A model of indirect crowding

William Foley[a], Jonas Radl[a, b]

[a] Universidad Carlos III de Madrid

[b] WZB Berlin Social Science Center

Corresponding author:

William Foley, wfoley@pa.uc3m.es

Department of Social Sciences
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)
España

Acknowledgments

The authors would like to gratefully acknowledge comments from Jan Stuhler, Philip Denter, and Madeline Swarr.

Declaration of conflicting interests

The authors report there are no competing interests to declare.

A model of indirect crowding

Introduction

The introduction of material rewards, penalties for non-compliance, and other such extrinsic incentives has often been suggested as a way of improving the participation of disadvantaged individuals. Such a policy has been discussed, for example, as a method of reducing socioeconomic inequalities in academic achievement (Riener and Wagner, 2022), political participation (Hill, 2006), and the diffusion of low-carbon technologies (Stewart, 2021), among other areas.

Such reward schemes are sometimes controversial as, inter alia, they are argued to corrupt the meaning or nature of the good in question (Sandel, 2020). In this paper, however, we are interested in potentially overlooked consequences of such policies, consequences which arise from the interaction between extrinsic rewards and individuals' intrinsic motivation. In this respect, we contribute to the "crowding" literature, both by distinguishing between two types of crowding and by elaborating a formal model.

First, we distinguish between "direct" and indirect" crowding. Direct crowding refers to an interaction between extrinsic rewards and intrinsic motivation that can lead to a decrease (or, more trivially, an increase) in effort through the weakening (or strengthening) of intrinsic motivation. Direct crowding is the more canonical form of crowding, as represented by Deci's pathbreaking work (1971), as well as Bénabou and Tirole's well-known ensuing contributions (2003, 2006). It has been widely discussed in part because direct crowding is usually concerned with "crowding out" – the seemingly counterintuitive phenomenon that introducing material benefits can actually decrease effort (by depleting or undermining intrinsic motivation). However, as this very counterintuitiveness suggests, the phenomenon is likely to obtain only in particular situations that may not be generally prevalent.

Indirect crowding refers to an interaction between extrinsic rewards and intrinsic motivation which affects the relative degree of effort between individuals of greater and lesser intrinsic motivation, without directly altering intrinsic motivation. This is a phenomenon of more general relevance – particularly for contexts where we expect introducing or increasing material rewards to boost effort. We would not expect, for example, that introducing extrinsic rewards or penalties for academic achievement, political participation, or the adoption of carbon technology would *decrease* participation in education, politics, or environmental

schemes. But it may lead to wider inequalities in the field, as the most academically, politically, or environmentally motivated respond more elastically to incentives.

We therefore elaborate a simple model of *indirect* crowding, building both on social and cognitive scientific modelling of effort and motivation. We show that modifying typical models of effort through including decreasing marginal rewards (alongside increasing marginal benefit) can lead to both indirect crowding in or out, but *only* if the intrinsic and extrinsic benefits of effort are additive. If the benefit of effort is the product of intrinsic and extrinsic benefit, then effort can only be indirectly crowded *in*.

Substantively speaking, assuming that total benefit of effort is an additive function of intrinsic and extrinsic benefit, indirect crowding *out* occurs when optimal effort is a concave function of extrinsic reward, making it increasingly less attractive to raise effort. While all individuals will increase effort as a function of rewards, less intrinsically motivated individuals will increase effort at a greater rate, allowing them to "catch up" with the intrinsically motivated. Conversely, indirect crowding *in* occurs when optimal effort is a convex function of extrinsic reward, making it increasingly attractive to raise effort. The highly motivated will speed away from the unmotivated.

The model has general implications for the crowding and motivation literatures. Policy interventions which aim to boost participation or achievement may inadvertently widen inequality, depending on the relative curvature of marginal cost and benefit curves, and whether total benefit is an additive or multiplicative aggregate of intrinsic and extrinsic benefit. It follows, for example, providing material benefits for political participation should also be coupled with interventions to boost political interest among the less interested.

More broadly, this study aims to showcase the relevance of motivational crowding for sociological thinking. By proposing a novel fourfold typology of crowding effects, it offers a useful heuristic that sheds new light on an established literature characterized by competing model formulations and inconsistent evidence. In addition, by stressing the heterogeneity and inertia of people's subjective orientations and preferences, our model emphasizes the role of population diversity that is sometimes neglected in economic or psychological approaches to motivation crowding.

Crowding effects

Exerting effort is costly since the human organism can only perform a limited number of tasks simultaneously. Neuroscientists and cognitive psychologists therefore theorize that effort feels "aversive" because it is the body's way of signaling that important cognitive functions are being engaged which could (potentially) be redirected to more fruitful purposes (Kurzban et al., 2013; Shenhav et al., 2017). It follows that it must benefit an agent to exert effort on a particular task, and that the degree of effort will be positively correlated with the degree of benefit.

