Don't Shoot The Messengers:

**A Response to Ellis and Hoskin's (2018) Commentary on the Prenatal Androgen
Hypothesis**

Travis C. Pratt

University of Cincinnati

Jillian J. Turanovic

Florida State University

Alex R. Piquero

University of Texas at Dallas

Francis T. Cullen

University of Cincinnati

In our two recently published meta-analyses of the literature assessing the “prenatal androgen hypothesis,” we failed to find supportive evidence in favor of the 2D:4D digit ratio—a rather crude proxy for exposure to fetal testosterone—being predictive of various forms of criminal, antisocial, and analogous behaviors (see Pratt, Turanovic, & Cullen, 2016; Turanovic, Pratt, & Piquero, 2017). Ellis and Hoskin (2018) recently published their own assessment of the hypothesis, which stood in part as a commentary on our previous work. Unlike the results from our meta-analyses, these authors argued that the weak to nonexistent effects of the digit ratio that we found in our studies were an artifact of certain methodological decisions that were made, and that our interpretation of the effect sizes that we calculated was unnecessarily pessimistic. As a result, Ellis and Hoskin (2018) maintained that evolutionary neuroandrogenic (ENA) theory—a perspective that relies heavily on the use of the 2D:4D digit ratio as a measure of its key construct—remains well supported in the face of negative evidence. We would like to take a little time to set the record straight about our methods and interpretations, and to do that we need to begin by explaining why we conducted these meta-analyses in the first place.

Meta-Analyses of the 2D:4D Digit Ratio and Criminal/Analogous Behavior

We were first introduced to the 2D:4D digit ratio with Hoskin and Ellis’s (2015) study linking the digit ratio to criminal behavior. Using data drawn a sample of several hundred U.S. college students along with other individuals recruited by these students, the authors found that across a wide array of criminal behaviors—interpersonal violent offenses like assault and robbery, non-contact property crimes like theft and fraud, and even drug use and gambling—their results showed some digit ratio effects that, in terms of magnitude, *dwarfed every other*

criminogenic risk factor known to us as criminologists. Not only were these effects large, but they were stunningly consistent (typically in the $r = .20$ range, with some extending higher and even upwards of $.40$ — effects that Hoskin and Ellis, 2015, p. 61 described as “substantial”; see Table 2 and 3 on pp. 62-63). Indeed, the associations between the digit ratio and criminal behavior were twice as large as those usually seen for self-control or deviant peers (particularly in the multivariate models)—two of the strongest and robust predictors of crime/deviance that have been found across hundreds of studies (see Pratt & Cullen, 2000; Pratt et al., 2010). The digit ratio effects were even larger than those typically seen for the relationship between prior criminal behavior and current criminal behavior (in other words: criminal behavior predicting criminal behavior, see Andrews & Bonta, 2010). The implications of these results were remarkable: over 50 years of criminological theories and statistical models aimed at predicting crime could be scrapped in favor of using a tape measure and long division. Their findings sparked our curiosity as to whether there might be a larger knowledge base regarding the link between the digit ratio and criminal/analogous behaviors, and whether that literature would end up revealing results similar to those found by Hoskin and Ellis.

And as it turned out, there was a larger literature out there with respect to the digit ratio and criminal (and crime-related) behavior. But nothing in that literature came close to demonstrating the kind of strong evidence as that presented by Hoskin and Ellis (2015). That was the point of the first meta-analysis by Pratt et al. (2016), which assessed 47 studies and 660 effect size estimates between the 2D:4D digit ratio and measures of crime and crime-analogous behaviors (e.g., aggression, impulsivity, risk seeking). As we discussed in our study, 76 percent of the effect sizes in this literature were null in their original studies. They were disproportionately clustered around zero; 85 percent were smaller than $.20$ in magnitude; and

they were not very consistent. Indeed, 35 of the 47 studies presented at least one correlation in the unexpected direction. With data that looked like this, it was not surprising that the overall bivariate mean effect size between the 2D:4D and crime, aggression, and risky/impulsive behavior was close to zero ($r = .047$).¹ And when we looked only at aggressive and violent behaviors in a follow-up meta-analysis (Turanovic et al., 2017), we found the overall mean effect size to be even weaker ($r = .036$). More concerning, the extensive moderator analyses we conducted and presented in our article revealed that there was little in the way of systematic variation in these effects that could be explained by any methodological differences observed across studies (e.g., the sampling frame, the hand measured, the race or age of the sample, or whether a caliper was used). These patterns looked a lot like random error—not at all what we would expect if the 2D:4D digit ratio was a valid biomarker of crime and analogous behaviors.²

