policing.

While the domestic violence experiment showed that experiments could be carried out successfully in policing, a study that sought to impact not upon specific incidents but the entire patrol effort of a police department presented new challenges. The fact that we were able to successfully implement the study, and to provide the required dosage of patrol to hotspots for most of the experimental period, provided evidence that large-scale randomized field trials could be implemented that impacted core elements of the police function. Its success was a major impetus behind the growth of randomized experiments in policing over the following decades (Braga et al. 2014).

The impact of the Minneapolis study confirmed our gut feeling that a randomized study could have great influence in overturning conventional thinking. The impact was

not immediate but, along with a series of other experimental studies developed after Minneapolis, upended the underlying views of scholars and practitioners about the potential for police to be effective in preventing crime. By 2004, the US National Research Council/National Academy of Sciences could conclude the following: "[S]tudies that focused police resources on crime hot spots provide the strongest collective evidence of police effectiveness that is now available. On the basis of a series of randomized experimental studies, we conclude that the practice described as hot-spots policing is effective in reducing crime and disorder" (Skogan and Frydl 2004, p. 250). In turn, today, most large American police agencies employ hotspots policing approaches (Weisburd and Majmundar 2018).

Why are experiments so persuasive as a tool for advancing public policy? Simply stated, when experiments are conducted with fidelity, they provide convincing and easy to understand evidence of program effectiveness. In this case, the most rigorous approach to evaluating programs or practices is also the easiest to understand. This duality of scientific credibility and simplicity of interpretation are key reasons why experiments have a unique "power" for advancing public policies. Dan Nagin and I made this point in an essay published in *Criminology and Public Policy* a few years ago:

> In scientific settings, high evidentiary value is given to findings from randomized experiments because it is widely understood that, if properly conducted, randomized experiments provide convincing evidence of causality. This feature of experimental findings increases their evidentiary standing in policy settings as well; however, it is not the only reason for their influential status in policy settings. Another is their transparency. Findings from randomized experiments are far easier for policy makers and practitioners to understand than findings from more technically elaborate statistical analyses, for example, various forms of regression analysis of nonexperimental data. The transparency of experiments increases the comfort level of policy makers in acting on experiment-based findings (Nagin and Weisburd 2013, p. 653).

What I learned early in my career was that randomized experiments have a prestige all of their own. Despite much opposition in criminology, experiments carry tremendous weight. Despite the fact that experiments are difficult to carry out, and take tremendous effort and time to develop, they are worth their weight in gold—so to speak, if you want to influence public policy.

While recognizing the power of experiments as a tool to influence public policy, I want to end with a few cautionary comments on the future of experimental criminology. I am worried by a type of ritualization in experimental science in criminology. When we were advocating experimental research in the face of large-scale opposition from many criminologists, we sometimes had to fight hard against prevailing evaluation paradigms. And supporters of those paradigms often had vested interests in pushing back against experimental research. If we did not get up and shout about the advantages of experiments, we would not have advanced as far as we have.

But now that we have come so far, I worry that we do not think enough about the limitations of experimental research, or indeed recognize enough the value of nonexperimental methods. I am not arguing against the advantages of experimental studies. What I am suggesting is that we must be intelligent and thoughtful about the

contributions of experimental designs. Experiments can be designed poorly, or examine the wrong questions. Simply because they are experiments does not mean that they trump findings from other research methods. The gold standard provided by experimental research is that, when conducted with fidelity, they provide more believable answers than any other method. That does not mean that they always provide more believability. This is a complexity I think that we need to recognize and embrace.

Advancing quasi-experimentation in criminology (by Anthony A. Braga)

Randomized controlled trials are more common in criminology today, compared with the 1980s. However, these designs continue to represent a small share of the total number of impact evaluations conducted in criminal justice policy areas on a yearly basis. Field settings often limit the use of randomized experiments in criminal justice program evaluations. Too frequently, little or no thought is invested in evaluation plans when programs are implemented. Political, ethical, and/or practical concerns sometimes lead evaluators to conclude that random assignment is not feasible when evaluation plans are made (see, e.g., Clarke and Cornish 1972; Erez 1986). The continued growth of criminological experiments in a broad range of real-world settings that have been carried out in an ethical manner demonstrates that many of these concerns are, in most cases, based in folklore rather than facts (Weisburd 2010). Nevertheless, quasi-experiments are often used as alternative approaches in program evaluations. While experimental criminologists, and organizations such as the AEC, have led the call for the increased use of randomized controlled trials, they have also been leading advocates for the advancement of rigorous quasi-experimental methods in crime and justice research.

A quasi-experimental design seeks to approximate characteristics of a true experiment without the benefit of random allocation of units to treatment and control conditions (Shadish et al. 2002). The ability of criminologists to use more rigorous quasi-experimental designs continues to improve over time, with advances in statistical techniques and improvements in the availability and quality of criminal justice datasets. Statistical matching techniques, such as propensity scores (Rubin 1990), represent an important advance in the ability of evaluators to develop equivalent comparison groups. Crime mapping technologies and the construction of databases to examine crime events at very small levels of aggregations, such as specific street segments and intersections, allow greater flexibility in developing treatment and control units in quasi-experiments (see, e.g., Braga et al. 2011). Other quasi-experimental approaches that suggest the potential for gaining higher levels of internal validity are becoming more popular in criminology. For example, the regression discontinuity design takes advantage of treatments that are administered at a specific cutting point in the data to see whether the expected regression lines below and above the data are similar (see, e.g., Ludwig and Miller 2007).

