In Defense of Elster's Mechanisms
Kimberly Chuang, University of Michigan

Abstract

Mechanisms are a species of causal explanation that apply in situations where there is unavoidable indeterminacy. In this paper, I defend Jon Elster's account of mechanisms against Johannes Persson's recently published critiques. Persson claims to have identified a dilemma to which Elster is committed. Elster stipulates that mechanisms must be indeterminate in either their triggering conditions or their consequences. Persson argues that we can resolve the indeterminacies in certain mechanisms. Upon doing so, we no longer, by definition, have mechanisms. At the same time, these resolved mechanisms remain only explanatorily local, and so fall short of the explanatory strength of laws. Persson concludes that by Elster's account, eliminating the indeterminacies of mechanisms actually leaves us at an explanatory deficit: we are left with something that is no longer a mechanism, but that still falls short of law-like explanatory strength. I counter that in his argument, Persson has overlooked the distinction between improved explanatory strength in a purely local sense and improved explanatory strength in a generalized sense. In addition, Persson has also overlooked the distinction between individual instances, or applications of mechanisms, and mechanisms themselves. I conclude that Persson has not, in fact, discovered any dilemma in Elster. Persson's argument occurs at the level of mere applications of mechanisms. His challenges to Elster pertain to improved explanatory strength in a purely local sense. What he needed to have done, in order to complete his case against Elster, was show that the alleged shortcomings of Elster's mechanistic account occurred at the level of mechanisms themselves, and pertained to improved explanatory strength in a generalized sense.

In this paper I defend Jon Elster's account of explanatory mechanisms against Johannes Persson's recently published critiques. Elster argues that explanatory mechanisms must be indeterminate in some respect: for a type A mechanism, we are unable to know the conditions that will trigger it; for a type B mechanism, we are unable to know the consequences, or net effect, that will ensue from it. Persson believes that the deficiencies in Elster's account are brought to the fore whenever we resolve the respective indeterminacies of mechanisms; that is, when we come to know the triggering conditions for a type A mechanism, or the consequences that ensue from a type B mechanism. As soon as we do, Persson argues, we must have something more explanatorily powerful than a mechanism. Since mechanisms are characterized by their indeterminacy, we would no longer have, by definition, a mechanism. Nor do we get laws when we clarify the indeterminacies of a mechanism – these clarified mechanisms will still remain short of the explanatory strength needed for laws. Consequently, Persson claims that Elster is on the horns of a dilemma. Whenever we are able to account for mechanistic indeterminacies, we no longer have mechanisms, but we also do not have laws. By resolving mechanistic indeterminacies — something we think would *strengthen* our ability to explanatorily account for things — we have actually lost explanatory ability.

My defense of Elster will be centered on the claim that Persson's argument inappropriately overlooks the difference between *applications* of mechanisms rather than *mechanisms themselves*, and the difference between explanatory strength as achieved through *specificity*, rather than explanatory strength achieved through *generalization*.

Throughout the paper, I will refer to bundles of triggering conditions and triggered mechanism, or mechanism and determinate consequences — what Persson intends as more explanatorily powerful than ordinary mechanisms — interchangeably as 'elucidated mechanisms,' 'disambiguated mechanisms,' or 'mechanisms whose indeterminacies have been resolved.'

The paper's first section will elaborate on the mechanisms with which I am working. The elaboration is extensive so as to give a better picture of the sophistication and utility of mechanistic explanation on Elster's account. The second section will be a preliminary discussion of the main ways laws and mechanisms differ. An exegesis of Persson's argument will be given in the third section, followed by my response — a defense of Elster — in the fourth section.

1. Elster on explanation

Causal explanation, quite generally, accounts for why something happened. Mechanisms, on which the bulk of this paper will be focused, are a species of causal explanation. Causal explanations do not merely answer 'why' questions. For example, when we ordinarily ask why a publication failed to cite a central piece of literature in its field, a sufficient answer may be that the author overlooked it (Elster 2007, 28). For Elster, this sort of "negative explanans"— which cites a nonevent — is inadequate for a causal explanation (13). Instead, what we want in an answer, and, likewise what would suffice for an explanation, are responses articulated in terms of agents and their actions — a sort of methodological individualism – when agents are involved in the situations under analysis. Causal explanation should ideally take the form of positive events – decisions that agents make they make or psychological facts about them that undergird those decisions they make (14).

For Elster, correlative explanation and statistical explanation do not, for Elster, constitute satisfactory causal explanation. While we might be in a position to conclude that an event of one kind is always accompanied by an event of a second kind, we can never on the basis of this correlation alone, conclude a causal relationship. The correlated events might be both consequences of a third, unperceived cause. And we certainly cannot determine the direction of causality by looking at correlated events. Statistical explanation fares similarly. What we ideally want for explanation are events at the level of individual decision-making. We would make a mistake in applying statistical explanations —which generalize about groups — to individual cases. Furthermore, statistical explanation presupposes the structure of causal explanation (27). When we are interested in why a particular phenomenon, such as democratization, takes place, we may find several accompanying correlates – the degree of industrialization, the equatorial latitude, and so forth. In determining which correlates can figure in a statistical explanation, we make a judgment about which correlate is more likely to be causally explanatory.

Mechanisms account for why things happened. They do this in a slightly different way from ordinary causal explanation. The general requirements for good causal explanation, however,

still apply to mechanisms. Once we figure out whether something is causally responsible for our phenomenon of interest, we can then look for *how* it is responsible. Explaining *how* something is responsible is accomplished by "showing it to be an instance of a general causal pattern, even if we cannot explain why that pattern occurred" (37). This general causal pattern is a mechanism. To show that something is an instance of the pattern, or that the mechanism obtains, is to furnish a mechanistic explanation. Insofar as we are "[subsuming] an individual instance under a more general causal pattern" (37) we are still performing explanation in a way consistent with the foregoing requirements of causal explanation.

1a. Type B mechanisms and type A mechanisms

Elster (2007, 36) gives the following definition of mechanisms:

> Roughly speaking, mechanisms are frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences.