Benefit can be decomposed into extrinsic and intrinsic benefit. Extrinsic benefit refers to external rewards such as money or status, which are contingently related to performance of a task. But exerting effort on a task may also be intrinsically beneficial – an individual may find the achievement of a goal purposeful (e.g. volunteering) or the performance of a task satisfying (e.g. sudoku) in its own right (Deci, 1971). It is logical that the greater the extrinsic (intrinsic) benefit an individual derives from a task, the more extrinsically (intrinsically) motivated they will be to do it – and the more effort they will exert (Kuvaas et al., 2017).

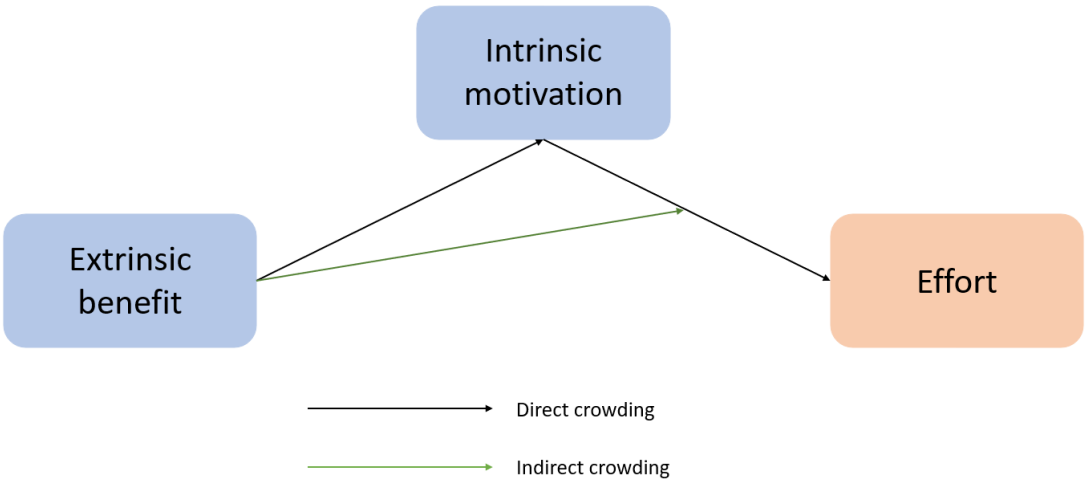
However, the provision of extrinsic benefits can also lead to the "crowding out" of intrinsic motivation, potentially even to the point of reducing overall effort (Frey and Jegen, 2001). That is to say, extrinsic rewards may deplete intrinsic motivation *and/or* weaken its association with effort. It is also possible that increasing extrinsic rewards may "crowd in" intrinsic motivation by enhancing intrinsic motivation or its association with effort (Frey and Jegen, 2001), though this crowding type has attracted less attention.

The "crowding" moniker covers a variety of phenomena and mechanisms. We identify a distinction between direct and indirect crowding. Direct crowding can be considered the "canonical" mechanism of dependence of intrinsic motivation on extrinsic rewards. Increases in intrinsic rewards either reduce or increase intrinsic motivation, which in turn may either reduce or increase absolute effort levels. This type of phenomenon is the one discussed by Deci (1971), who theorized that intrinsic motivation decreased after an extrinsic reward was introduced. This type of direct crowding out – or "undermining effect" – has generated much debate throughout the social science literature in recent decades, partly because it seems to violate the economistic axiom that "introducing extrinsic incentives cannot lower effort levels" (Kreps, 1997: 360).

The second type of crowding is indirect crowding. This crowding mechanism does not stipulate a direct effect of extrinsic rewards on the level of intrinsic motivation. Rather, rewards modulate the association between intrinsic motivation and effort. In the context of a population of individuals with heterogeneous levels of intrinsic motivation, indirect crowding entails an effect of extrinsic benefits on the relative degree of effort between individuals with greater and lesser intrinsic motivation, without directly modifying their levels of motivation.

The distinction between the two is summarized in Figure 1. Direct crowding operates through the black arrows running from extrinsic benefit to intrinsic motivation to effort. Indirect crowding operates through the green arrow which modifies the arrow linking intrinsic motivation and effort. (Since the figure is illustrating *crowding* effects, the direct arrow between extrinsic benefit and effort is omitted).

Figure 1. Illustration of direct versus indirect crowding



As an example of indirect crowding, consider the case of a firm where some of the workers are highly intrinsically motivated and others are not, for instance because they vary in their personal identification with the firms' business model. Standard principle-agent theory would suppose that raising wages should increase effort – indeed this is the logic behind efficiency wages (Murphy and Topel, 1990). Two scenarios may then occur. In the first, highly motivated workers increase their effort by X%, but workers with low motivation increase their effort by more than X%. Hence, the effort gap narrows, as does the correlation between effort and intrinsic motivation – intrinsic motivation is crowded *out*. In the second scenario, highly motivated workers increase their effort by X%, but workers with low motivation increase their effort by less than X%. Hence, the effort gap increases, as does the correlation

between effort and intrinsic motivation – intrinsic motivation is crowded *in*. (There is also the "knife-edge" scenario of a proportionally equal increase across worker types). Table 1 summarizes the distinction between direct and indirect crowding, for both crowding in and out phenomena.