Coming to the Digit Ratio's Defense

In their defense of the use of the digit ratio in criminological research, Ellis and Hoskin (2018) took issue with our meta-analyses on a couple of fronts. First, they argued that we diluted the digit ratio effects by including studies that assessed non-criminal outcomes (e.g., physical aggression, verbal aggression, impulsivity, and risk taking).³ We included such studies with

¹ In each meta-analysis the effects were reverse-coded so that positive values represented effects in the theoretically-expected direction (e.g., more exposure to fetal testosterone leading to more problematic behavior).

² Ellis and Hoskin (2018) were quick to concede that the digit ratio is a “crude,” “weak,” and “noisy” indirect indicator of exposure to fetal testosterone. Yet, they later argued that if digit ratio effects are to be found using such an unreliable measure, then the effects must actually be really “strong” (p. 42). But conflating a measure’s weakness with its potential predictive strength obscures the reality that unreliable measures tend, more often than not, to produce results that are just that: unreliable.

³ Ellis and Hoskin (2018) contend that such outcomes should only be assessed if they exhibit sex differences. But in their own work (Hoskin & Ellis, 2015) nearly a third of the criminal outcomes they examined (ranging from various

various outcomes for three reasons. First, the broader literature on exposure to fetal testosterone (or at least the digit ratio) indicates that it should have “general” effects. To be sure, far from being limited to particular forms of criminal behavior (e.g., interpersonal violence), scholars have also argued that the digit ratio should predict not only multiple forms of aggression and impulsive/risky behavior, but also a whole roster of human behaviors from success in sumo wrestling (Tamiya, Lee, & Ohtake, 2012) to the popularity of Polish actors (Karwowksi & Lebuda, 2014) to the belief in the paranormal and superstitions (Voracek, 2009). This is not a perspective that lacks ambition. So clearly there is a precedent for looking at non-criminal (yet crime-analogous) behaviors.

Second, there is a long-standing concept in criminology of the “generality of deviance”—that is, the tendency of those who engage in crime to also engage in several other forms of questionable behavior (even if such behavior does not have a legal rule against it; Gottfredson & Hirschi, 1990; Hirschi & Gottfredson, 1994). For instance, those who engage in crime are also more likely to do poorly in school, to drink too much and do drugs, to mess up romantic relationships, to find themselves in financial trouble, to alienate their friends, and so on (Pratt, Barnes, Cullen, & Turanovic, 2016). In fact, the most serious of offenders, or what Moffitt (1993) has referred to as life-course-persistent offenders, tend to be involved in a wide range of antisocial, criminal, and analogous behaviors throughout their lives. We therefore have no theoretical or empirical reason to believe that someone who engages in violent crime will somehow shy away from all other forms of problematic behavior. This is especially the case since there is virtually no evidence to support the existence of violent specialists (see Piquero,

forms of assault and theft) did not demonstrate sex differences to begin with (see Table 1, p. 61), yet they tested for the effect of the digit ratio on these outcomes anyway. If they had adhered to their own argument none of these relationships—relationships upon which they placed considerable substantive importance—should have been estimated in the first place.

Farrington, & Blumstein, 2003; Piquero, Jennings, & Barnes, 2012). As such, nearly three decades of criminological scholarship should have made our decision to include studies that examined aggression, impulsivity, and risk-taking non-controversial.