Examining this definition will allow us to fully understand the subtlety of Elster's account of mechanisms. Elster posits two kinds of mechanisms — type A mechanisms are the first kind— "those triggered under generally unknown conditions"— and type B mechanisms the second— those "with indeterminate consequences" (36). Examples illustrate the distinction more clearly. Type B mechanisms are generally pairs of opposed mechanisms: for instance, the "endowment effect" versus the "contrast effect." Both are mechanisms that account for instances of past experience impacting present welfare (Elster 2005, 55). With the endowment effect, "a memory of a good experience is a good memory," and the "experience of a bad [memory] is a bad memory" (56). When an endowment effect is at work, the lingering effects of a good experience will improve similar present experiences that we might have, and the effect of a bad experience will make similar present experiences worse. With the contrast effect, our memories of past experiences will have an opposing effect on our present experiences. A negative contrast effect occurs when our prior experience of something good leads us to "devalue less good" present experiences of a similar kind (56). Likewise, a positive contrast effect can lead us to regard a present experience more favorably than we might have otherwise, because we had previously had a highly negative experience of something similar. The present experience, then, seems more favorable by comparison than it might have without that highly negative prior experience.

Both the endowment effect and the contrast effect are potential ways by which past experiences could come to influence present welfare. They also both exhibit the essential feature of type B mechanisms — their consequences must be indeterminate. This indeterminacy can mean one of two things. The first is that we cannot tell, "in the case of several [type B] mechanisms [operating] simultaneously or successively, what their net effect will be" (51). The first kind of indeterminacy is particularly true if we consider net effect on welfare as indexed to particular times. A negative contrast effect, ensuing from a highly positive prior experience, might have a correspondingly negative net effect on later welfare, but the overall net effect — measured at a time point much later — might still be positive. The "positive utility from the [initial] experience itself [would offset] the negative net effect at the later times" (56), but we could not have expected this to result from a negative contrast experience alone.

This same example also illustrates the second way indeterminacy in the consequences of type B mechanisms might manifest. Were we to just look at the overall positive net effect, we would be unable to tell what mechanisms were responsible for it. We might assume that positive contrast effects or positive endowment effects were constitutive of the overall positive net effect, but we would be wrong.[1] As we have seen, negative contrast effects are just as likely to be responsible for an overall positive net effect. The consequences of type B mechanisms can thus be indeterminate in two ways. We might be unable to say what the net effect that ensues from our mechanism will be. Or, we might be unable, given a net effect value, to say with certainty what type B mechanism was responsible for it.

For type A mechanisms, indeterminacy exists at another point in the chain of causal events. Rather than having two, distinct, causal chains of events that affect a single independent variable in divergent ways, type A mechanisms account for situations in which different consequences ensue from the same initiating circumstances. The "compensation effect" versus the "spillover effect" provides an example (61). According to Tocqueville, "compensation" or "spillover" are two ways politics and religion might come to influence one another in a democratically organized society. The starting conditions from which both "compensation" and "spillover" effects are equally likely to ensue are those of a democratically organized society. The "spillover effect" describes situations in which an agent who "follows a particular pattern of behavior *P* in one sphere of life, *X*, will also follow *P* in sphere *Y*" (54). An agent accustomed to a democratic society, then, might, according to the "spillover effect," be disinclined to participate in organized religion. Religion might demand an obeisance to authority that the agent finds undesirable. The "compensation effect" describes an opposing phenomenon ensuing from the same starting conditions of a democratically organized society. The agent, instead of being made averse to organized religion by his experiences in a democratic civil society, comes to crave the authority offered by religion (63). The agent 'compensates' for the comparative dearth of authority in his political life by participating in religion.[2]

Seemingly true of both A and B type mechanisms are that they come in opposing pairs. Each mechanism corresponds to a complementary mechanism that describes an entirely opposing phenomenon. If true, however, then only one mechanism of an opposed pair could be at work at any given point in time; that is, the presence of a "spillover effect" must be to the exclusion of a concurrent "compensation effect". For type B mechanisms, at least, such is often not the case. Elster acknowledges a further distinction in the types of B type mechanisms available. While cases certainly remain where only one type B mechanism is at work, there are also type B_1 mechanisms, in which two opposing type B mechanisms are "triggered simultaneously by the same cause" (50). In addition, type B_2 mechanisms occur when one type B mechanism subsequently triggers its opposing, mechanistic complement (56). The existence of type B_1 and B_2 mechanisms offers the possibility of another kind of indeterminacy specific to type B

---

[1] It would be convenient to visualize the indeterminate consequences of these B type mechanisms, and B type mechanisms in general, in the following way:
     (i)   endowment effect → indeterminate effect on overall net effect
     (ii) contrast effect → indeterminate effect on overall net effect
[2] Also better visualized as:
     (i)   democratically organized society → spillover effect
     (ii) democratically organized society → compensation effect
where what triggers one type A mechanism over another is unknown and unpredictable.

mechanisms, and an elaboration on the net effect indeterminacies discussed earlier: we can never discern from a combined net effect which B_1 or B_2 mechanisms were responsible for that net effect.

This indeterminacy of type B_1 and B_2 mechanisms is in a direction opposite to what we had earlier discussed, but it remains consistent with at least the second kind of indeterminacy in single type B mechanisms; that is, our being unable to predict what net effect ensues from a single type B mechanism. Our inability to assess the net effect that ensues from a type B mechanism is an indeterminacy in the causal direction of – type B mechanism → net effect. The kinds of indeterminacy that generally emerge with type B_1 and B_2 mechanisms, however, are in the direction of — net effect → type B mechanism. While it is true that we still cannot assess the net effect that ensues from pairs of opposing type B_1 or B_2 mechanisms – such as the net effect that ensues when an endowment effect triggers a contrast effect, or when they are triggered simultaneously — this is compatible with the new kind of indeterminacy identified in B_1 or B_2 mechanisms. If we cannot know whether some endowment effect/contrast effect pair will bring about a positive net effect rather than a negative one, we cannot work backwards, deducing which antecedent B type mechanisms were responsible for that net effect. Being uncertain of the contribution a given mechanism will make prevents us from knowing which mechanisms contributed to a particular net effect. Even if we had both a net effect value and a set of mechanisms we knew had contributed to that net effect, we would still be unable to tell what magnitude of contribution was made by each mechanism.