Table 1. Conceptual overview of crowding effects

	Crowding in	**Crowding out**
Direct crowding	External rewards directly increase intrinsic motivation, which leads to greater effort. *Example*: Non-monetary publicly awarded rewards boost motivation to perform by promoting virtuous attitude (Bruni et al., 2020).	External rewards directly decrease intrinsic motivation, which leads to less effort. *Example*: Money offered for blood donations reduces motivation to donate because former donators worry they no longer signal altruism (Bénabou and Tirole, 2006).
Indirect crowding	External rewards strengthen the effect of intrinsic motivation on effort. *Example*: Performance bonus payments reduce effort among less intrinsically motivated workers, increasing the effort differential with highly motivated workers (Sliwka, 2007).	External rewards weaken the effect of intrinsic motivation on effort. *Example*: Increased material rewards lead to relatively higher effort increases among individuals with low intrinsic motivation (Dorner and Lancsar, 2023).

Most work on motivation crowding has focused on direct crowding. Deci's seminal "cognitive evaluation theory" (Deci, 1971, 1976) proposed that extrinsic benefits crowd out intrinsic motivation by shifting the locus of control from internal (agents believing they themselves are in control) to external (agents believing they are not in control). Another influential early theory was the "overjustification hypothesis" (Lepper et al., 1973), which stipulates that extrinsic rewards may signal to an individual that they are pursuing a goal for

the sake of the extrinsic reward rather than an intrinsic interest, causing them to downplay their own intrinsic motivation.

Subsequent theories have tended to focus on the informational content of the extrinsic reward – whether as a self-signal (as in overjustification theory), or as a signal of others' private information or otherwise opaque/unknown beliefs. Bénabou and Tirole (2003) model a framework where a principal has private information on the costliness of a task to an agent. A high reward offered by the principal signals to the agent that the task is not intrinsically interesting. Bolle and Otto analyze a similar setting where extrinsic rewards signal to the agent they had previously overestimated the task's value, causing them to scale down their effort. (Bolle and Otto, 2010). In another paper, Bénabou and Tirole (2006) examine a setting where the introduction of extrinsic benefits causes an agent to worry that performing an altruistic task, which had previously been unincentivized, would now send the wrong signal to third parties regarding the agent's own intentions (i.e. that the agent is motivated by monetary gain rather than altruism).

Other theories of direct crowding (out) include Frey's (1994) argument that the introduction of extrinsic rewards may cause an agent to reinterpret their relationship to a principal in utilitarian terms, diminishing their intrinsic motivation. Kreps has argued (1997) that extrinsic benefits – especially when used to incentivize ambiguous and multifaceted activities – may cause agents to focus on the wrong component of the activity, at the expense of the more intrinsically motivated component.

Less prominent are theories of (direct) crowding *in*. It is reasonable to suppose that crowding in may occur through the same mechanisms as crowding out, but working in the reverse direction. For example, symmetric to Bolle and Otto's mechanism (2010), providing high rewards for a task may signal to agents that the task was previously *under*valued. And, symmetric to Bénabou and Tirole's model (2006), concerns about image signaling may improve contribution to a collective good (Weibel et al., 2014). Indeed, Bruni and colleagues (2020) predict, and experimentally verify, that – unlike money – non-monetary prizes crowd in intrinsic motivation by enhancing agents' motivations to appear altruistic. Awards may also enhance loyalty between principals and agents, boosting the intrinsic motivation of the latter (Frey and Gallus, 2016).

Direct rowding in may also occur when an extrinsic reward is delivered in such a way as to acknowledge or enhance an individual's sense of agency, by acknowledging the agent's

motivation (Frey, 1994), or allowing the agent to choose a performance-contingent contract (Bonner and Sprinkle, 2002).

Formally modelled theories of *indirect* crowding – where extrinsic benefits interact with, but do not decrease or amplify intrinsic motivation – are somewhat less common, at least in the literature that explicitly addresses crowding phenomena. Sliwka (2007) analyses a setting where a subgroup of the workforce are "conformists" (i.e. not intrinsically motivated) and another "fairness" orientated (i.e. intrinsically motivated). The latter only provide more than the "selfish" level of effort if they believe the median colleague is intrinsically motivated. Hence, the introduction of effort-contingent bonuses signal to conformists that the median worker is not intrinsically motivated, causing them to reduce effort.

The direct crowding mechanism is more headline-grabbing because the proposition that introducing extrinsic benefits would reduce effort seems counterintuitive. (The focus of theories of direct crowding tends to be on the intriguing crowding *out* phenomenon). But indeed the intuition that is being countered is itself intuitive for a reason – direct crowding out is unlikely to be a generalized phenomenon because it "hinges on several assumptions unlikely to hold in many nonlaboratory contexts" (Cerasoli et al., 2014: 2). For example, while assumptions vary across models and theories, direct crowding mechanisms logically imply that individuals' intrinsic preferences are endogenous to extrinsic reward, a relationship which will hold only in a subset of contexts.