Quasi-experimental designs are usually not able to produce treatment and control groups that are alike on all possible characteristics and, as such, quasi-experimental findings are generally regarded as having lower internal validity compared with randomized controlled trials. Some scholars suggest that quasi-experiments combining the use of a control group with time series data can sometimes produce results that are of similar quality to randomized controlled trials (Lipsey and Wilson 1993). For instance,

Berk et al. (2010) found that the results from a randomized controlled trial were essentially identical to the results from a regression discontinuity quasi-experiment when both designs were implemented in the same setting. Others, however, have reported that even strongly designed quasi-experiments produce less valid outcomes compared with well-executed randomized controlled trials (see Weisburd et al. 2001). In general, the persuasiveness of quasi-experiments should be judged on a case-by-case basis.

An ongoing Campbell Collaboration systematic review of focused deterrence policing programs supports the position that criminologists have increasingly used more rigorous quasi-experimental designs to estimate program impacts over time. The original version of the focused deterrence review found that less than one third of the eligible studies used quasi-experimental designs with matched comparison groups (Braga and Weisburd 2012). The less rigorous studies tended to use conveniently selected large areas, such as cities, as the treated and untreated units of analysis. In contrast, almost two-thirds of the newly identified studies in a recently updated version of the focused deterrence review used these more rigorous controlled designs (Braga et al. 2018). The more rigorous studies tended to use much smaller in-city units of analysis, such as gangs or census tracts, to form matched treated and untreated groups. The updated systematic review found that the nonequivalent quasi-experimental designs were associated with a much larger within-group effect size (0.703) relative to the matched quasi-experimental designs (0.194) (Braga et al. 2018). In both versions of the focused deterrence review, none of the eligible studies used a randomized experimental design. However, shortly after the completion of the updated review, the first randomized experiment evaluating focused deterrence was completed (Hamilton et al. 2018). Similar to the overall conclusion of the updated systematic review, the randomized experiment revealed that focused deterrence produced a modest crime control impact.

Rigorous quasi-experimental evaluation techniques are important options in the toolkits of experimental criminologists. However, well-executed randomized controlled trials provide superior confidence in program evaluation findings, given their high levels of internal validity relative to common quasi-experimental methods such as matched pairs and propensity score techniques. Indeed, the landmark Cambridge-Somerville Youth Study longitudinal randomized experiment would not have detected the harmful effects of the program if quasi-experimental evaluation methods had been used (McCord 2003). The words of a former AEC president, the late Joan McCord, remain a rallying cry for experimental criminologists. As she observed (McCord 2003, p. 29), "whenever possible," evaluation studies "should employ random assignment."

In a recent article, Nagin and Sampson (2019, p. 123) argued that the "counterfactual worlds that matter most to social science and policy" go beyond the particular units of analysis included in short-term experimental studies. They suggest that criminologists should also test different treatment regimens as applied to all eligible population members over a sustained period of time and claim that "police executives are done a disservice if they are led to believe that the experimental findings answer the question that should be paramount to them—will this work system-wide?" (Nagin and Sampson 2019, p. 136). While I agree with their position, it is worth noting that inappropriately lowering or raising the expectations of police executives about the benefits associated with particular crime prevention programs is equally problematic. Studies with lower levels of internal validity, such as observational analyses and weak quasi-experiments,

are more likely to lead to such faulty conclusions. Indeed, the city-level focused deterrence evaluations (some of which are considered to be very rigorous tests involving longer observation periods; see Morgan and Winship 2007) tended to inflate the crime control benefits of these programs.

Nagin and Sampson's (2019) argument boils down to the well-known problem of external validity (Shadish et al. 2002). A study can have very high internal validity but be relevant only to a very limited number of contexts or problems. I do not know of any empirical investigation that has demonstrated a systematic relationship between external validity and various kinds of study designs (observational, quasi-experimental, randomized experimental). Observational studies of citywide policy interventions implemented in particular jurisdictions and using extended observation periods seem likely to face generalization challenges rooted in concerns over unique and idiosyncratic study settings, procedures, and participants, relative to other populations and conditions. The problem of external validity should be kept in mind when reviewing all study findings, regardless of the design used. I am sure that Nagin and Sampson would concur. Looking forward, experimental criminologists should focus on the robust implementation of evaluations with very high levels of internal validity such as randomized experiments and rigorous quasi-experiments.