Ambiguities about the magnitude of mechanistic contribution are not just a matter of proportion; that is, determining what proportion of the overall makeup of the net effect is due to which mechanism—ambiguities about magnitude also fall back into problems of attributing the correct mechanisms at all. Indeterminacy about the proportion of net effect attributable to each mechanism is very much related to indeterminacy about which type B mechanism is at work. If we could know that some type B mechanism consistently had a negative or positive contribution to net effect, then we would be better positioned to discern that an overwhelmingly negative net effect was probably the contribution of that type B mechanism known to produce negative net effects. Since we cannot even predict what sorts of contribution to net effect type B mechanisms will make, we certainly cannot reliably work backwards, deducing culpable mechanisms from overall net effect. This indeterminacy of contributing mechanisms when we look at net effect is the other major sort of indeterminacy that characterizes type B mechanisms.

Although thinking that mechanisms always come in opposing pairs might be more elegant, we must keep in mind that this often will not be the case. For instance, the ways in which past experience can influence present welfare are not just restricted to endowment and contrast effects. Certainly cases exist where neither effect is triggered, and additionally, cases complicated by very ordinary features like the valuation we attach to expected, or future, experiences. In these latter types of cases, my prior experience — that of having had an excellent French meal — might be expected to engender a contrast effect when I partake of my present, inferior French meal. If I intend to intermittently fast in the coming days though, that anticipated contrast effect might be altogether absent. Or, the anticipated contrast effect might manifest in a novel way – as an endowment effect, for example, in which I find my present French meal gratifying, but in virtue of not just my prior experiences (which allows me to appreciate aesthetic or gustatory features of my current meal to which I had been inattentive previously) but also in virtue of the

expected experience of fasting. This endowment effect would probably be of a different kind, one that took into consideration future experiences as well, but the point remains the same. B type mechanisms need not be modeled on the schema, 'if ϕ_1 or ϕ_2 (where ϕ_1 and ϕ_2 are opposing type B mechanisms) then sometimes z.' Or, likewise, for type A mechanisms, that 'if z then sometimes ψ_1 or ψ_2 (where ψ_1 and ψ_2 are opposing type A mechanisms).' Rather, it could very well be the case that we have more than two candidate type B mechanisms—'if ϕ_1, ϕ_2, ϕ_3... ϕ_n then sometimes z'—or more than two candidate type A mechanisms—'if z then sometimes ψ_1, ψ_2, ψ_3... ψ_n.' (2005, 49).

Lastly, type A and type B mechanisms closely interact. If I start with an exceedingly negative experience, without any further situational markers that would favor the issuance of a particular mechanism, I might re-describe the experience to myself more positively. I may engage, then, in wishful thinking — a type A mechanism. I might decide that the over overall utility I gained from the negative experience was in fact more valuable (even if that is not in fact the case). But doing so would impact any neighboring endowment effect – a type B mechanism — that would have ensued from the negative experience. Rather than that initial negative experience casting a pall on future similar experiences—I would instead have a positive endowment effect take place, in which my upvaluing the prior experience in turn upvalues present similar experiences.

2. Laws vs. mechanisms

Elster distinguishes his mechanisms from laws. Laws, according to Elster, "are general propositions that allow us to infer the truth of one statement at one time from the truth of another statement at some earlier time" (2007, 8). Laws are explanatory generalizations considerably stronger than mechanisms. Explanation in terms of law entails "[citing] a set of initial conditions together with a statement to the effect that whenever those conditions obtain an event of that type follows" (32). In order for explanation in terms of law to proceed, there cannot be any of the indeterminacies present and required by mechanistic explanation.

There are, then, at least two ostensible differences between laws and mechanisms. The first is that the starting conditions for a law are generally known — rather than "If conditions C_1, C_2...C_n obtain, then *sometimes* explanation E," as would be the case for a mechanism, we have "If conditions C_1, C_2...C_n obtain, then *always* explanation E" (2005, 48). This distinction is perhaps most pronounced with type A mechanisms, which are triggered under generally unknown conditions. Exceptions to laws are certainly recognized on Elster's account, but as they figure into my response to Persson, I will address these later.

The second difference, which concerns the certainty with which a mechanism will occur, is another way of looking at the first difference. Since we already established that identifying when a mechanism has occurred is to furnish a mechanistic explanation, this second difference is also about the certainty with which we know that a particular mechanistic explanation applies. While the having the starting conditions C_1, C_2...C_n ensure that explanation E will be the case when the situation is law-governed, a mechanism will only sometimes occur, even when starting conditions known to generally give rise to it are in place. This principle is well-illustrated by type A mechanisms. Having the conditions under which a mechanism is known to generally occur does not give us certainty about whether a particular mechanism will arise or even which mechanism will arise. Since starting conditions for a particular phenomenon are usually, by

definition, sufficient conditions for the phenomenon's occurrence, the unknowability of a mechanism's starting conditions, and the uncertainty of a mechanism occurring are related features of mechanisms.

When we *do* manage to "identify the triggering conditions" of a type A mechanism, or resolve the contribution of each type B mechanism to the otherwise mechanistically indeterminate net effect, according to Elster, our mechanism is "replaced by a law, albeit usually a weak one" (2007, 44). Initially, this claim seems overly ambitious and perhaps an unpleasant, unavoidable commitment that follows from how Elster has set up his theory of mechanisms. By distinguishing them from laws on the sole basis of their indeterminacy, resolving the indeterminacy of a type B or type A mechanism must turn that mechanism into a law. The distinction Elster draws between laws and mechanisms on the basis of indeterminacy, however, is not merely a stipulated difference. This distinction serves as an important recognition of the unusual explanatory role occupied by mechanisms — one that exists a step beneath the rigor of laws, but that still furnishes causal explanation.