Finally, Hoskin and Ellis's (2015) own theoretical model explicitly states that exposure to fetal testosterone should result in "acting out," "emotional volatility," "aggression," "externalizing behavior," "hyperactivity," "risk-taking", to be "impulsive," and to engage in forms of "antisocial behavior" that may not be criminal (p. 56). It is a heightened predisposition for such behaviors that, in turn, should ultimately lead to criminal behavior down the line. Thus, our inclusion of studies that assessed these outcomes was wholly consistent with the theory that Ellis and Hoskin themselves laid out (see e.g., the theoretical figure in Ellis, 2005 that was reproduced in Hoskin & Ellis, 2015). Our decision is also reinforced by the work of Ellis, Beaver, and Wright (2009), whose discussion of the effects of testosterone on criminal behavior reviewed several studies linking testosterone to "various forms of aggression" (p. 210), where the positive effects were generally most consistent. They also treated as substantively important in their review studies that linked testosterone to other non-aggressive forms of criminal behavior, like illegal drug use, to "general" forms of criminal and delinquent behavior that make no distinction between aggressive and non-aggressive acts (Table 8.7.2a, p. 209), and to "clinical or personality traits associated with criminality," including externalizing behavior, sensation seeking, impulsivity, and callous-unemotionality (Table 8.7.2b, p. 211).

Still, a look at each of our meta-analyses will reveal that we did, in fact, present dozens of additional analyses *separately by type of outcome*. The Pratt et al. (2016) meta-analysis split the data in multiple ways—presenting both mean effect sizes and moderator analyses by type of outcome (by aggression, risky/impulsive behavior, and criminal behavior, just as Ellis et al.

[2009] did), by sex of the sample, and by scale of the outcome variable. The Turanovic et al. (2017) meta-analysis also presented mean effect sizes by sex and by type of outcome (by general aggression, physical aggression, verbal aggression, and violent offending; see p. 56, Table 3), and split the data into a total of 12 different subsamples in additional moderator analyses. In total, the Pratt et al. (2016) meta-analysis presented 107 regression models (25 in the main document, and 82 in appendices) and the Turanovic et al. (2017) study presented 23 (and summarized findings from an additional 118, see p. 57). Yet despite these analyses, Ellis and Hoskin (2018, p. 31) stated that, “Unfortunately, a coefficient was not reported for the 16 studies dealing exclusively with physical aggression” in the Turanovic et al. (2017) meta-analysis. This is simply incorrect. Physical aggression mean effect sizes can be readily found in Turanovic et al. (2017, p. 56, Table 3), and are presented for the overall sample ($r = .032$, n.s.), for males ($r = .032$, n.s.), and for females ($r = .007$, n.s.). Moderator analyses also showed that general aggression, physical aggression, verbal aggression, and violent offending effect sizes did not differ significantly from one another (see Turanovic et al., 2017, p. 57, Table 4).

But in the end it really did not matter. Whether the mean effect sizes included indicators of aggression or risk-seeking/impulsive behaviors or not, or whether they were composite scales, sex-specific, or broken down into physical or verbal aggression, the broad empirical pattern was the same: the effect sizes for the 2D:4D were weak, with little in the way of systematic variation attributable to methodological differences across studies. The point here is that the digit ratio effect was never diluted and it was never concealed. The aggression/offending-specific analyses conducted by Turanovic et al. (2017; Table 3) could not have been presented any clearer for readers to easily see how the mean effect size varied (or did not) through the different model

permutations. The bottom line is that the effects were almost universally nonexistent however the data were analyzed.

The Digit Ratio and Crime

The second issue raised by Ellis and Hoskin (2018) concerned a single mean effect size presented in the Pratt et al. (2016) meta-analysis for the relationship between the digit ratio and crime. The mean effect size in question was based on 6 studies and 88 effect sizes, with a magnitude of $r = .104$ (see Pratt et al., 2016, p. 601, Table 2). While this mean effect size was statistically significant, we placed little confidence in it in our original work. Ellis and Hoskin (2018) take issue with this decision and argue that we were overly dismissive of this effect—one that they deem as substantively important. And part of Ellis and Hoskin's (2018) justification for isolating the .104 effect size was that it stood in stark contrast to the non-significant effects for other behaviors like risk-taking. In particular, they (p. 31) noted that the overall mean effect size for risk-taking in Pratt et al. (2016) was “-.077 (NS).” Table 2 in the Pratt et al. (2016) study (p. 601), however, shows that the risk-taking effect *was, in fact, statistically significant* (that was what the asterisks next to the coefficient of .077 meant; see the legend in Table 2), and the moderator analyses in the very next table (p. 604) indicated that the risk-taking effects were significantly stronger than the effects for aggression.