Conclusions (by Friedrich Lösel and David P. Farrington)

The contributions from past presidents of the Academy of Experimental Criminology in this article raise several important issues. In accordance with the mission of the AEC, there is now a substantial body of experimental and sound quasi-experimental studies in criminology. This is not only indicated by the success of the *Journal of Experimental Criminology*, but many experimental studies have also been published in other outlets. In comparison with the rarity of criminological experiments in previous times (Farrington 2003a), experiments are nowadays more common. Twenty years ago, there was more resistance to experiments in both criminological research and practice. Some opponents may have never read Campbell's (1969) seminal article on "Reforms as experiments" that was a guide to how to evaluate programs as validly as possible in different circumstances. The contributors to the present article discussed a broad range of experimental and sound quasi-experimental studies on topics such as policing, developmental prevention, offender rehabilitation, moral behavior and dishonesty, and biosocial origins of violence.

The above contributions also show that the former AEC presidents are not advocating experiments as a ritualized "gold standard" of criminological research, but recommend the application of RCTs whenever they are feasible under given theoretical, practical, legal, political, and sometimes ethical circumstances. The polarization of experimental versus nonexperimental criminologists does not acknowledge the methodological breadth of many researchers. For example, various contributors to this article have carried out RCTs, but are also engaged in prospective longitudinal studies with correlational designs and in the ASC Division of Developmental and Life-Course Criminology.

The authors of this article have addressed problems of both internal and external validity in field experiments and there is a clear message that RCTs should be preferred

when they are feasible because they are the most robust design to establish valid causal inference. Internal validity is the basis of causal inference and this is the reason why the natural sciences, engineering, and large parts of medicine successfully (but not only) rely on experiments. Of course, in the natural sciences, it is possible to hold constant extraneous variables by physical methods, although a substantial number of studies could not even be replicated in these fields of "hard science" (e.g., Baker 2016). In the social sciences, we need to equate extraneous variables in experimental and control conditions by randomization, but there can be some threats to valid causal inference even in RCTs (see Farrington 2003c; Lösel 2007).

As an example, Maltz (2006) pointed out that a significant decrease in crime in an area after a program was implemented could be driven by a very small number of offenders. He contrasted two situations. In the first case, a few hundred offenders each committed one or two crimes per year before the time period. Then, the program was implemented, causing about 100 offenders to quit or go elsewhere or be arrested, leading to a reduction in crimes from 500 before to 400 after. In the second case, five high-rate offenders each committed 100 or so offenses per year in the before time period. Then, the program was implemented, causing one offender to quit or go elsewhere or be arrested, and again, the number of crimes reduced from 500 to 400. In the second case, the lack of independence of the crimes would threaten the statistical significance of the decrease. Maltz argued that it was essential to talk to people on the ground (e.g., police and/or offenders) to find out what was really happening in this kind of study.

Although internal validity is the precondition of causal inference in the social sciences, there is also the need for external validity to real-life situations. Fellows of the AEC strongly recommend RCTs, but are aware that scientific progress can result from other approaches as well. For example, the first heart transplant was a courageous case study that may have received a Nobel Prize if Christiaan Barnard (as he assumed) had not been a citizen of South Africa with its then apartheid regime. From a scientific perspective, it needs to be emphasized that this achievement was only possible because of previous experimental research on cell reactions, appropriate medication, etc. The famous research on congenital trigger mechanisms, for which Konrad Lorenz received a Nobel Prize, began with naturalistic observations of geese in a pond, but the robustness of the theory and the underlying mechanisms was systematically validated later in experiments. There are many other examples of productive relations between rigorous (quantitative) experiments, correlational designs, and sometimes even qualitative research (e.g., Lösel 1985).

Controversies about experiments in criminology often neglect the fact that causality is a rather complex construct that has various facets (Bunge 1959). Most experimental researchers in criminology adhere to the principles of causality that Donald Campbell, Thomas Cook, and others have developed since the 1960s: Two variables must be correlated and must have a clear time ordering, a change in variable *X* must lead to a change in variable *Y*, and alternative explanations of a finding (threats to internal validity) must be excluded (see Shadish et al. 2002).

When internal validity has been demonstrated, further research should carry out thorough tests on the external validity of a finding, that is, its generalization across individuals (or units), interventions, measurements, situational contexts, and times. The contributors to this article are aware that some experiments in criminology are not yet

so well replicated that they allow wide-ranging generalization of findings (Farrington et al. 2019; Lösel 2018). However, the same applies to correlational research and other nonexperimental intervention designs. In the latter case, one even has to ask: what can be generalized if the research is not internally valid? As various contributors to this article have emphasized, sound and replicated experiments are particularly helpful for policy making. Systematic reviews and meta-analyses also play a key role in advancing knowledge. Accordingly, these reviews are well represented in the *Journal of Experimental Criminology*.

The above-mentioned theoretical and conceptual issues should be considered in controversial discussions about the advantages of experimental criminology. In their ASC session at Atlanta and in their published article, Nagin and Sampson (2019) questioned whether experiments are the "gold standard" for criminological research. Most participants in the AEC's 20th anniversary session would agree with at least some of their arguments, and this is also reflected in the above contributions. For example, Nagin and Sampson emphasized the importance of studying theoretical concepts and counterfactual outcomes, instead of a ritualized application of experimental designs. The present article clearly shows that experimental criminologists are flexible and not closed-minded ritualists. It should also be mentioned that Nagin and Sampson are not against RCTs. For example, Sampson (2008) commended the Moving to Opportunity experiment on housing programs. Nagin and Pogarsky (2003) carried out an RCT in which they tested conditions of cheating; this laboratory study had some similarities to the use of more externally valid field experiments on dishonesty that David Farrington recommended above.