The indeterminacies of type A and B mechanisms are not just due to epistemic limitations. Although situations exist in which we are forced to use mechanisms as explanatory placeholders until more information can be had, there are also situations where indeterminacy seems genuinely unavoidable. Nancy Cartwright's example of camellias is particularly helpful here (1983, 51-52).

Cartwright explains that while camellias are known to need rich soil, their roots are also averse to high temperatures. So when she planted her camellias in a bed of composted manure, and only a fraction of the camellias originally planted survived despite perfect care, she knew that the others had died from heat. But since all camellias are benefitted by rich soil and averse to heat, it would have been impossible to predict, at the outset, which camellias would survive, and which would die. This situation illustrates what necessitates a type A mechanism since the triggering conditions for camellia death rather than camellia survival are indeterminate:

> (i) Camellia plant + manure + heat $\rightarrow$ (type A mechanism) $\rightarrow$ camellia plant dies
> (ii) Camellia plant + manure + heat $\rightarrow$ (type A mechanism) $\rightarrow$ camellia plant lives

But no amount of biological scrutiny is going to give us the ability to predict whether a given camellia lives or dies under the initial conditions. The indeterminacy in triggering conditions essential to type A mechanisms will inevitably remain. We thus should not consider indeterminacy in an explanatory mechanism uniquely intended to accommodate indeterminacy as a defect of the explanatory device, or as something that awaits clarification. Sometimes clarification is not forthcoming and indeterminacy in an explanatory device is valuable.

I will admit, in keeping with Daniel Little (2012, 3), that the indeterminacy of a given Elster mechanism will sometimes be a matter of when, and how, we are looking at the situation that warrants mechanistic explanation. Sometimes mechanistic indeterminacy will be due to epistemic limitations. As soon as we are furnished with more information, or able to look at the situation in hindsight, there may no longer be indeterminacy in the ways necessary to constitute type A or type B mechanisms. Still, in these cases, while we might lose what is necessary, by

definition, for a mechanism to be in place, we do not lose explanatory strength. Rather, mechanisms whose indeterminacies in specific situations can be accounted for become descriptive forms of causal explanation — namely, facts. This outcome holds true, however, only if we are dealing with applications of particular mechanisms, instead of attempting to resolve the indeterminacy of *the mechanism itself*.

Once we have resolved the indeterminacy of a particular mechanism in a particular situation — that is, an *instance or application* of the mechanism — we can construct a causal narrative about what has happened. I think an equivocation between resolving the indeterminacy of an *instance* of a mechanism, rather than *the mechanism itself*, is present in Persson's criticisms of Elster. While Persson believes, throughout the counterexamples he presents, that he has resolved the indeterminacy in *the mechanism itself,* what he has actually done is accounted for the indeterminacy of an *application* of a mechanism.

3. Persson's argument

Persson's overall objection to Elster's account of mechanisms can be reconstructed from the two case studies used as vehicles for his argument. Each case study approaches Persson's objection differently. The first case can be seen as an attack on the concept of type B mechanisms. The second case an attack on type A mechanisms. Both cases, however, ultimately conclude that Elster's mechanisms must, by definition, lose their explanatory force as mechanisms whenever we resolve their respective indeterminacies. As soon as we resolve these indeterminacies, they cease to be mechanisms. For type B mechanisms, resolution occurs when we are able to identify the exact contribution of each distinct, causal chain of events to the eventual net effect. Previously, Elster had stipulated that type B mechanisms produce resulting net effects in which the mechanisms at work could not be discerned.

Resolving the indeterminacy of a type A mechanism involves determining the conditions that trigger it. The second half of Persson's argument concerns the non-exportability of type A mechanisms for which triggering conditions have been resolved. This argument allows him to conclude that mechanisms whose indeterminacies have been resolved remain explanatorily inferior to laws; these mechanisms lack the feature of generalizability—applicability to other, similar, contexts—that laws must have.

In his first case study, Persson takes issue with Elster's interpretation of the proverb, "the best swimmers drown". In effect, the proverb claims that the superior capabilities of swimmers who are "best" makes them more susceptible to drowning. However, if we take the proverb to express a mechanism, it becomes considerably more meaningful. While Elster never specifies the type of mechanism at work, Persson argues that "the best swimmers drown" would, on Elster's account, constitute a type B mechanism (2007, 10). That is, in order to furnish a mechanistic explanation for *why* the best swimmers drown, while avoiding the implausible *prima facie* explanation that the best swimmers drown because of their skill, there must be another byproduct of being well-trained in swimming that is not just the cultivation of a swimming skill — that of overconfidence (27). According to Persson, we thus begin with two distinct causal chains of events — the minimal number of initiating causal chains of events necessary for a type B Elster – one of skill-cultivation resulting from swim-training and the other of overconfidence (also ensuing from swim-training). Persson intends something like the following schema:

(a) Swim training $\rightarrow$ cultivation of skill $\rightarrow$ (intervening explanatory mechanism) $\rightarrow$ drowning

(b) Swim training $\rightarrow$ overconfidence $\rightarrow$ (intervening explanatory mechanism) $\rightarrow$ drowning

Even though we start with two different results of training for swimming — two causal chains — they both contribute to the same overall net effect: the drowning of swimmers who have been "best" trained. This net effect to which each causal chain contributes is indeterminate in exactly the way that it must be for a type B mechanism. By looking at the net effect of drowned, trained swimmers alone, we would be unable to tell which swimmers died in genuine accidents and which died for reasons relating to their overconfidence.

While I intend to call Persson's reading of "the best swimmers drown" as a type B mechanism into question, there are good reasons for being persuaded by Persson's interpretation. While the starting conditions of each causal chain might not be truly opposed to the other — swimmers who cultivate skill from their training do not necessarily do so to the exclusion of being overconfident — this is an acceptable situation for a type B mechanism. As we recall from Elster's initial characterization of type B mechanisms, there are also B_1 and B_2 subspecies of type B mechanisms. B_1 mechanisms would allow for situations in which overconfidence and skill were coextensive in a swimmer, and B_2 mechanisms situations in which skill in swimming entailed overconfidence.

