What economics can contribute to the addiction sciences

Jonathan P. Caulkins1 & Nancy Nicosia2
Carnegie Mellon University Heinz College and Qatar Campus, Pittsburgh, PA, USA1 and RAND Drug Policy Research Center, Boston, MA, USA2

ABSTRACT

Aims The addiction sciences are intrinsically multi-disciplinary, and economics is among the disciplines that offer useful perspectives on the complex behaviors surrounding substance abuse. This paper summarizes contributions economics has made in the past and could make in the future towards understanding how illegal markets operate, how prices affect use, how use generates various consequences, and how policy shapes all three. Methods Review of literature, concentrating on illegal drugs as insights concerning markets are particularly salient, although we also mention relevant studies from the alcohol and tobacco fields. Findings and Conclusions Economics offers tools and topical expertise that usefully complement other disciplines associated traditionally with the addiction sciences. Its value goes far beyond the ability to monetize non-monetary outcomes or to calculate a cost-benefit ratio.

Keywords Addiction, demand elasticity, drug policy, economics.

INTRODUCTION

Compulsive behaviors that harm self and others are inherently complex. When those behaviors involve the consumption of goods produced and distributed through illegal markets, that complexity is amplified. This complexity implies that a full understanding of behaviors surrounding illegal addictive substances must be rooted in the perspectives of multiple disciplines.

The field of economics offers insight into this domain. Given others’ past contributions [1,2], we do not attempt to be comprehensive. We focus instead upon how economics can contribute to understanding in four key areas: (i) the demand for drugs by ‘consumers’ (users); (ii) the supply of drugs by producers, traffickers and sellers; (iii) the consequences of use; and (iv) the effectiveness of policies designed to ameliorate those consequences. We focus consciously upon upside potential. Someone could, instead, fill an article with missteps made in the name of the discipline, but multi-disciplinarity succeeds by drawing on the best of each discipline.

DEMAND: UNDERSTANDING INDIVIDUAL DRUG-USER BEHAVIOR

Perhaps the most fundamental insight gained from economics is that drugs can be thought of as consumer goods. In particular, drug consumption obeys the ‘Law of Demand’, meaning that the higher the price, the less product people will buy. At one time that statement was controversial. Intuitively, people thought casual users might respond to price, but the prevailing wisdom was that ‘addicts have to get their fix’. However, the evidence shows that even indicators associated with heavy or dependent users also respond to price, perhaps because drug use often accounts for a large share of a heavy user’s budget. For example, negative correlations with price have been documented for treatment admissions, emergency department mentions, ambulance call-outs and urinalysis results among arrestees [3–6].

A more precise way to think about this relationship is in terms of the ‘elasticity of demand’, defined as the percentage change in consumption associated with a 1% increase in price. If the percentage change in consumption is greater in absolute value than the percentage change in price, then the good is ‘price-elastic’. For price-elastic goods, driving up price reduces not only drug use, but also drug spending. Conversely, if consumption responds less than proportionately, the good is ‘price-inelastic’. Inelastic demand does not mean that consumption is impervious to price; rather, it means the change in consumption is smaller proportionally than the change in price. In this case, increases in price drive up total

INTRODUCTION

Compulsive behaviors that harm self and others are inherently complex. When those behaviors involve the consumption of goods produced and distributed through illegal markets, that complexity is amplified. This complexity implies that a full understanding of behaviors surrounding illegal addictive substances must be rooted in the perspectives of multiple disciplines.

The field of economics offers insight into this domain. Given others’ past contributions [1,2], we do not attempt to be comprehensive. We focus instead upon how economics can contribute to understanding in four key areas: (i) the demand for drugs by ‘consumers’ (users); (ii) the supply of drugs by producers, traffickers and sellers; (iii) the consequences of use; and (iv) the effectiveness of policies designed to ameliorate those consequences. We focus consciously upon upside potential. Someone could, instead, fill an article with missteps made in the name of the discipline, but multi-disciplinarity succeeds by drawing on the best of each discipline.