On the other hand, as we will show, indirect crowding entails more plausible scope conditions. Besides some particular formal assumptions about functional form, the type of indirect crowding studied here only requires a small number of highly plausible assumptions to generate indirect crowding in *and* out effects: namely, that at least one of the benefit and cost curves exhibits curvature (i.e. marginally increasing and/or decreasing costs), that the curvatures are not identical, and that the benefit and cost curves are power functions.

The less restrictive conditions for indirect crowding suggest that this phenomenon is probably of greater prevalence. Indeed, one meta-analysis of the crowding work has come to the conclusion that there is little evidence for (direct) crowding out (Cameron and Pierce, 1994). Though it should be said that a subsequent meta-analysis, conducted by proponents of crowding theory, argued that Cameron and Pierce's method was flawed and found evidence in favor of (direct) crowding (Deci et al., 1999). Either way, the greater prevalence of indirect crowding arguably lends it greater sociological import.

Indeed, without specifically invoking the crowing label, large swathes of the empirical social scientific literature have tackled questions of motivation and incentives that best fall under the rubric of indirect crowding – from gender inequality in academic performance (Levitt et al., 2016), to the productivity of public sector employees (Belle and Cantarelli, 2015), to work quality in casual labor markets (Rogstadius et al., 2011). The majority of the 150+ papers reviewed in Cerasoli, Nicklin, and Ford's meta-analysis (2014) would fall into the indirect crowding category. Further, papers focused on studying direct crowding also often find indirect crowding phenomena as well (Dorner and Lancsar, 2023).

Model

A typical formalization of the utility of effort (Bénabou and Tirole, 2003; Breen, 1999; DellaVigna and Pope, 2018; Frey, 1994; Piketty, 1995), expressed in general terms, will be the following:

$$u(\varepsilon) = f(\varepsilon) - g(\varepsilon) \quad (1)$$

Where ε represents effort, $f(\varepsilon)$ is the benefit of effort and $g(\varepsilon)$ is the cost of effort. In literature engaging with crowding, $f(\varepsilon)$ is often specified so that the benefit of effort can be decomposed into intrinsic benefit, i, and extrinsic reward, r. For example, (substituting our own notation) dellaVigna and Pope (2018) set $f(\varepsilon) = (r + i)\varepsilon$. Decomposing the benefit of effort into the *sum* of intrinsic and extrinsic benefit is a plausible formalization from a neuroscientific perspective, although it is not the only possible one (Inzlicht et al., 2018).

Typically the cost of effort is held to be convex and increasing on effort, such that $g'(\varepsilon) > 0$ and $g''(\varepsilon) > 0$ (DellaVigna and Pope, 2018; Frey, 1994; Sliwka, 2007). More particularly, $g(\varepsilon)$ is usually modelled as a power function (DellaVigna and Pope, 2018; James, 2005; Piketty, 1995; Sliwka, 2007), of the form $g(\varepsilon) = c\varepsilon^m$, where naturally $m > 1$. This functional form has empirical support (DellaVigna and Pope, 2018).

The benefit of effort, $f(\varepsilon)$ is often set to be linear in effort (Breen, 1999; James, 2005; Piketty, 1995; Sliwka, 2007). We assume, however, that the benefit function is also a power function: $f(\varepsilon) = (r + i)\varepsilon^n$, where i represent intrinisc benefit and r extrinsic reward. We also assume, though it is not strictly necessary for the main result, that there are diminishing marginal benefits to effort. Hence, $n \in (0, 1)$, and hence $f'(\varepsilon) > 0$ and $f''(\varepsilon) < 0$.

The restrictions on the values of n and m imply that $m \neq n$. While it is not strictly necessary for our result that there be decreasing marginal returns and increasing marginal benefit, it is

necessary that $m \neq n$, otherwise there is no interior solution. This requirement also implies at least one of the benefit and cost curves is nonlinear.

We hence rewrite (1) as:

$$u(\varepsilon) = (r + i)\varepsilon^n - c\varepsilon^m \quad (2)$$

Where $r > 0, r + i > 0, c > 0$ and $\varepsilon \in [0, k]$ where k is some upper limit denoting the maximum level of effort an agent can expend – substantively equivalent to the maximum bandwith of an individual's execution function (Kurzban et al., 2013). For the sake of elucidating our central point we assume here that the optimal level of effort is always less than k, so the constraint does not bind and an interior solution is possible. Indeed, effort models do not typically assume an upper bound on ε (for example DellaVigna and Pope 2018; Piketty 1995; Sliwka 2007).