Experimental criminologists endorse the importance of theory, and well-implemented experiments are the most rigorous method of testing causal hypotheses. Although criminology is a rather small discipline, we have numerous theories that fill thousands of pages in books and journals. Criminology had a phase of mainly theoretical discussions in the 1950s–1960s and then a strong increase in empirical research, sometimes without much theoretical foundation. However, even widely acknowledged theories, such as that on self-control, have rather limited explanatory power (Lösel 2017), and there is a clear need to reduce the proliferation of fragmented theories in criminology (Bruinsma 2016). As Bernard (1990) pointed out, no criminological theory has ever been falsified, and few theories have been revised in light of new findings and new empirical tests. If criminology is to be a science, robust quantitative predictions need to be derived from criminological theories and tested in the most convincing designs, preferably experimental (Farrington et al. 2016).

The great advantage of large-scale RCTs is that they control not only for all measured variables but also for all possible unmeasured variables. Since criminologists cannot control extraneous variables physically, an RCT is the best method of controlling them and hence of maximizing internal validity and demonstrating that *X* causes *Y*. This is the main justification for experimental criminology and the mission of AEC. In addition, it is also important to obtain sound knowledge about what findings can be safely generalized. Although studies with nonexperimental and internally less valid designs may particularly emphasize external validity, they have the same problems of generalization (e.g., Weisburd 2010). For these and other reasons, policy makers and practitioners should seek to carry out experiments whenever feasible and preferably *before* implementing expensive programs.

References

Alem, Y., Eggert, H., Kocher, M. G., & Ruhinduka, R. D. (2018). Why (field) experiments on unethical behavior are important: comparing stated and revealed behavior. *Journal of Economic Behavior and Organization, 156*, 71–85.

Andrée Löfholm, C., Brannstrom, L., Bergstöm, M., & Hansson, K. (2013). Treatment-as-usual in effectiveness studies: what is it and does it matter? *International Journal of Social Welfare, 22*, 25–34.

Ares, C. E., Rankin, A., & Sturz, H. (1963). The Manhattan Bail Project: an interim report on the use of pretrial parole. *NYU Law Review, 38*, 67.

Asscher, J. J., Deković, M., Manders, W., van der Laan, P. H., Prins, P. J. M., & Arum, S. (2014). Sustainability of the effects of multisystemic therapy for juvenile delinquents in the Netherlands: effects on delinquency and recidivism. *Journal of Experimental Criminology, 10*, 227–243.

Baker, M. (2016). Is there a reproducibility crisis? *Nature, 533*, 452–455.

Bayley, D. H. (1994). *Police for the future*. New York: Oxford University Press.

Berk, R., Barnes, G., Ahlman, L., & Kurtz, E. (2010). When second best is good enough: a comparison between a true experiment and a regression discontinuity quasi-experiment. *Journal of Experimental Criminology, 6*, 159–190.

Bernard, T. J. (1990). Twenty years of testing theories: what have we learned and why? *Journal of Research in Crime and Delinquency, 27*, 325–347.

Braga, A. A., Hureau, D., & Papachristos, A. V. (2011). An ex-post-facto evaluation framework for place-based police interventions. *Evaluation Review, 35*, 592–626.

Braga, A. A., & Weisburd, D. (2012). The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. *Journal of Research in Crime and Delinquency, 49*, 323–358.

Braga, A. A., Weisburd, D., & Turchan, B. (2018). Focused deterrence strategies and crime control: an updated systematic review and meta-analysis of the empirical evidence. *Criminology & Public Policy, 17*, 205–250.

Braga, A. A., Welsh, B. C., Papachristos, A. V., Schnell, C., & Grossman, L. (2014). The growth of randomized experiments in policing: the vital few and the salience of mentoring. *Journal of Experimental Criminolology, 10*, 1–28.

Bruinsma, G. J. N. (2016). Proliferation of crime causation theories in an era of fragmentation: reflections on the current state of criminological theory. *European Journal of Criminology, 13*, 659–666.

Buikhuisen, W. (1974). General deterrence: research and theory. *Abstracts in Criminology and Penology, 14*, 285–298.

Bunge, M. (1959). *Causality: the place of the causal principle in modern science*. Cambridge: Harvard University Press.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409–429.

Choy, O., Raine, A., & Hamilton, R. H. (2018). Stimulation of the prefrontal cortex reduces intentions to commit aggression: a randomized, double-blind, placebo-controlled, stratified, parallel-group trial. *Journal of Neuroscience, 38*, 6505–6512.