DEMAND: UNDERSTANDING INDIVIDUAL DRUG-USER BEHAVIOR

Perhaps the most fundamental insight gained from economics is that drugs can be thought of as consumer goods. In particular, drug consumption obeys the ‘Law of Demand’, meaning that the higher the price, the less product people will buy. At one time that statement was controversial. Intuitively, people thought casual users might respond to price, but the prevailing wisdom was that ‘addicts have to get their fix’. However, the evidence shows that even indicators associated with heavy or dependent users also respond to price, perhaps because drug use often accounts for a large share of a heavy user’s budget. For example, negative correlations with price have been documented for treatment admissions, emergency department mentions, ambulance call-outs and urinalysis results among arrestees [3–6].

A more precise way to think about this relationship is in terms of the ‘elasticity of demand’, defined as the percentage change in consumption associated with a 1% increase in price. If the percentage change in consumption is greater in absolute value than the percentage change in price, then the good is ‘price-elastic’. For price-elastic goods, driving up price reduces not only drug use, but also drug spending. Conversely, if consumption responds less than proportionately, the good is ‘price-inelastic’. Inelastic demand does not mean that consumption is impervious to price; rather, it means the change in consumption is smaller proportionally than the change in price. In this case, increases in price drive up total

INTRODUCTION

Compulsive behaviors that harm self and others are inherently complex. When those behaviors involve the consumption of goods produced and distributed through illegal markets, that complexity is amplified. This complexity implies that a full understanding of behaviors surrounding illegal addictive substances must be rooted in the perspectives of multiple disciplines.

The field of economics offers insight into this domain. Given others’ past contributions [1,2], we do not attempt to be comprehensive. We focus instead upon how economics can contribute to understanding in four key areas: (i) the demand for drugs by ‘consumers’ (users); (ii) the supply of drugs by producers, traffickers and sellers; (iii) the consequences of use; and (iv) the effectiveness of policies designed to ameliorate those consequences. We focus consciously upon upside potential. Someone could, instead, fill an article with missteps made in the name of the discipline, but multi-disciplinarity succeeds by drawing on the best of each discipline.

DEMAND: UNDERSTANDING INDIVIDUAL DRUG-USER BEHAVIOR

Perhaps the most fundamental insight gained from economics is that drugs can be thought of as consumer goods. In particular, drug consumption obeys the ‘Law of Demand’, meaning that the higher the price, the less product people will buy. At one time that statement was controversial. Intuitively, people thought casual users might respond to price, but the prevailing wisdom was that ‘addicts have to get their fix’. However, the evidence shows that even indicators associated with heavy or dependent users also respond to price, perhaps because drug use often accounts for a large share of a heavy user’s budget. For example, negative correlations with price have been documented for treatment admissions, emergency department mentions, ambulance call-outs and urinalysis results among arrestees [3–6].

A more precise way to think about this relationship is in terms of the ‘elasticity of demand’, defined as the percentage change in consumption associated with a 1% increase in price. If the percentage change in consumption is greater in absolute value than the percentage change in price, then the good is ‘price-elastic’. For price-elastic goods, driving up price reduces not only drug use, but also drug spending. Conversely, if consumption responds less than proportionately, the good is ‘price-inelastic’. Inelastic demand does not mean that consumption is impervious to price; rather, it means the change in consumption is smaller proportionally than the change in price. In this case, increases in price drive up total
spending on drugs, potentially exacerabting spending-related harms such as property crime and/or impoverishment of users.