A utility-maximizing individual exerts effort level $\bar{\varepsilon}$, the value of ε that maximises $u(\varepsilon)$:

$$\bar{\varepsilon} = \left(\frac{n(r + i)}{mc}\right)^{(m-n)^{-1}} \quad (3)$$

Crucial to our account is that $\bar{\varepsilon}$ is a strictly concave function on r if $m - n > 1$, and strictly convex if < 1. This leads to our first proposition:

Proposition 1. If $m - n > 1$, increasing the extrinsic reward r, leads to indirect crowding *out*. If $m - n < 1$, increasing the extrinsic reward r, leads to indirect crowding *in*.

Indirect crowding occurs when a change in the extrinsic reward changes the relative effort levels between highly intrinsically motivated and lesser intrinsically motivated subgroups in a population. We consider a setting with two individuals, l and h, who differ only in terms of their intrinsic motivation, i. Their parameter values for m, n, and c are equal, and they face a common piece rate of r for a unit of effort, ε, that is exerted. Their utility depends only on their own effort – i.e. the setting is not competitive so the others' effort does not enter into their utility function. Ceteris paribus, h is more intrinsically motivated, and thus should provide more effort than l. However, if effort is indirectly crowded out as extrinsic rewards increase, then the gap in optimal effort between h and l should decline (i.e. intrinsic motivation matters "less") – and vice versa if effort is crowded in (i.e. intrinsic motivation matters "more").

Modifying (3), define $\bar{\varepsilon}_j(\mathrm{r}) = \left(\frac{n(r+ai)}{mc}\right)^{(m-n)^{-1}}$ to be the optimal level of effort for individual j. The new parameter a is a multiplier on intrinsic benefit i. The value of a is greater for more intrinsically motivated individuals. For the individual with low intrinsic motivation l, we set $a = 1$, and for the individual with high intrinsic motivation h, we set $a > 1$. Let there be a function $\delta(r)$, which gives the difference in optimal effort between l and h as a function of r:

$$\delta(r) = \bar{\varepsilon}_h(r) - \bar{\varepsilon}_l(r) \tag{4}$$

If $\delta'(r) > 0$ then the gap in effort as a function of r increases between high and low motivated individuals, and if $\delta'(r) < 0$, it decreases:

$$\delta'(r) = \frac{n}{(m-n)mc}\left(\frac{n(r+ai)}{mc}\right)^{\frac{1-(m-n)}{m-n}} - \frac{n}{(m-n)mc}\left(\frac{n(r+i)}{mc}\right)^{\frac{1-(m-n)}{m-n}} \tag{5}$$

$\delta'(r)$ is greater than zero if:

$$\left(\frac{n(r+ai)}{mc}\right)^{\frac{1-(m-n)}{m-n}} > \left(\frac{n(r+i)}{mc}\right)^{\frac{1-(m-n)}{m-n}} \tag{6}$$

This inequality holds if $m - n < 1$, i.e. if optimal effort, $\bar{\varepsilon}$, is a strictly convex function of r. In this case, indirect crowding *in* occurs, and the highly motivated individual accelerates away from the lesser motivated one.

On the other hand, the inequality in (6) reverses direction (meaning $\delta'(r)$ is less than zero) if $m - n > 1$. In this case, $\bar{\varepsilon}$, is a strictly concave function on r. Indirect crowding *out* occurs, and the gap between the highly motivated individual and the lesser motivated one shrinks.

The gap in effort between individuals with high and low intrinsic motivation, $\delta(r)$, when total benefit is an additive function of intrinsic and extrinsic benefit (i.e. $f(\varepsilon) = (r+i)\varepsilon$) is graphed in Figure 2a. The blue line represents a situation where $\bar{\varepsilon}$ is strictly convex on r meaning the gap in effort between highly and lesser motivated individuals grows as a function of r. The red line represents a situation where $\bar{\varepsilon}$ is strictly concave on r, meaning the the gap in effort between highly and lesser motivated individuals declines as a function of r.

While models of effort typically treat the benefit of effort as an additive function of intrinsic and extrinsic benefit, cognitive scientists have also discussed the possibility that total benefit

could be the product of intrinsic and extrinsic benefit (Inzlicht et al., 2018) – i.e the benefit of effort $f(\varepsilon) = (ri)\varepsilon^n$. If this is the case, there can be no indirect crowding *out* of effort, only crowding *in*, i.e. the highly motivated accelerate away from the lesser motivated.

Proposition 2. If the benefit of effort function $f(\varepsilon) = (ri)\varepsilon^n$, there can only be indirect crowding *in*.