In addition, Ellis and Hoskin (2018) take issue with our description of the small mean effect size for offending vis-à-vis our characterization of the mean effect size in a previous meta-analysis of a biosocial criminological issue: the relationship between maternal cigarette smoking during pregnancy and criminal/delinquent behavior in children (Pratt, McGloin, & Fearn, 2006). Ellis and Hoskin (2018) stated that Pratt et al. (2006) found a mean effect size of $r = .078$ for this

relationship—an effect was large enough to be referred to as substantively important. Ellis and Hoskin go on to state that “different standards” are being used when it comes to the digit ratio (p. 32). We are confused by this assertion for two reasons. First, the overall mean effect size in question in the Pratt et al. (2006) study was actually .117 (and not 0.78, see the first line of Table 2, p. 680). The value of .078 that Ellis and Hoskin specify in their critique—a figure that is noted three separate times in their commentary—*never even appears once* in the study by Pratt et al. (2006), leaving us perplexed as to where this number was drawn from. And second, the mean effect size of .117 was based on a larger N of 18 studies, not the 6 for the digit ratio finding in question. It was also remarkably consistent within the sample of studies assessed, and the effect sizes were statistically significant in 69 percent of their original models. We were therefore much more confident that the effect size in Pratt et al. (2006) was “real” (i.e., not a methodological artifact) even with the rather modest mean effect size.

The reason we placed little confidence in the mean effect size of .104 with respect to the digit ratio and crime had less to do with its magnitude, and more to do with the fact that it was unstable. Not only did our moderator analyses show that *there were no statistically significant differences between the effect sizes for crime and the effect sizes for aggression* (which were the weakest of the bunch; see Pratt et al., 2016, p. 604, Table 3), but in nearly all of our supplemental analyses—85 percent of them—the mean effect sizes for the digit ratio on crime were null (see Pratt et al., 2016, Tables S.2., S.4., S.5., and S.6.). The magnitude of the .104 mean effect size also appeared to be driven entirely by the inclusion of the Hoskin and Ellis (2015) study. Without it, the mean effect size for the digit ratio and crime was only .036 and was not statistically significant.⁴ We were therefore hesitant to give credence to a single finding

⁴ It is also curious that Ellis and Hoskin’s defense of the .104 effect size is rooted in their argument that ENA theory really only pertains to “physical aggression” and criminal acts that stem from it (hence their position that indicators

(among hundreds that were estimated) that was so unstable that it could disappear with the removal of a single study.⁵ Put simply, the data did not support that there was a robust or stable relationship between the digit ratio and criminal behavior.

Concluding Remarks

The meta-analyses by Pratt et al. (2016) and Turanovic et al. (2017) were not the first to point out problems with the use of the 2D:4D digit ratio as an indirect proxy for exposure to fetal testosterone (see Honekopp & Watson, 2011). The failure of the digit ratio to consistently correlate with direct measures of fetal testosterone is now well documented (there was a single positive study that was followed by 16 replication failures, see Voracek, 2014). Our meta-analyses did, however, shed light on just how unique Hoskin and Ellis's (2015) findings were from the rest of the literature. If, for example, a given study in our data found a link between the digit ratio and aggressive or criminal behavior that was in the $r = .20$ range, it was likely present in a correlation matrix where most of the other effects were null or even in the wrong direction (e.g., for full scales and subscales of the outcome measures, for right hand and left hand measures of the digit ratio, for full samples and subsamples disaggregated by gender—analyses that are common in this literature). Thus, a scatterplot of the results from every other study in our

of verbal aggression or impulsivity or risk taking should not have been assessed in our meta-analyses). Yet, much of what is driving the mean of .104 are Hoskin and Ellis's (2015) effect size estimates predicting forms of criminal behavior that do not involve any form of physical aggression at all (e.g., theft, financial crime, property damage, drug dealing, gambling, car racing). It appears that non-violent (and non-aggressive) forms of criminal or analogous behaviors only "count" for Ellis and Hoskin if they reveal significant effects (see, e.g., Ellis et al., 2009's enthusiastic discussion of significant testosterone effects for measures of drug use and prison violations).