A large and increasingly international literature seeks to estimate various price elasticities of drug use, including ‘participation elasticity’ (the ratio of change in number of users to change in price) and ‘conditional elasticity’ (the ratio of change in quantity consumed among users to change in price). For example, Chaloupka et al. [7] estimate a past-year participation elasticity for cocaine among youth of −0.89 and an overall elasticity of −1.28 when adding the intensity response. Price responsiveness can vary with the type of user and context. The price elasticity for cannabis observed among US high school seniors (−0.33) is lower than among Australian adults (−0.88) [8,9]; differences are also observed across age and gender groups within Australia [10]. Legal status may also matter: elasticity estimates were higher for regulated, legal opium [11,12] than for illegal heroin [5,6,13]. Estimates also differ depending upon whether income is constrained artificially (as in some laboratory studies) or whether users can respond at least partially to price changes by increasing or reducing income [14]. Economists also study ‘cross price-elasticities’, meaning how changes in the price of one drug affects consumption of another. This literature is still young, but tends to find complementary relationships across different types of substances, although valium and, to some extent, marijuana appear to be substitutes for heroin and cocaine [15]. Complementarity versus substitution may also depend upon the time-frame observed. Some drug pairs may appear to be substitutes in the short term if polydrug users, for example, adjust their use in response to relative price changes even though the drugs are complements in the long term.

Consumption responding to price does not imply that users have perfect foresight, maximize utility successfully, or are otherwise rational [16]. Rational addiction theory (RAT), pioneered famously by Becker & Murphy [17] and extended by Dockner, Engelbert & Feichtinger [18], among others, showed that many behaviors previously considered irrational can be consistent with fully informed optimal choice with stable preferences. For economists, RAT is important because it shows that a key assumption underpinning conventional welfare economics (that consumers make self-interested choices that further their own welfare) is not necessarily threatened by addiction.

For non-economists, the greater benefit has been the follow-on work relaxing certain aspects of perfect rationality, while preserving the idea that drug users are making choices, albeit constrained and imperfect ones [19]. This newer work draws upon the pioneering contributions of Herrnstein (the matching law) [20], Simon (bounded rationality) [21] and Kahneman & Tversky (prospect theory) [22]; the latter two ideas led to Nobel Prizes in economics. Their key insight was that people are incapable of perfect choices, so they follow rules of thumb or heuristics. Usually heuristics lead to good outcomes, but as Skog [16] observes, addictive substances can bring out the worst in our heuristic decision-making.

Perhaps the best-documented example of a heuristic is that people discount future outcomes relative to immediate consequences [23], as economic theory suggests, but that discounting is hyperbolic rather than exponential [24]. Hyperbolic discounting implies that the rate of substitution between two consecutive time-periods in the near term is smaller than in any successive periods, so that immediate consequences are weighted more heavily. For example, a user may be more likely to adjust behavior to avoid an additional week in prison starting today than to avoid an additional month tacked onto the end of a 5-year sentence.

Gruber & Köszegi [25] show that modeling consumption as forward-looking but time-inconsistent is just as compatible with the empirical evidence as RAT, but with dramatically different implications. Whereas, if consumers are fully rational, the government’s role is limited to ensuring that participants are fully informed and do not impose negative ‘externalities’ on others, this modest change raises the possibility that government interventions may also improve social welfare by reducing the ‘internalities’ that users themselves potentially suffer.

Hyperbolic discounting is just one form of time inconsistency. Bretteville-Jensen [26] suggests that being a current injection user can increase the discount rate relative to former or never users. Badger and colleagues [27] find similar effects from craving; and there are other ways, beyond time inconsistency, in which decision-making can depart from the rationality that liberal political economy assumes when delegating government to a minimalist role. A more fundamental issue may be individuals’ inability to imagine alternate cognitive states. It is easy to believe that someone considering trying drugs for the first time might not understand what it feels like to be addicted. However, Lowenstein et al. [28] argue convincingly that this ‘projection bias’ is pervasive. People even have difficulty recalling states they already experienced, which may explain the patterns of recurring relapse and temporary abstinence common to addiction.

Merging psychologically plausible models of decision-making into an economic framework generates exciting possibilities for developing more effective interventions [29]. Pioneering ideas include token economies and contingency management, which might be characterized loosely as methods to induce treatment compliance by offering small, immediate, positive incentives [30], and
coerced or mandated abstinence, which is similar with respect to immediacy and modest stakes, but which sanctions non-compliance [31]. Both address the limitations of self-control while incorporating—indeed embracing—users’ autonomy.