Another reason exists for believing that the initial distinction between skillful swimmers and overconfident swimmers is justifiably drawn. On Persson's own account of overconfidence, there is no such thing as an overconfident, skilled swimmer. A skilled swimmer could have confidence — a consistently high subjective prediction of his own success rate — but since his predictions would be matched with a comparable swimming skill, the deficits in skill by which overconfidence is characterized would never come about. I will get to this later. For the time being, Persson uses this interpretation of "the best swimmers drown" as a type B mechanism as a starting point for his argument against Elster — an argument that will be next furnished. However, even at this early state of the argument one might disagree with Persson.

Dislodging the first premise on which Persson's objection to Elster rests – that is, Persson's reading of "the best swimmers drown" as a type B mechanism — will help in dismantling Persson's overall argument against type B mechanisms. After all, Elster never settles on whether "the best swimmers drown" describes a type A or a type B mechanism (39). Contra Persson, we can also interpret "the best swimmers drown" as a type A mechanism:

(1) Best swimmers $\rightarrow$ (intervening explanatory mechanism) $\rightarrow$ drowning

(2) Best swimmers $\rightarrow$ (intervening explanatory mechanism) $\rightarrow$ not drowning

Assuming that we are again looking beyond the *prima facie* absurdity of "the best swimmers drown", the proverb, on this alternative interpretation, states that only *some* of the best swimmers drown. Rather than starting out with two distinct causal chains, then, we have a single set of starting conditions — swimmers with the degree of swimming training necessary to constitute "bestness" — from which two opposing consequences ensue: one in which a subset of those best

swimmers drown, and another in which they do not drown. Under this interpretation mechanisms furnish explanations for why both drowning, and not drowning, are potential consequences of the same starting conditions. This is consistent with the role of type A mechanisms as things that indeterminately ensue from the same set of starting conditions, as well as things that, in hindsight, explanatorily account for why each determinate consequence does ensue. Having shown that the adage "the best swimmers drown," is compatible with a reading in terms of type A mechanisms — and keeping this alternate interpretation in mind — I can now give the rest of Persson's argument.

After identifying "the best swimmers drown" as an instance of a type B mechanism, Persson proceeds to exploit the vulnerability identified earlier: that whenever the indeterminacies of a mechanism are resolved, that mechanism must, by Elster's own lights, cease to be a mechanism and become a law. Since type B mechanisms account for situations in which distinct sets of starting conditions converge upon an indeterminate net effect, all Persson must show is that a single causal chain *can* in fact be traced from start to finish. Persson must show that we *can* pick out the contributions made by distinct type B mechanisms — and the phenomena they account for — from the resulting net effect, by identifying a causal chain of events that remains continuous and distinct from its starting condition beginnings, to its eventuation in that net effect. Once Persson has done this, he can claim either that (1) the determinate contributions by type B mechanisms that he has identified in the resulting net effect are actually such disparate aspects of the case at hand that assessing a net effect is in fact inappropriate,[3] or (2) the more devastating contention that Elster is committed to the existence of a law in place of any mechanism whose indeterminacies have been elucidated. Persson is most interested in the second contention.

A mechanism whose indeterminacies have been merely elucidated will still be short of the explanatory strength of a law. Since the triggering conditions or consequences that have been identified will be, at best, local, the conjunction of a mechanism with either the conditions from which it is known to ensue, or the consequences that result from it, will still lack the (i) wide scope and (ii) degree of invariance under change required of laws (Woodward 2002, 369). A mechanism for which we know just the triggering conditions or consequences is unlikely to be exportable to other contexts. The mechanistic relationship we have revealed will generally obtain in a limited number of contexts. Even though we know, for instance, that during periods of political upheaval, internal struggles for power will often be present in a revolutionary movement or faction — where a "competition for power" is the explanatory mechanism that fits between starting conditions of an ideologically fragmented revolutionary group, and the emergence of a particular revolutionary group and ideology on top — this phenomenon is not generalizable to all

---

[3] Imagine that the "endowment" and "contrast" effects from earlier weren't both mechanisms of past effect on present welfare in quite the same way; as it stands, neither the endowment nor the contrast effect are going to arise when the present experiences are categorically different from the relevant past experiences. Elster argues, for instance, that the excellent French meal I had earlier might lead me to "devalue" subsequent, inferior French meals I take (contrast effect), but have no impact on the subsequent Chinese meals I might take (2005, 57). If I assess the net effect of total endowment effects ensuing from an excellent French meal I had, it would be inappropriate — indeed, a category mistake — if that net effect included contributions from subsequent French and Chinese meals I had. This is an argumentative route available to Persson, but one that he does not undertake. Not much more will be said about this in the paper.

instances of political upheaval (McAdam 2001, 68).[4] Sometimes revolutionary movements that are initially fragmented along ideological lines will unite for the sake of a common revolutionary interest.[5]

Furthermore, these more fine-grained mechanisms will be inflexible in a way that laws are not. A feature common to both laws and mechanisms is that they can both admit of counterfactual situations. We can substitute different values for the component variables of a mechanism or law, for instance, and still expect the overall mechanism or law to obtain. That is to say, we expect the causal relationship posited by a law or mechanism to withstand a certain amount of change. This quality of laws or mechanisms will be referred to as "invariance" under change, and the situations involving modified variables under which a particular causal relationship is still expected to obtain will be called "counterfactual" situations (Woodward 2002, 370). When a law or mechanism is invariant, then it "supports", or admits of counterfactuals (369). Although invariance and the admittance of counterfactuals is commonly seen in, and demonstrated by, scientific laws and mechanisms, they are qualities present in social science mechanisms and laws too. The compensation mechanism that Tocqueville identifies in democratic societies, for instance, also manifests — though with different effects — during times of revolution. When repression, the "[suppression] of either contentious acts or groups and organizations responsible for them" is coupled with a compensation effect, "generally stiffening resistance on the part of threatened communities" ensues (McAdam 2001, 69). This can be summarized by the following:

(i) repression → compensation effect → further radicalization

The compensation effect, which accounts for why triggering conditions of repression often result in situations of further radicalization (being constrained in the modes of contention available to them, ideological groups will find other outlets for expression) has held true in a number of circumstances. The compensation effect could be observed, for example, during the formation of the Black Panthers. The Black Panthers came about, in no small part, because of the increasing threat of a white police force. To lend another example, during the French Revolution, Western French peasants, facing a military draft, joined the "clerical-legitimist resistance" (69). Despite differences in geography, culture, and historical context, the compensation effect was present in these, and a number of other revolutionary events. As a mechanism, the compensation effect can withstand tweaking of the very component elements necessary to secure its occurrence in the first place. Indeed, the very fact that the compensation effect holds in both democratic societies, as Tocqueville had intended it, and in revolutionary periods, is itself a demonstration of invariance by the mechanism.