Choy, O., Raine, A., Portnoy, J., Rudo-Hutt, A., Gao, Y., & Soyfer, L. (2015). The mediating role of heart rate on the social adversity-antisocial behavior relationship: a social neurocriminology perspective. *Journal of Research in Crime and Delinquency, 52*, 303–341.

Clarke, R. V., & Cornish, D. (1972). *The controlled trial in institutional research*. London: H.M. Stationery Office.

Cornet, L. J. M., de Kogel, C. H., Nijman, H. L. I., Raine, A., & van der Laan, P. H. (2014). Neurobiological factors as predictors of cognitive-behavioral therapy outcome in individuals with antisocial behavior: a review of the literature. *International Journal of Offender Therapy and Comparative Criminology, 58*, 1279–1296.

Cornet, L. J. M., de Kogel, C. H., Nijman, H. L. I., Raine, A., & van der Laan, P. H. (2015). Neurobiological changes after intervention in individuals with anti-social behaviour: a literature review. *Criminal Behaviour and Mental Health, 25*, 10–27.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward a reform of program evaluation*. San Francisco: Jossey Bass.

Erez, E. (1986). Randomized experiments in correctional context: legal, ethical, and practical concerns. *Journal of Criminal Justice, 14*, 389–400.

Farrington, D. P. (1979). Experiments on deviance with special reference to dishonesty. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 207–252). New York: Academic Press.

Farrington, D. P. (1980). External validity: a problem for social psychology. In R. F. Kidd & M. J. Saks (Eds.), *Advances in applied social psychology* (Vol. 1, pp. 184–186). Hillsdale: Lawrence Erlbaum.

Farrington, D. P. (1983). Randomized experiments on crime and justice. In M. Tonry & N. Morris (Eds.), *Crime and justice* (Vol. 4, pp. 257–308). Chicago: University of Chicago Press.

Farrington, D. P. (2003a). A short history of randomized experiments in criminology: a meager feast. *Evaluation Review, 27*, 218–227.

Farrington, D. P. (2003b). British randomized experiments on crime and justice. *The Annals of the American Academy of Political and Social Science, 589*, 150–167.

Farrington, D. P. (2003c). Methodological quality standards for evaluation research. *The Annals of the American Academy of Political and Social Science, 587*, 49–68.

Farrington, D. P. (2013). Longitudinal and experimental research in criminology. In M. Tonry (Ed.), *Crime and justice in America 1975–2025* (pp. 453–527). Chicago: University of Chicago Press.

Farrington, D. P., & Kidd, R. F. (1977). Is financial dishonesty a rational decision? *The British Journal of Social and Clinical Psychology, 16*, 139–146.

Farrington, D. P., Knapp, W. S., Erickson, B. E., & Knight, B. J. (1980). Words and deeds in the study of stealing. *Journal of Adolescence, 3*, 35–49.

Farrington, D. P., & Knight, B. J. (1979). Two non-reactive field experiments on stealing from a "lost" letter. *The British Journal of Social and Clinical Psychology, 18*, 277–284.

Farrington, D. P., & Knight, B. J. (1980a). Four studies of stealing as a risky decision. In P. D. Lipsitt & B. D. Sales (Eds.), *New directions in psycholegal research* (pp. 26–50). New York: van Nostrand Reinhold.

Farrington, D. P., & Knight, B. J. (1980b). Stealing from a "lost" letter: effects of victim characteristics. *Criminal Justice and Behavior, 7*, 423–436.

Farrington, D. P., Lösel, F., Boruch, R. F., Gottfredson, D. C., Mazerolle, L., Sherman, L. W., & Weisburd, D. (2019). Advancing knowledge about replication in criminology. *Journal of Experimental Criminology, 15*, 373–396 .

Farrington, D. P., MacLeod, J. F., & Piquero, A. R. (2016). Mathematical models of criminal careers: deriving and testing quantitative predictions. *Journal of Research in Crime and Delinquency, 53*, 336–355.

Farrington, D. P., & Petrosino, A. (2001). The Campbell collaboration crime and justice group. *The Annals of the American Academy of Political and Social Science, 578*, 35–49.

Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: what have we learned in the last two decades? *Journal of Experimental Criminology, 1*, 9–38.

Farrington, D. P., & Welsh, B. C. (2006). A half-century of randomized experiments on crime and justice. In M. Tonry (Ed.) *Crime and Justice* (Vol. 34, pp. 55-132). Chicago: University of Chicago Press.

Franzen, A., & Pointner, S. (2013). The external validity of giving in the dictator game: a field experiment using the misdirected letter technique. *Experimental Economics, 16*, 155–169.

Gabor, T., & Barker, T. (1989). Probing the public's honesty: a field experiment using the "lost letter" technique. *Deviant Behavior, 10*, 387–399.

Gabor, T., Strean, J., Singh, G., & Varis, D. (1986). Public deviance: an experimental study. *Canadian Journal of Criminology, 28*, 17–29.

Gajos, J. M., & Beaver, K. M. (2016). The effect of omega-3 fatty acids on aggression: a meta-analysis. *Neuroscience and Biobehavioral Reviews, 69*, 147–158.