**SUPPLY: UNDERSTANDING DRUG DISTRIBUTORS AND DRUG MARKETS**

Comprehensive understanding of drug markets must also consider the supply side, and economics is well equipped to aid in that understanding. The literature on licit addictive substances, such as alcohol and tobacco, pays considerable attention to retail availability, advertising and market structure [32,33]. When society prohibits a substance, the supply side matters even more. As Kleiman [34] notes, prohibition changes the nature of the problem from primarily pertaining to use and dependence to pertaining substantially to black markets and drug control. Drug markets cause enormous harms that are often ignored by cost of illness studies [35]. Also, supply-control spending dominates drug control budgets even in countries that embrace harm reduction, such as the Netherlands [36] and Australia [37].

Economists’ contributions begin with characterizing the ‘structure, conduct, and performance’ for production [38–41], high-level trafficking [42,43] and street-level distribution [44,45]. This work has punctured many myths, such as that most drug dealing is dominated by powerful monopolies; roughly 90% of incarcerated drug sellers in the United States reported that they were not part of an organized group [46]. The work has also resolved conundrums, such as why production is concentrated in so few countries [47], and helped to explain why globalization makes it difficult to thwart international distribution [48,49].

Prohibition dramatically increases transaction costs, defined as search and information costs, bargaining costs and policing and enforcement costs [50–52]. Even experienced heroin users may spend more than half an hour acquiring the drug and do so several times a day [53]. Evidence suggests that moral hazard influences the nature of transactions and relationships between buyers and sellers [54]. Prohibition’s impact on purity-adjusted price is even more dramatic. Heroin and cocaine are semi-refined agricultural products, akin to sugar or coffee; yet they sell for many times their weight in gold. The opium needed to produce 1 g of pure heroin costs only 55 cents in Afghanistan, but fetches 250 times as much in Europe [55]. The increase is due largely to prohibition and enforcement’s impact on distribution costs: shipping a kilogram of cocaine from Colombia to Washington DC costs $13,000 compared to only $42 for licit goods [55,56].

Our understanding of what drives variation in price within a prohibitionary regime is more limited. However, it appears that once prohibition is backed by enough enforcement to impose its structural consequences [57], further policy changes produce relatively modest effects on price and use [58]. Modeling studies [59,60] and econometric analysis [61] show that even the United States’ dramatic expansion in supply control increased prices only modestly.

Frailties of human judgement can help to explain why enforcement has limited deterrent effect on sellers [62]. Some go further, arguing that increasing enforcement might actually increase use [63,64] and that basing sanctions on the quantity possessed might have perverse consequences on use [65], violence [66,67] and property crimes [68]. The fundamental explanation is that drug markets, like all markets, innovate in response to changing circumstances; even interventions that disrupt supply successfully usually engender only temporary reductions in use and consequences [69].

Similarly, one of economics’ early contributions was to point out that quantity seized is a poor performance metric of enforcement success [70]. Contrary to the old ‘physical flow’ model, seizing a kilogram does not reduce availability by a kilogram; rather, it creates an incentive for greater production. However, seized drugs and incarcerated workers are probably not replaced at a one-for-one rate. Suppliers can recoup some of the cost of replacing seized drugs by increasing prices, but as discussed above, higher prices reduce consumption [71]. This insight may explain why a number of studies document enforcement successes [61,72–76]. The effects are sometimes transient and may not be large enough to justify their costs, but the markets nevertheless react as expected.

**CONSEQUENCES OF ADDICTION**

Considerable research seeks to understand the societal consequences of drug use, including the consequences of early drug use on later drug use [77]. For example, cost-of-illness studies consider the direct health costs, the burden of dependence and the indirect costs due to crime, mortality, lower productivity and other indirect consequences of drug use [78–81], although some question the extent to which these studies are useful to inform policy [82]. The correlations noted in these studies may not imply causality, because third factors such as proclivity for taking risks may determine simultaneously both substance use and these consequences.