We rewrite equation (3):

$$\bar{\varepsilon}_j(\mathrm{r}) = \left(\frac{nrai}{mc}\right)^{(m-n)^{-1}} \tag{7}$$

Where $a = 1$ if $j = l$, and $a > 1 \;\; j = h$. Then we substitute (7) into (4), and take the derivative:

$$\delta'(r) = \frac{nai}{(m-n)mc}\left(\frac{nrai}{mc}\right)^{\frac{1-(m-n)}{m-n}} - \frac{ni}{(m-n)mc}\left(\frac{nri}{mc}\right)^{\frac{1-(m-n)}{m-n}} \tag{8}$$

If indirect crowding *out* is to occur, then $\delta'(r) < 0$ and the following inequality must hold:

$$a\left(\frac{nrai}{mc}\right)^{\frac{1+(m-n)}{m-n}} < \left(\frac{nri}{mc}\right)^{\frac{1+(m-n)}{m-n}} \tag{9}$$

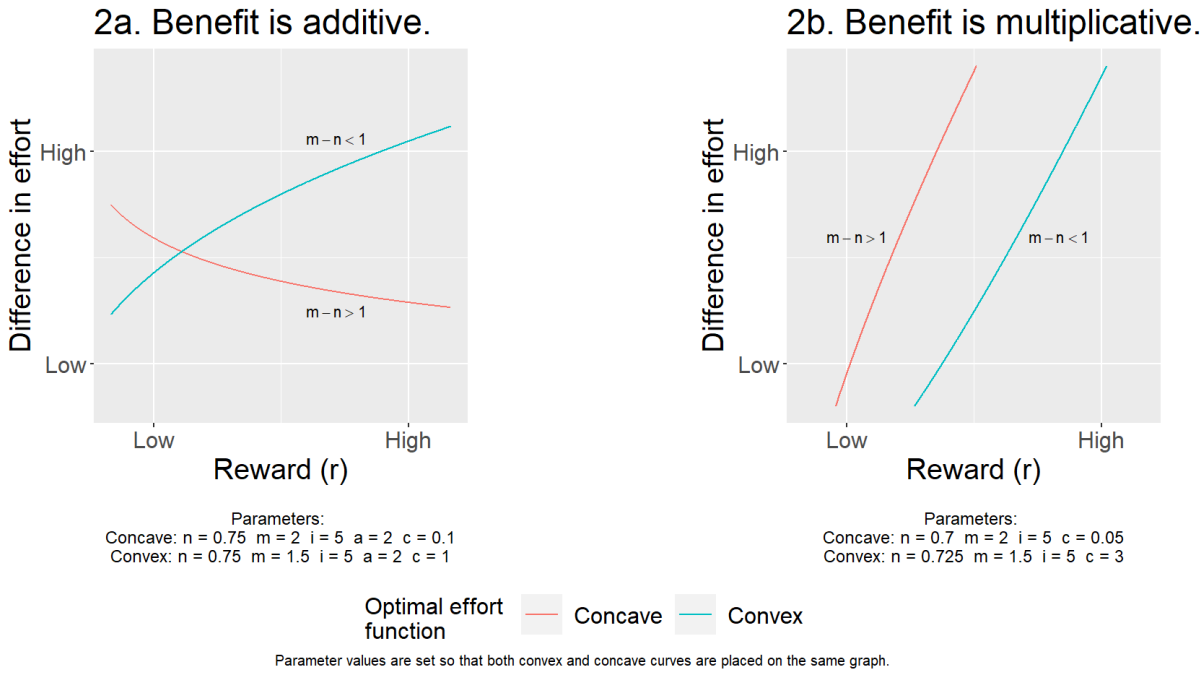
Or:

$$a < \left(\frac{1}{a}\right)^{\frac{1+(m-n)}{m-n}} \tag{10}$$

By assumption $m - n \in (0, \infty)$. Let's consider how the exponent of the RHS of (10) behaves as $m - n$ (monotonically) traverses this interval. Approaching the lower bound: $\lim_{(m-n)\to 0} \frac{1+(m-n)}{m-n} \to \infty$. Therefore, since $a > 1$, the quantity on the RHS of (10) approaches zero as $m - n$ approaches its lower bound of 0. As $m - n$ increases without bound $\lim_{(m-n)\to\infty} \frac{1+(m-n)}{m-n} \to 1$. Therefore, the quantity on the RHS of (10) approaches a maximum of $\frac{1}{a}$ as $m - n$ approaches ∞. Hence, the maximum value that the RHS of (10) can have over the feasible values of $m - n$ is $\frac{1}{a}$. And since $a > 1$, the inequality in (10) can never obtain. Hence crowding *out* will never occur. In fact, even if we let $m - n < 0$,

$\frac{1+(m-n)}{m-n}$ can still only reaching a maximum value of 1. There is no real value $m-n$ can take that will satisfy (10).

For the sake of illustration, the gap in effort between someone with high and low extrinsic motivation, $\delta(r)$, is graphed for the multiplicative case in Figure 2b.



Discussion

In the context of a population of individuals with heterogeneous levels of motivation, *indirect* crowding affects the relative degree of effort between individuals with greater and lesser motivation without directly altering their level of motivation.

While indirect crowding is less headline-hogging than its direct counterpart, it is likely to be a phenomenon of more general interest when considering the interaction between extrinsic benefits and motivation. For indirect crowding to obtain, one does not have to assume that a shift in rewards changes their personal interests and preferences.