The compensation effect, however, remains merely a mechanism. This means that it will be invariant over a "limited spatiotemporal interval," and admit a very finite number of

---

[4] McAdam, Tarrow, and Tilly give the specific example of the more "moderate branch" of the American Civil Rights Movement and its leadership under Martin Luther King Jr. More radical groups during the Civil Rights Movement such as SNCC and CORE opposed the perceived conciliatory attitudes of the moderate branches.

[5] I am thinking of the French Revolution, in particular, during which the commoners, nobility, and clergy were all united in forming the Third Estate. Although each subgroup had specific interests — the clergy in the preservation of a class of religious employees in the form of civil servants, the nobility and commoners in property redistribution, etc. — each of their interests was thought to be best achieved through the deposal of King Louis XVI. They were thus able to temporarily set aside ideological differences.

counterfactuals (Woodward 2002, 371). Not all ideological groups in situations of political upheaval, for instance, will respond to repression via further radicalization. Sometimes such groups will succumb to repression rather than respond with a compensation effect. Laws, on the other hand, admit of far more counterfactuals than mechanisms, and therefore exhibit significantly more invariance. There will be very few exceptions — counterfactual situations — to a law; for the most part, the causal relationship posited by a law will hold under the relevant conditions.

Wide scope and invariance are both markers of explanatory strength. And in fact, wide scope and invariance are deeply related explanatory virtues. When a generalization is widely applicable, that means that it obtains in a variety of different situations; it means that it admits a significant number of counterfactual situations. And when a generalization is "stable under some changes and interventions" that means that it is invariant in those situations (371). Because mechanisms lack the explanatory strength of laws in these two important respects of wide scope and invariance — or, rather, the comparative explanatory weaknesses of mechanisms can be measured by their comparative deficiencies in these two respects — Persson hopes to show that even the more fine-grained explanation that results from resolving the indeterminacies of a particular mechanism — that is, when we know its triggering conditions or identify its consequences — is still short of the explanatory strength of a law.

Persson argues that our distinct starting conditions of (a) and (b) types of swim-training — assuming that we have agreed to parse "the best swimmers drown" as a type B mechanism — and their respective results of (a) cultivating skill at swimming or (b) cultivating overconfidence at swimming, each correspond to a different type of risk. On the basis of these distinct kinds of risk leading up to drowning, we can divide the otherwise indeterminate net effect of drowning in general, into different kinds of drowning ensuing from different kinds of risk. Elster, according to Persson, has overlooked the distinction between drowning resulting from "outcome risk" and drowning resulting from "epistemic risk."

"Outcome risks" inevitably accompany any "well-defined event" (Persson 2007, 14). Even a well-trained swimmer who has cultivated genuine skill, instead of simple overconfidence, will be susceptible to outcome risks – the possibility of being overtaken by an unexpectedly strong tide, or that of being knocked unconscious and falling from a boat. "Epistemic risks," by contrast, are risks that result from the deliberative processes of an agent (14). Such risks may be foreseen, or unforeseen. A swimmer might decide that his training, which prepares him exclusively for indoor swimming, inadequately prepares him to tackle swimming in the ocean. His self-perceived preparedness for swimming in the ocean is an epistemic risk consideration. Another swimmer with the same primarily indoor swimming experience might fail to recognize the same risk. He might fail to recognize the difference, and increased difficulty, of swimming in the ocean and believe that his training in fact adequately prepares him for ocean swimming.

From this example, we see that epistemic risk is related to "epistemic reliability" — the actual probability distributions that inhere in events (21). Assessments of epistemic risk consist in the probabilities in certain events that are taken seriously, notwithstanding of the actual probability distributions themselves. The risk of lethality in swimming in the ocean untrained might be equally great for both aforementioned swimmers, but only the first swimmer takes the risk

seriously. The second swimmer, in his overconfidence, will either suppress consideration of the risk, or assign an unrealistically low likelihood to its occurrence.

Now, according to Persson, we also have a definition for overconfidence — that which is speculated to ensue as a secondary consequence of swim training, aside from skill — in terms of epistemic risk and epistemic reliability. Overconfidence is the "mean subjective probability" — the assessment of epistemic reliability in terms of epistemic risk — minus the actual probability distribution. The overconfidence of a swimmer can be thus measured by the difference between his perceived risk in swimming in the ocean and the actual risk of his swimming in the ocean. This can be visualized by taking x as the subjective probability of success — the probability of success predicted by the swimmer, that is, his perceived risk — and c as the actual rate of success. Subtracting c from x will indicate that a swimmer is overconfident if $x-c > 0$, and under confident if $x-c < 0$ (16).

From this account of overconfidence, we can now also see how Persson is maintains that swimmers who cultivate skill, and swimmers who cultivate overconfidence instead, are genuinely opposed causal chains. Persson defines overconfidence as the sort of thing that genuinely skilled swimmers could never have. A confident, but skilled swimmer would estimate that his "subjective probability" of success in a given body of water was high, but since his actual skill would be sufficient to secure his success in swimming, the difference of $x-c = 0$ (16).

Not only does Persson posit two distinct starting conditions — one in which swim training cultivates genuine swimming skill and the other in which swim training results in overconfidence — but he also traces each of these starting conditions to a determinate result, and fills in a unique explanatory mechanism for each.