⁵ Meta-analytic methods have long recognized the importance of identifying potential outliers. As Viechtbauer and Cheung (2010, p. 122) note, "When the removal of just one or two studies has a considerable impact on the conclusions from a meta-analysis, then these conclusions must be stated with some caution." It was this advice that led us to be skeptical of the mean effect of .104 (see also Higgins & Thompson, 2002). We still encourage that very caution.

meta-analytic data looked like a shotgun blast from several meters away. Hoskin and Ellis's (2015) results, on the other hand, looked like a used target from a sniper rifle—a tightly-grouped cluster of effects that rarely deviated from the bullseye.

Given those findings, we suggested that it might be time to bring a healthy dose of what Merton (1973) referred to as “organized skepticism” to the study of the consequences of the 2D:4D digit ratio. This does not mean that we set out to conduct these meta-analyses with any sort of agenda in mind, nor were we trying to use our meta-analyses to try to undermine in any way the biosocial tradition in criminology—an area of study that all of us view with respect for its impressive theoretical and empirical record. In fact, our own (and collaborative) work(s) have often tested—and have found consistent support for—several different biological/biosocial hypotheses (e.g., Barnes, Beaver, & Piquero, 2011; McGloin, Pratt, & Piquero, 2006; Tibbetts and Piquero, 1999; Unnever, Cullen, & Pratt, 2003). Instead, our focus was centered on whether the 2D:4D digit ratio—as a proxy measure of a core theoretical concept—could be reliably linked to crime and analogous behaviors. That the digit ratio did not perform well is simply what the data revealed.

It is also important to note that in conducting our meta-analyses we were not attempting to tear down or to otherwise critique ENA theory as an explanation of either criminal or aggressive behavior. As social scientists, we remain open to the idea that exposure to fetal testosterone is an important developmental phenomenon that could well influence a wide array of behaviors over the life course, whether they be criminal, aggressive, risky, impulsive, and/or antisocial. This is still an important—and in our estimation, open—empirical question and one that criminologists should continue to investigate (we made this point explicitly as well in Pratt et al., 2016). We are simply skeptical that measuring fingers is going to amount to a valid and

reliable way of testing ENA theory's core propositions. As researchers, we need to be like baseball umpires, calling balls and strikes when we see them. Sometimes, people disagree with the ump's call, but the data are what the data are.

Either way, we want to close this response by discussing the role of theory and research more generally. We have no strong stance or dog in the fight regarding the veracity of ENA theory, or any other criminological theory for that matter. Sure, we have our preferences, but none of us are trying to protect any theoretical turf here. Nor do we believe that any one theorist nor any one academic discipline "owns" a given theory (see, e.g., the excellent discussion by Tittle 1995). Theories instead belong to the research community. And we need to be willing to refine these theories when new ideas push us to do so, and when new and credible empirical research challenges those ideas that we hold dear.