In this paper we present a simple model of indirect crowding. The model stipulates that, assuming the benefit of effort is an additive function of intrinsic and extrinsic benefit, optimal effort will be a concave function of extrinsic rewards when the ratio of marginal costs to benefit rises very steeply. Every unit increase in extrinsic reward will yield marginally declining increases in optimal effort, meaning that a less intrinsically motivated individual

will "catch up" with their more intrinsically motivated counterpart – a "crowding out" effect. Effects consistent with this mechanism are visible in empirical explorations of motivation crowding, where individuals with lower intrinsic motivation show greater improvements in effort in response to increased rewards (Dorner and Lancsar, 2023).

The other possibility demonstrated by the model is that intrinsic motivation is "crowded in". If the ratio of marginal cost to benefit does not rise too quickly, optimal effort will be a *convex* function of extrinsic rewards. As rewards increase, the highly motivated accelerate away from the low motivated as a function of rewards. According to Cerasoli and colleagues' meta-analysis (2014), most studies show that extrinsic incentives *strengthen* the association between intrinsic motivation and effort, though there are substantial heterogeneities across studies. This suggests that, at least in the sample of studies investigated by Cerasoli et al, indirect crowding in dominates crowding out.

These results are conditional on the benefit of effort being an *additive* function of intrinsic and extrinsic benefit – i.e. $f(\varepsilon) = (r + i)\varepsilon^n$, as assumed, for example, in dellaVigna and Pope (2018). If the function is a *product* of intrinsic and extrinsic benefit – i.e. $f(\varepsilon) = (ri)\varepsilon^n$ – then effort is only crowded in by increasing extrinsic reward. If benefit is more often than not an additive function, this could also (partly) explain the findings of Cerasoli et al (2014).

The model presented here is simple. It builds on typical models of effort and incorporates the highly plausible assumption of decreasing marginal benefit and increasing marginal return. Indeed, the assumption of decreasing marginal benefits is not strictly required – benefit could be linear in effort, as assumed by Bénabou and Tirole (2003) for example. At the same time, the model eschews the highly contingent assumption, a feature of direct crowding models, that extrinsic benefits alter intrinsic motivation. The assumption presupposes that (often small) changes in material rewards can significantly alter people's preferences. But this assumption can be questioned on the grounds of the inconsistent and heterogeneous effects often found, for instance, in survey experiments assessing information effects on policy preferences (Fernández et al., 2023; Finseraas et al., 2017), attitudes toward inequality (Engelhardt and Wagener, 2018), or educational choices (Ballarino et al., 2022).

The model naturally has several limitations. First, while indirect crowding stresses population heterogeneity in intrinsic motivation, the model doesn't symmetrically feature heterogeneity in responsiveness of extrinsic motivation. Yet, there are obviously differences in the value individuals place on material rewards (Inglehart and Baker, 2000; Moors and Vermunt,

2007). Hence, formally incorporating heterogeneous sensitivity to extrinsic reward would be one possible extension of the model. Second, the results are reliant on the assumption that both the benefit and effort functions are power functions. There is some empirical evidence for this, provided by della Vigna and Pope's paper (2018), but other functional specifications are possible.

Nonetheless, the model is rather general and based on typical models of effort, and hence has significant implications for the design of motivation schemes. Returning to the examples given at the beginning, introducing or increasing material incentives to boost the participation of (often socioeconomically disadvantaged) lesser motivated individuals may have unexpected consequences. For example, the introduction of immediate and direct material benefits for educational success (e.g. "cash for grades") may backfire and increase educational inequality if the marginal cost to benefit ratio is not too steep, or if total benefit is the product of intrinsic and extrinsic benefit. Indeed, there is evidence that introducing material benefits increases inequality in academic outcomes (Leuven et al., 2010). Hence, policymakers who seek to boost participation in some good, and who do not wish to simultaneously increase inequality, should couple extrinsic rewards with schemes to boost intrinsic motivation and should consider the structure of cost and benefit curves.

References

Ballarino G, Filippin A, Abbiati G, Argentin G, Barone C, and Schizzerotto A (2022) The effects of an information campaign beyond university enrolment: A large-scale field experiment on the choices of high school students. *Economics of Education Review* 91: 102308.

Belle N and Cantarelli P (2015) Monetary Incentives, Motivation, and Job Effort in the Public Sector: An Experimental Study With Italian Government Executives. *Review of Public Personnel Administration* 35(2): 99–123.

Bénabou R and Tirole J (2003) Intrinsic and Extrinsic Motivation. *The Review of Economic Studies* 70(3): 489–520.

Bénabou R and Tirole J (2006) Incentives and Prosocial Behavior. *The American Economic Review* 96(5): 1652–1678.

Bolle F and Otto PE (2010) A price is a signal: On intrinsic motivation, crowding-out, and crowding-in. *Kyklos* 63(1): 9–22.

Bonner SE and Sprinkle GB (2002) The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27(4): 303–345.

Breen R (1999) Beliefs, rational choice and Bayesian learning. *Rationality and Society* 11(4): 463–479.