> (a*) Swim training → cultivation of skill → "[exposure] to outcome risk" → drowning
> (b*) Swim training → overconfidence → agent decision-making based on "monitored outcome risk and [ensuing] epistemic risk" judgment → drowning
> (27)

In (b*), swimmers who grow overconfident from their training are more susceptible to errors in their assessment of epistemic risk. As such, their drowning deaths are explanatorily accounted for by some mechanism of erroneous epistemic risk assessment ensuing from overconfidence. Swimmers who do not cultivate overconfidence from their training, but who nevertheless drown — the causal chain (a*) — have their deaths differently accounted for by outcome risk — those risks, out of their control, to which they would be inevitably exposed by swimming.

Elster might counter that we still cannot tell what will become of each set of swimmers. Even if we start with the two prospective type B mechanisms of swimming overconfidence and the cultivation of genuine swimming skill, we cannot predict what the net effect of either will be – the overconfident swimmers are as likely to drown as they are to remain unscathed. Persson claims that we can. Given the different kinds of risk he has identified, Persson argues that we can distinguish the deaths of best swimmers that result from outcome risks, from deaths that result from epistemic risk. This approach resolves one of the two characteristic indeterminacies of type

B mechanisms described earlier. We can, according to Persson, make the indeterminate consequences of type B mechanisms determinate and known – we can know the net effect that will ensue from a given mechanism. Best swimmers who have cultivated skill rather than overconfidence will be more likely to drown for reasons related to outcome risk exposure. Overconfident best swimmers will be susceptible to drowning deaths related to erroneous assessments of epistemic risk.

Knowing the outcomes that ensue from type B mechanisms also resolves the second kind of indeterminacy by which type B mechanisms are characterized: our inability to discern from the net effect alone which mechanisms were antecedently responsible for that net effect. According to Persson, were we to be given the net effect of a group of drowned best swimmers, we would simply look for the subgroup of swimmers exposed to outcome risk, and segregate them from the subgroup of swimmers who died of circumstances related to poor judgments of epistemic risk. This strategy also allows us to determine the exact contribution made by each antecedent mechanism of overconfidence-cultivation and swimming-skill-cultivation. It is trivially the case that the proportion of total drowned best swimmers who died of causes relating to overconfidence will be simply those who died of causes related to overconfidence; likewise for those swimmers who were genuinely skilled.

I am not certain how Persson is supposed to avoid the objection that his solutions to mechanistic indeterminacy inappropriately exploit the advantage of having hindsight knowledge (as discussed earlier). What he fills in between the triggering conditions of "best swimmers" and their drowning, in order to trace a causal narrative, is knowledge that can only be had with hindsight. Persson seems to ignore the distinction between resolving the indeterminacy of a mechanistic *application,* and resolving the indeterminacy of the *mechanism itself*. As we recall, for those mechanistic indeterminacies that can be cleared up with knowledge we gain over time, in hindsight — indeterminacies that result from epistemic limitations — what we are usually dealing with is a mere application of a mechanism, rather than the mechanism itself. In filling in the indeterminacy for an instance, or application, of a mechanism, we still gain explanatory strength even when the mechanism ceases to be a mechanism. I will elaborate on this point in the next section by distinguishing the different ways in which increased explanatory strength can be achieved – distinctions that Persson ostensibly fails to recognize.

We *can* fill in the indeterminacies of *mechanisms themselves* with knowledge garnered from hindsight; that is, find general features that triggering conditions or determinate consequences must have in order for their concomitant mechanisms to be elicited. But Persson's case study does not address this approach. He gives us strictly local instances of resolved mechanistic indeterminacy. If we expect to resolve indeterminacies of *mechanisms themselves*, we need to have many more situations in which we find the triggering conditions or determinate consequences coextensive with the mechanism of interest. A single, local instance – in which a causal relationship between the determinate triggering condition or determinate consequence and the mechanism of interest is seen to obtain – appears insufficient. When we can resolve the indeterminacies of mechanisms themselves, we will have something much closer to a genuine explanatory law than the mere elucidated *applications* of mechanisms that Persson attempts to pass off as covering laws.

Persson's analysis of the swimmer's proverb as a type B mechanism furnishes the first half of his argument. For Persson, we can, in fact, clarify mechanistic indeterminacies by filling in what happens between either the indeterminate starting conditions and mechanism triggered (type A) or between the mechanism and its consequences (type B).

The second half of Persson's argument, given against type A mechanisms, is intended to strengthen his preceding case. While the case study of overconfident swimmers shows that clarifying the indeterminacies of mechanisms is possible – and that by Elster's own lights, these couplets of triggering conditions/mechanism or mechanism/determinate consequences can no longer be mechanisms – the second half of Persson's argument attempts to show that though these more explanatorily powerful, elucidated mechanisms must be laws, they cannot be those either.

Persson gives two main examples to this end. The first is that of "terror management theory" (TMT) (Persson 2012, 111). TMT addresses triggering conditions in which a population is "[reminded] of their own [mortality]," and hypothesizes a number of mechanisms that can conceivably ensue from those starting conditions (111). According to Persson, social scientists seem content with the idea that TMT is inapplicable to contexts other than that of the United States, or Western contexts in general. Persson's second example of an elucidated mechanism that still falls short of law-like explanatory strength is the "monetarist thesis." The monetarist thesis argues that inflation — as the "rate of growth of consumer prices"— is a "function of growth rates in money supply, real income, and the expected cost of holding money" (111). The monetarist thesis has been found inapplicable to "developing countries."

What Persson suggests by these examples is that even when triggering conditions of particular mechanisms are known in the social sciences — or, rather, when candidate triggering conditions have been eliminated — these elucidated mechanisms are regularly non-exportable.[6] Such is equally true of type B mechanisms. Mechanisms that have been elucidated in this way fail to meet either criterion of wide scope or invariance. They thus cannot be laws. Since Elster's earlier statements ostensibly commit him to the notion that mechanisms for which the triggering conditions are revealed come to be replaced by "weak laws," Persson concludes that Elster is on the horns of a dilemma. Mechanisms for which we have resolved the indeterminacies are no longer, by definition, mechanisms. They are also too weak to be laws. In the process of achieving more fine-grained explanations by resolving the situational indeterminacies of mechanisms, then, we lose explanatory strength.