References

- Andrews, D.A., & Bonta, J. (2010). *The psychology of criminal conduct*. 5th edition. New York: Routledge.
- Barnes, J.C., Beaver, K.M., & Piquero, A.R. (2011). A test of Moffitt's hypothesis of delinquency abstention. *Criminal Justice and Behavior*, 38, 690-709.
- Ellis, L. (2005). A theory explaining biological correlates of criminality. *European Journal of Criminology*, 2, 287-315.
- Ellis, L., Beaver, K.M., & Wright, J.P. (2009). *Handbook of crime correlates*. San Diego, CA: Academic Press.
- Ellis, L., & Hoskin, A.W. (2018). Status of the prenatal androgen hypothesis after two meta-analyses reported little support: A commentary. *Aggression and Violent Behavior*, 42, 29-34.
- Gottfredson, M.R., & Hirschi, T. (1990). *A general theory of crime*. Palo Alto, CA: Stanford University Press.
- Higgins, J.P.T., & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.
- Hirschi, T., & Gottfredson, M.R. (1994). *The generality of deviance*. New York: Routledge.
- Honekopp, J., & Watson, S. (2011). Meta-analysis of the relationship between digit-ratio 2D:4D and aggression. *Personality and Individual Differences*, 51, 381-386.
- Karwowski, M., & Lebuda, I. (2014). Digit ratio predicts eminence of Polish actors. *Personality and Individual Differences*, 64, 30-34.
- McGloin, J.M., Pratt, T.C., & Piquero, A.R. (2006). A life-course analysis of the criminogenic effects of maternal cigarette smoking during pregnancy: A research note on the mediating impact of neuropsychological deficit. *Journal of Research in Crime and Delinquency*, 43, 412-426.
- Merton, R.K. (1973). *The sociology of science: Theoretical and empirical investigations*. Storer, N.W. (ed). Chicago: University of Chicago Press.
- Moffitt, T.E. (1993). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674-701.
- Piquero, A.R., Farrington, D.P., & Blumstein, A. (2003). The criminal career paradigm. *Crime and Justice: A Review of Research*, 30, 359-506.

- Piquero, A.R., Jennings, W.G., & Barnes, J.C. (2012). Violence and criminal careers: A review of the literature from a developmental life-course perspective. *Aggression and Violent Behavior, 17*, 171-179.
- Pratt, T.C., Barnes, J.C., Cullen, F.T., & Turanovic, J.J. (2016). "I suck at everything": Crime, arrest, and the generality of failure. *Deviant Behavior, 37*, 837-851.
- Pratt, T.C., & Cullen, F.T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology, 38*, 931-964.
- Pratt, T.C., Cullen, F.T., Sellers, C.S., Winfree, L.T., Madensen, T.D., Daigle, L.E., Fearn, N.E., & Gau, J.M. (2010). The empirical status of social learning theory: A meta-analysis. *Justice Quarterly, 27*, 765-802.
- Pratt, T.C., McGloin, J.M., & Fearn, N.E. (2006). Maternal cigarette smoking during pregnancy and criminal/deviant behavior: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology, 50*, 672-690.
- Pratt, T.C., Turanovic, J.J., & Cullen, F.T. (2016). Revisiting the criminological consequences of exposure to fetal testosterone: A meta-analysis of the 2D:4D digit ratio. *Criminology, 54*, 587-620.
- Schmidt, F.L., & Hunter, J.E. (2015). *Methods of meta-analysis*. 3rd edition. Thousand Oaks, CA: Sage.
- Tamiya, R., Lee, S.Y., & Ohtake, F. (2012). Second to fourth digit ratio and the sporting success of sumo wrestlers. *Evolution and Human Behavior, 33*, 130-136.
- Tibbetts, S.G., & Piquero, A.R. (1999). The influence of gender, low birth weight, and disadvantaged environment in predicting early onset of offending: A test of Moffitt's interactional hypothesis. *Criminology, 37*, 843-878.
- Tittle, C.R. (1995). *Control balance: Toward a general theory of deviance*. Boulder, CO: Westview Press.
- Turanovic, J.J., Pratt, T.C., & Piquero, A.R. (2017). Exposure to fetal testosterone, aggression, and violent behavior: A meta-analysis of the 2D:4D digit ratio. *Aggression and Violent Behavior, 33*, 51-61.
- Unnever, J.D., Cullen, F.T., & Pratt, T.C. (2003). Parental management, ADHD, and delinquent involvement: Reassessing Gottfredson and Hirschi's general theory. *Justice Quarterly, 20*, 471-500.
- Viechtbauer, W., & Cheung, M.W.L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*, 112-125.

- Voracek, M. (2009). Who wants to believe? Associations between digit ratio (2D:4D) and paranormal and superstitious beliefs. *Personality and Individual differences*, 37, 1533-1541.
- Voracek, M. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D:4D): A comprehensive meta-analysis and critical evaluation of research. *Evolution and Human Behavior*, 35, 430-437.