Bruni L, Pelligra V, Reggiani T, and Rizzolli M (2020) The Pied Piper: Prizes, Incentives, and Motivation Crowding-in. *Journal of Business Ethics* 166(3): 643–658.

Cameron J and Pierce WD (1994) Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Review of Educational Research* 64(3): 363–423.

Cerasoli CP, Nicklin JM and Ford MT (2014) Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin* 140(4): 980–1008.

Deci EL (1971) Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18: 105–115.

Deci EL (1976) Notes on the theory and metatheory of intrinsic motivation. *Organizational Behavior and Human Performance* 15: 130–145.

Deci EL, Koestner R and Ryan RM (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125: 627–668.

DellaVigna S and Pope D (2018) What Motivates Effort? Evidence and Expert Forecasts. *The Review of Economic Studies* 85(2): 1029–1069.

Dorner Z and Lancsar E (2023) Don't pay the highly motivated too much. *Journal of Behavioral and Experimental Economics* 103: 101972.

Engelhardt C and Wagener A (2018) What do Germans think and know about income inequality? A survey experiment. *Socio-Economic Review* 16(4): 743–767.

Fernández JJ, García-Albacete G, Jaime-Castillo AM, and Radl J (2023) Priming or learning? The influence of pension policy information on individual preferences in Germany, Spain and the United States. *Journal of European Social Policy* 33(3): 337–352.

Finseraas H, Jakobsson N and Svensson M (2017) Do knowledge gains from public information campaigns persist over time? Results from a survey experiment on the Norwegian pension reform. *Journal of Pension Economics & Finance* 16(1): 108–117.

Frey BS (1994) How Intrinsic Motivation is Crowded out and in. *Rationality and Society* 6(3): 334–352.

Frey BS and Gallus J (2016) Honors: A rational choice analysis of award bestowals. *Rationality and Society* 28(3): 255–269.

Frey BS and Jegen R (2001) Motivation Crowding Theory. *Journal of Economic Surveys* 15(5): 589–611.

Hill L (2006) Low Voter Turnout in the United States: Is Compulsory Voting a Viable Solution? *Journal of Theoretical Politics* 18(2): 207–232.

Inglehart R and Baker WE (2000) Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review* 65(1): 19–51.

Inzlicht M, Shenhav A and Olivola CY (2018) The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences* 22(4): 337–349.

James HS (2005) Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance. *Journal of Economic Psychology* 26(4): 549–566.

Kreps DM (1997) Intrinsic Motivation and Extrinsic Incentives. *The American Economic Review* 87(2): 359–364.

Kurzban R, Duckworth A, Kable JW and Myers J (2013) An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences* 36(6): 661–679.

Kuvaas B, Buch R, Weibel A, Dysvik A and Nerstad CGL (2017) Do intrinsic and extrinsic motivation relate differently to employee outcomes? *Journal of Economic Psychology* 61: 244–258.

Lepper MR, Greene D and Nisbett RE (1973) Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology* 28(1): 129–137.

Leuven E, Oosterbeek H and van der Klaauw B (2010) The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment. *Journal of the European Economic Association* 8(6): 1243–1265.

Levitt SD, List JA, Neckermann S and Sadoff S (2016) The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy* 8(4): 183–219.

Moors G and Vermunt J (2007) Heterogeneity in Post-materialist Value Priorities. Evidence from a Latent Class Discrete Choice Approach. *European Sociological Review* 23(5): 631–648.

Murphy KM and Topel RH (1990) Efficiency Wages Reconsidered: Theory and Evidence. In: Weiss Y and Fishelson G (eds) *Advances in the Theory and Measurement of Unemployment*. London: Palgrave Macmillan UK, pp. 204–240. Available at: https://doi.org/10.1007/978-1-349-10688-2_8 (accessed 14 September 2023).

Piketty T (1995) Social Mobility and Redistributive Politics. *The Quarterly Journal of Economics* 110(3): 551–584.

Riener G and Wagner V (2022) Non-monetary rewards in education. *Educational Psychology* 42(2): 222–239.

Rogstadius J, Kostakos V, Kittur A, Smus B, Laredo J and Vukovic M (2011) An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. *Proceedings of the International AAAI Conference on Web and Social Media* 5(1). 1: 321–328.

Sandel MJ (2020) *The Tyranny of Merit: What's Become of the Common Good?* London: Penguin UK.

Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD and Botvinick MM (2017) Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience* 40(1): 99–124.

Sliwka D (2007) Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes. *The American Economic Review* 97(3). American Economic Association: 999–1012.

Stewart F (2021) All for sun, sun for all: Can community energy help to overcome socioeconomic inequalities in low-carbon technology subsidies? *Energy Policy* 157: 112512.

Weibel A, Wiemann M and Osterloh M (2014) A behavioral economics perspective on the overjustification effect: Crowding-in and crowding-out of intrinsic motivation. In: Gagné M (ed.) *The Oxford Handbook of Work Engagement, Motivation, and Self-Determination Theory.* Oxford University Press New York, NY, pp. 72–84.