Economics has a particularly powerful toolkit for identifying causal relationships using non-experimental data [83,84]. Non-experimental methods were developed because randomized controlled trials, the gold standard
for inferring causality, are often infeasible. (We would not usually allocate subjects randomly to different levels of use or cities to different levels of drug enforcement.) Here, we discuss how these methods have been applied to understanding the indirect consequences of both alcohol and illicit drug use.

Substance use is substantially higher among offenders and many report committing crimes while under the influence [85,86], but how many of these crimes would not have been committed in the absence of use? Carpenter [87] assesses the causal relationship between alcohol and crime by leveraging exogenous variation across US states in the adoption of zero-tolerance drunk-driving laws. Zero-tolerance laws decreased the fraction of nuisance and property crime arrests attributable to 18–20-year-olds, but had no impact on violent crime. The fact that these same policy changes had no effect on older cohorts provides a credible falsification test. Using an instrumental variables strategy, Dobkin & Nicosia [69] examined a disruption to methamphetamine supply in California. That the study found little evidence of declines in property and violent crime despite a substantial reduction in proxies for use suggests either that use does not cause a large share of crime, or that the intervention did not reduce use and/or spending significantly among criminally involved users. Norstrom & Skog’s [88] work on Sweden takes advantage of a phase-in of retail alcohol sales on Saturdays, with similar findings. While alcohol sales increased in the relevant regions as a result of the increased availability, crime-related harms, including assaults and drunk driving, did not. Conversely, leveraging a plausibly exogenous change in alcohol consumption due to a tax cut in Sweden, Andreasson et al. found that increased consumption resulted in 1627 additional assaults [89].

Concerns about selection also apply to mortality: individuals may act recklessly under the influence, but selection makes the extent of the causal relationship unclear. Carpenter & Dobkin [90] leveraged the exogenous increase in alcohol access that occurs at the minimum legal drinking age (MLDA). Comparing individuals just above and below the MLDA, they found a 21% increase in the number of days individuals drink and, consequently, a 9% increase in mortality due to accidents, overdoses, suicides and other external causes. Using exogenous shocks to income based on the timing of Supplemental Security Income disbursements in the United States, another study found a 2.2% increase in mortality among drug users [91]. Similar approaches have been used to study mortality outside the United States. For example, Andreasson et al. [89], described above, linked increased alcohol use to increased deaths in Sweden. Similarly, research has linked the large but temporary reductions in alcohol consumption during the Gorbachev anti-alcohol campaign with substantial declines in alcohol-related deaths, although there are concerns about data and methods [92].

These methods have also been used to examine the extent to which drug use reduces the academic achievement and productivity of users and their children. Kremer & Levy [93] evaluated the consequence of alcohol use on academic achievement by exploiting the random assignment of college room-mates via lottery. Random assignment to a room-mate who drank alcohol prior to college reduced male students’ grade point averages (GPAs) among those at the low end of the GPA distribution. Similarly, using a fixed-effects design to address unobservable differences across high-risk youth, Engberg & Morral [94] found evidence that alcohol and stimulant use reduces school attendance, as van Ours & Williams [95] do vis-à-vis cannabis. More recent studies have examined the longer-term effects on the children of users. Nilsson [96], for example, examines whether a natural experiment creating regional variation in Sweden’s alcohol availability resulted in poorer educational and labor market outcomes among those exposed in utero.

Much more remains to be carried out beyond considering additional outcomes. Future research could consider whether the consequences differ across substance types, across casual versus heavy use levels, and in the short versus the long term. Addressing these questions will probably require similar non-experimental techniques because of similar concerns about selection.

**POLICY**

Once one understands the extent to which drug use causes various consequences, it is natural to ask where limited resources should be directed to produce the greatest reduction in harms. Cost-effectiveness analysis requires understanding the relative value of consequences, how effective policies are at reducing their incidence, and how much those policies cost [60,97,98]. Holding costs constant, a cost-effective policy may be one that deters large amounts of use or it may be one that reduces costly consequences such as mortality or incarceration (potentially despite minimal reductions in overall drug use).