Gesch, C. B., Hammond, S. M., Hampson, S. E., Eves, A., & Crowder, M. J. (2002). Influence of supplementary vitamins, minerals and essential fatty acids on the antisocial behaviour of young adult prisoners: randomised, placebo- controlled trial. *British Journal of Psychiatry, 181*, 22–28.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford: Stanford University Press.

Hamilton, B., Rosenfeld, R., & Levin, A. (2018). Opting out of treatment: self-selection bias in a randomized controlled study of a focused deterrence notification meeting. *Journal of Experimental Criminology, 14*, 1–17.

Karlsson, P., & Bergmark, A. (2015). Compared with what? An analysis of control-group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substances use disorders. *Addiction, 110*, 420–428.

Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. (1974). *The Kansas City preventive patrol experiment: a technical report*. Washington, DC: Police Foundation.

Leschied, A., & Cunningham, A. (2002). *Seeking effective interventions for serious young offenders: interim results of a four-year randomized study of multisystemic therapy in Ontario, Canada*. London, Canada: Centre for children and families in the justice system.

Lipsey, M. W. (2018). Effective use of the large body of research on the effectiveness of programs for juvenile offenders and the failure of the model programs approach. *Criminology & Public Policy, 17*, 1–10.

Lipsey, M. W., & Wilson, D. B. (1993). *Practical meta-analysis*. Thousand Oaks: Sage.

Lipton, D. S., Martinson, R., & Wilks, J. (1975). *The effectiveness of correctional treatment: a survey of treatment evaluation studies*. New York: Praeger.

Liu, J. H., Raine, A., Venables, P. H., Dalais, C., & Mednick, S. A. (2004). Malnutrition at age 3 years and externalizing behavior problems at ages 8, 11 and 17 years. *American Journal of Psychiatry, 161*, 2005–2013.

Lösel, F. (1985). Zur Kontroverse um eine gegenstandsangemessene psychologische forschung: Bemerkungen aus der sicht der forschungspraxis. [The controversy about adequate psychological research methodologies: remarks from the perspective of research practice.]. In W.-F. Kugemann, S. Preiser, & K. A. Schneewind (Eds.), *Psychologie und komplexe Lebenswirklichkeit. [Psychology and complex reality.]* (pp. 65–91). Göttingen: Hogrefe.

Lösel, F. (2007). Doing evaluation in criminology: balancing scientific and practical demands. In R. D. King & E. Wincup (Eds.), *Doing research on crime and justice* (2nd ed., pp. 141–170). Oxford: Oxford University Press.

Lösel, F. (2017). Self-control as a theory of crime: a brief stocktaking after 27/42 years. In C. Bijleveld & P. van der Laan (Eds.), *Liber amoricum for Gerben Bruinsma* (pp. 232–238). The Hague: Boom.

Lösel, F. (2018). Evidence comes by replication, but needs differentiation: the reproducibility issue in science and its relevance for criminology. *Journal of Experimental Criminology, 14*, 257–278.

Lösel, F., & Bliesener, T. (2003). *Aggression und delinquenz unter jugendlichen: Untersuchungen von kognitiven und sozialen bedingungen. [Aggression and delinquency in adolescence: studies on cognitive and social origins.]*. Neuwied: Luchterhand.

Lösel, F., & Egg, R. (1997). Social-therapeutic institutions in Germany: description and evaluation. In E. Cullen, L. Jones, & R. Woodward (Eds.), *Therapeutic communities in prisons* (pp. 181–203). Chichester: Wiley.

Lösel, F., & Köferl, P. (1989). Evaluation research on correctional treatment in West Germany: a meta-analysis. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system* (pp. 334–355). New York: Springer.

Lösel, F., & Pomplun, O. (1997). *Jugendhilfe statt Untersuchungshaft: Eine Evaluationsstudie zur Heimunterbringung. [Residential care instead of pretrial detention of juvenile offenders: an evaluation.]*. Pfaffenweiler: Centaurus Verlag.

Lösel, F., & Wittmann, W. (1989). The relationship of treatment integrity and intensity to outcome criteria. *New Directions for Program Evaluation, 42*, 97–107.

Lösel, F., Stemmler, M., & Bender, D. (2013). Long-term evaluation of a bimodal universal prevention program: effects from kindergarten to adolescence. *Journal of Experimental Criminology, 9*, 429–449.

Ludwig, J., & Miller, D. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics, 122*, 159–208.

Maltz, M. D. (2006). Some p-baked thoughts (p>0.5) on experiments and statistical significance. *Journal of Experimental Criminology, 2*, 211–226.

McCord, J. (2003). Cures that harm: unanticipated outcomes of crime prevention programs. *The Annals of the American Academy of Political and Social Science, 587*, 16–30.

Mihalic, S. F., & Elliott, D. S. (2015). Evidence-based programs registry: blueprints for healthy youth development. *Evaluation and Program Planning, 48*, 124–131.

Morgan, S., & Winship, C. (2007). *Counterfactuals and causal inference: methods and principles for social research*. New York: Cambridge University Press.

Nagin, D. S., & Pogarsky, G. (2003). An experimental investigation of deterrence: cheating self-serving bias, and impulsivity. *Criminology, 41*, 167–194.

Nagin, D., & Sampson, R. (2019). The real gold standard: measuring counterfactual worlds that matter most to social science and policy. *Annual Review of Criminology, 2*, 123–145.

Nagin, D., & Weisburd, D. (2013). Evidence and public policy: the example of evaluation research in policing. *Criminology & Public Policy, 12*, 651–679.

National Research Council. (1977). *Understanding crime: an evaluation of the National Institute of law enforcement and criminal justice*. Washington, DC: National Academies Press. https://doi.org/10.17226/13536.

Ogden, T., & Amlund Hagen, K. (2006). Multisystemic treatment of serious behaviour problems in youth: sustainability of effectiveness two years after intake. *Journal of Child and Adolescent Mental Health, 11*, 142–149.

Raine, A., Ang, R. P., Choy, O., Hibbeln, J. R., Ho, R., Lim, C. G., et al. (2018). Omega-3 (w-3) and social skills interventions for reactive aggression and childhood externalizing behavior problems: a randomized, stratified, double-blind, placebo-controlled, factorial trial. *Psychological Medicine, 49*, 335–344.

Raine, A., Cheney, R. A., Ho, R. G., Portnoy, J., Liu, J. L., Soyfer, L., et al. (2016). Nutritional supplementation to reduce child aggression: a randomized, stratified, single-blind, factorial trial. *Journal of Child Psychology and Psychiatry, 57*, 1038–1046.

Raine, A., Mellingen, K., Liu, J. H., Venables, P., & Mednick, S. A. (2003). Effects of environmental enrichment at ages 3-5 years on schizotypal personality and antisocial behavior at ages 17 and 23 years. *American Journal of Psychiatry, 160*, 1627–1635.

Raine, A., Portnoy, J., Liu, J. H., Mahoomed, T., & Hibbeln, J. R. (2015). Reduction in behavior problems with omega-3 supplementation in children aged 8-16 years: a randomized, double-blind, placebo-controlled, stratified, parallel-group trial. *Journal of Child Psychology and Psychiatry, 56*, 509–520.

Raine, A., Venables, P. H., Dalais, C., Mellingen, K., Reynolds, C., & Mednick, S. A. (2001). Early educational and health enrichment at age 3-5 years is associated with increased autonomic and central nervous system arousal and orienting at age 11 years: evidence from the Mauritius Child Health Project. *Psychophysiology, 38*, 254–266.

Rubin, D. B. (1990). Formal modes of statistical inferences for causal effects. *Journal of Statistical Planning and Inference, 25*, 279–292.

Sampson, R. J. (2008). Moving to inequality: neighborhood effects and experiments meet social structure. *American Journal of Sociology, 114*, 189–231.

Schmucker, M., & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: an international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology, 11*, 597–630.

Sechrest, L. B., White, S. O., & Brown, E. D. (1979). *The rehabilitation of criminal offenders: problems and prospects*. Washington, DC: National Academy of Sciences.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for general causal inference*. Boston: Houghton-Mifflin.

Sherman, L. W. (1992). *Policing domestic violence: experiments and dilemmas*. New York: Free Press.

Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review, 49*, 261–272.

Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (Eds.). (2002). *Evidence-based crime prevention*. New York: Routledge.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J. E., Reuter, P., & Bushway, S. D. (1997). *Preventing crime: what works, what doesn't, what's promising*. Washington, DC: National Institute of Justice.

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": a randomized, controlled trial. *Justice Quarterly, 12*, 625–648.

Sigelman, C. K., & Sigelman, L. (1976). Authority and conformity: violation of a traffic regulation. *Journal of Social Psychology, 100*, 35–43.

Skogan, W. G., & Frydl, K. (Eds.). (2004). *Fairness and effectiveness in policing: the evidence*. Washington, DC: National Academies Press.

Sundell, K., Hansson, K., Löfholm, C. A., Olsson, T., Gustle, L. H., & Kadesjö, C. (2008). Multisystemic therapy and traditional services for antisocial adolescents in Sweden: results from a randomized controlled trial after six months. *Journal of Family Psychology, 22*, 550–560.

Tong, L. S. J., & Farrington, D. P. (2006). How effective is the Reasoning and Rehabilitation programme in reducing offending? A meta-analysis of evaluations in four countries. *Psychology, Crime & Law, 12*, 3–24.

Umbach, R., Raine, A., & Leonard, N. R. (2018). Cognitive decline as a result of incarceration and the effects of a CBT/MT intervention: a cluster-randomized controlled trial. *Criminal Justice and Behavior, 45*, 31–55.

Weisburd, D. (2010). Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: challenging folklore in evaluation research in crime and justice. *Journal of Experimental Criminology, 6*, 209–227.

Weisburd, D., & Braga, A. A. (2006). Introduction: Understanding police innovation. In D. Weisburd & A. A. Braga (Eds.), *Police innovation: contrasting perspectives*. Cambridge: Cambridge University Press.

Weisburd, D., Lum, C., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science, 578*, 50–70.