Economists have conducted yeoman work providing guidance on how to compare the value (i.e. monetize) the many disparate consequences of use. The literature continues to wrestle with some methodological issues, such as whether and how to value pain and suffering, particularly that of the user [99]. Following other health domains where the key metric is quality-adjusted life year (QALY), initial efforts suggest that the losses experienced by dependent users may be substantial [78].

Program evaluation methods are particularly advanced for evaluating treatment interventions [100]. There is now an enormous literature arguing that treat-
ment, particularly opiate substitution therapies, offers a favorable return on investment from a societal perspective [101–103], although comparisons across treatment programs are sometimes less decisive [104,105].

However, a fundamental concern is that both supply and demand reduction efforts outside treatment are less amenable to these evaluation methods because randomized controlled trials are often not possible. For example, we generally cannot change the legal drinking age for a random subset of teenagers or increase border interdiction for randomly selected parts of a country. Here again, non-experimental methods can provide insight.

Broadly implemented policy changes can serve as natural experiments to evaluate effectiveness. Carpenter & Dobkin’s [90] results described above suggest that reducing the drinking age in the United States by 1 year would result in approximately 408 additional deaths among 20-year-olds. Similarly, Lu & McGuire [106] used an instrumental variables approach to isolate the effectiveness of out-patient care in reducing drug use. They found that treatment is effective for moderate and heavy drug users and identified the cut-off at which treatment becomes completely ineffective. Taking a different approach, Saffer et al. [107] estimated the effect of state expenditures on criminal justice and public health on drug use using a supply-and-demand model. They found that both drug treatment and enforcement reduce use, but that correctional spending does not.

There are other opportunities to employ these methods. From the enforcement perspective, this might encompass changes in legislation or penalties for possession or sale of illicit substances [10,72,108] or market shocks such as the Australian heroin drought [109]. From the treatment perspective, the implementation of substance abuse parity, expansion of drug courts and diversion programs and budget cuts induced by the recent recession provide opportunities to exploit a substantial change in treatment availability. For example, an evaluation of California’s Proposition 36 prison diversion policy found it to be cost-effective, but not necessarily effective. In essence, Proposition 36 reduced the costs of incarceration because drug offenders were diverted to less costly treatment services rather than prison, but the policy did not necessarily reduce drug use or crime, perhaps because the population was no longer incapacitated [110,111]. Further research is needed to understand the long-term consequences of these large-scale policy changes [112], but for reviews of what is understood at present, see [32,99,113].

CONCLUSION

Economics is often considered as the study of things related to money, so often economists are invited to the table only when outcomes need to be monetized as in cost–benefit studies. While it is true that economists have developed tools for monetizing diverse outcomes, that skill is only a small part of what economists have to offer addiction sciences.

A better, although still imperfect, characterization would be that economists can offer both methodological and substantive expertise. Economists’ toolkit includes powerful methods for inferring causal relations from non-experimental data, and economists’ topical expertise includes markets and how policy affects markets and market-related behavior. This makes economics a powerful complement to disciplines that focus more upon experimental data or for which the usual unit of analysis is the individual, not the society or governing jurisdiction. Economics also embraces both the positive (what is) and normative (what should be), so it is well suited to informing policy decision-making.

There are challenges. For example, economics can be criticized as being somewhat insular. Papers in economics journals rarely cite research from other disciplines, and young economists are often counseled to focus upon economics journals. Nevertheless, the potential fruits of recruiting more economists into studying addiction-related problems are great.

Declarations of interest

None.

Acknowledgements

This work was supported in part by the Qatar Foundation, a Robert Wood Johnson Foundation Health Investigator Award, and the National Institute on Drug Abuse.

References

23. Kirby K. N., Petry N. M. Heroin and cocaine abusers have
19. Heyman G. M.
18. Dockner E., Feichtinger G. Cyclical consumption patterns