Weisburd, D., & Majmundar, M. K. (Eds.). (2018). *Proactive policing: effects on crime and communities*. Washington, DC: National Academies Press.

Weisburd, D., Mazerolle, L., & Petrosino, A. (2007). The Academy of Experimental Criminology: advancing randomized trials in crime and justice. *The Criminologist, 32*(3), 1–7.

Zaalberg, A., Nijman, H., Bulten, E., Stroosma, L., & van der Staak, C. (2010). Effects of nutritional supplements on aggression, rule-breaking, and psychopathology among young adult prisoners. *Aggressive Behavior, 36*, 117–126.

David P. Farrington is an Emeritus Professor of Psychological Criminology at Cambridge University. He has received the Stockholm Prize in Criminology and he has been the President of the American Society of Criminology. His major research interest is in developmental criminology, and he is the Director of the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 61. In addition to 799 published journal articles and book chapters on criminological and psychological topics, he has published 111 books, monographs, and government publications and 156 shorter publications (total = 1066).

Friedrich Lösel is an Emeritus Professor and past Director at the Institute of Psychology, University of Erlangen-Nuremberg (Germany), and at the Institute of Criminology, Cambridge University (UK). He carried out research on topics such as juvenile delinquency, offender treatment, prisoners, football hooliganism, bullying, psychopathy, resilience, developmental prevention, and extremism. His oeuvre contains 415 articles or chapters and 38 monographs, edited volumes, and special issues. He is a recipient of the ASC Sellin-Glueck Award, Lifetime Award of the European Association of Psychology and Law, Jerry Lee Award of the ASC Division of Experimental Criminology, Lifetime Award of the ASC Division of Developmental & Life-Course Criminology, Joan McCord Award of the Academy of Experimental Criminology, German Psychology Prize, and Stockholm Prize in Criminology.

Anthony A. Braga is a distinguished Professor and Director of the School of Criminology and Criminal Justice at Northeastern University. He is a Fellow of the American Society of Criminology. He is also a past President of the Academy of Experimental Criminology and the 2014 recipient of its Joan McCord Award.

Lorraine Mazerolle is an Australian Research Council Laureate Fellow (2010–2015), a Professor of Criminology in the School of Social Science at the University of Queensland, and a Chief Investigator with the Centre of Excellence for Children and Families over the Life Course. Her research interests are in experimental criminology, policing, drug law enforcement, regulatory crime control, and crime prevention. Professor Mazerolle is the recipient of the 2018 ASC Sellin-Glueck Award, the 2016 ASC Division of Policing Distinguished Scholar Award, the 2013 AEC Joan McCord Award, and the 2010 ASC Division of International Criminology Freda Adler Distinguished Scholar Award.

Adrian Raine is the Richard Perry University Professor of Criminology, Psychiatry, and Psychology at the University of Pennsylvania. He has published over 450 journal articles, book chapters, and books on the etiology and prevention of antisocial, criminal, and psychopathic behavior in children and adults. He is a past President of the Academy of Experimental Criminology, the co-founding Chair of the Division of Biopsychosocial Criminology at ASC, and recipient of both an honorary degree from the University of York (UK), and also the Lifetime Achievement Award from the Society for the Scientific Study of Psychopathy.

Lawrence W. Sherman is a Wolfson Professor of Criminology Emeritus at the University of Cambridge and a Distinguished University Professor at the University of Maryland. He has received the Benjamin Franklin Medal of the Royal Society of Arts and served as the President of the American Society of Criminology and the American Academy of Political and Social Science. His major research interest is in evidence-based policing. Being the Director of the Jerry Lee Centre of Experimental Criminology at Cambridge, his most cited work includes the Minneapolis Domestic Violence Experiment, Hot Spots of Predatory Crime, and Defiance Theory.

David Weisburd is a distinguished Professor of Criminology, Law, and Society at George Mason University and the Walter E. Meyer Professor of Law and Criminal Justice at the Hebrew University. His primary research interests in recent years have been focused on policing, experimental criminology, evidence-based

policy, and crime at place. He is the 2010 recipient of the Stockholm Prize in Criminology, the 2014 recipient of the Sutherland Award, the 2015 recipient of the Israel Prize, and the 2017 recipient of the Vollmer Award. He received his Ph.D. in sociology from Yale University.

Affiliations

David P. Farrington[1] · Friedrich Lösel[2,3] · Anthony A. Braga[4] · Lorraine Mazerolle[5] · Adrian Raine[6] · Lawrence W. Sherman[2,7] · David Weisburd[8,9]

1 Institute of Criminology, Cambridge University, Sidgwick Avenue, Cambridge CB3 9DA, UK

2 Cambridge University, Cambridge, UK

3 University of Erlangen-Nuremberg, Erlangen, Germany

4 Northeastern University, Boston, USA

5 University of Queensland, Brisbane, Australia

6 University of Pennsylvania, Philadelphia, USA

7 University of Maryland, College Park, USA

8 George Mason University, Fairfax, USA

9 Hebrew University, Jerusalem, Israel