4. A defense of Elster

Now I pick up the alternative reading of Persson's B type mechanism case study begun earlier. If Persson's rejection of Elster-type-mechanisms is in part, or even solely, dependent on his analysis of drowning, well-trained swimmers as type B mechanisms, and his type B mechanistic reading is erroneous, then we have a starting point by which to jettison Persson's critique of Elster. My initial response to Persson, like my overall defense of Elster's mechanisms, will be

---

[6] I have my doubts that merely eliminating candidate triggering conditions is in fact quite the same as discovering consistent triggering conditions—which seems to be the level of rigor necessary for Persson's argument to go through.

based on the argument that even when we have ostensibly disambiguated the indeterminacies in a mechanism, we have not really done so.

Persson argues that at least for the type B mechanisms he works with, once we have resolved the indeterminacies, we have something that is by definition not a mechanism, but still less than a law. I will argue that for each of Persson's case studies, enough indeterminacies of the right kind remain to still warrant calling the explanatory devices mechanisms. I do not want to purposely introduce ambiguity, or hinder the possibility of more fine-grained explanation; rather, Persson has not given us enough reason to think that we have something more powerful than mechanistic explanation, let alone law-governed explanation.

If we read "the best swimmers drown" as a type A mechanism and plug in Persson's mechanistic explanation in terms of risk, we get the following:

> (1) Best swimmers → whose swim training gives rise to overconfident swimmers whose epistemic risk judgments are flawed→ drowning
> (2) Best swimmers → whose swim training gives rise to skilled swimmers who are exposed to outcome risk → not drowning

This formulation reveals the persisting indeterminacy that Persson's reading of "the best swimmers drown" as a type B mechanism conceals. If "best swimmers" are still those who have mastered a certain set of swimming skills, then there must be a type A mechanism present that accounts for why mastery of the same skill set — with presumably some shared curriculum by which that skill set was picked up — would produce overconfidence in some, and genuine skill in others. It would have to be a type A mechanism that accounted for this indeterminacy, because once again, we have a situation in which it is impossible to tell, given a set of starting conditions, what elicits one result rather than another.

Persson might claim that he is not explanatorily obligated to account for the type A mechanism at work because those events irrelevantly precede what he is interested in — the immediate events that give rise to best swimmers drowning. As such, he can cordon off two groups of best swimmers at the very outset — those whose training has lead them to cultivate genuine skill, and those whose training has resulted in their overconfidence. This move presupposes the indeterminacies of a type A mechanistic analysis.

The problem with this reading, however, is that if Persson merely assumes that there is a group of "best swimmers" whose training has resulted more significantly in their overconfidence than in their cultivation of genuine skill, then that subgroup of overconfident swimmers cannot be considered best — especially if our metric of "best" is one of general swimming skill. With a type A mechanistic analysis, we can show how "best swimmers" might cease to be best swimmers by becoming overconfident swimmers. In failing to first sort out how "best swimmers" might become overconfident swimmers through type A mechanistic analysis, we can only assume that both of Persson's initial categories of genuinely skilled swimmers, and overconfident swimmers, must still be considered "best swimmers" under the original conditions of the proverb.

If Persson means something else, perhaps that there are swimmers who are coextensively overconfident *and* skilled, he is going to have to (1) account for how this can be, given his definition of overconfidence and skill as mutually exclusive, and (2) in the process give us some sort of causal etiology for how overconfidence and skill can be coextensively cultivated. I suspect that (2) would result in a third possibility branching off of the two existing mechanistic possibilities under the type A mechanistic reading. Persson must either revise his interpretation of the proverb — potentially threatening his current analysis, which allows him to allegedly clarify type B mechanistic indeterminacies — or he must concede that there remain type A mechanistic indeterminacies in even his type B reading of the proverb. Either of these is necessary for him to maintain internal consistency in his analysis.

Persson's failure to elucidate the preceding type A mechanistic indeterminacies also comes to be problematic in his effort to clarify type B mechanisms. It is unclear how Persson would classify swimmers who were overconfident, by his own lights, but who die of exposure to outcome risks unrelated to the overconfident decisions they made. An overconfident swimmer can think himself perfectly capable of swimming in the ocean when he is in fact not, but never intend to actually step off his boat. This overconfident swimmer can nevertheless drown from being struck by the boom that attaches to his vessel's mast. Would this drowning death be considered a consequence of exposure to outcome risk rather than epistemic risk? The answer is further muddled if one argues, quite validly, that our overconfident swimmer would have never found himself in the boating situation if he did not think himself capable of dealing with the accompanying outcome risks. The distinction between type A and type B mechanisms is flexible; type B mechanisms will often necessitate type A mechanistic clarification, or decompose into multiple type A mechanistic scenarios, and vice versa. As such, there is rarely ever a hard and fast mechanistic explanation, or a single isolated mechanistic explanation in the way that Persson wants there to be.

While reading the proverb as a type A mechanism indicates that there remain crucial indeterminacies — enough to at least show that the proverb "the best swimmers drown" describes a type A mechanism with all of its accompanying ambiguities — the type B mechanistic reading that Persson favors can also be directly attacked. In order for something to be more explanatorily powerful than mechanistic explanation, it must fulfill the criteria for mechanistic explanation and more. As we recall, causal explanation, ideally, is articulated in terms of individuals and their actions — where articulation in this form is applicable or possible. This approach is how the most fine-grained kind of causal explanation is achieved. Even if we discount the type A mechanistic indeterminacies, and accept Persson's categories of 'best swimmers whose training promotes erroneous assessments of epistemic risk' and 'best swimmers whose training promotes reliable assessments of outcome risk,' any type B explanation can only be causally explanatory if it accounts for how an individual swimmer arrived at the specific overconfident decision subsequently responsible for his drowning.

What Persson must show us is that his disambiguated mechanisms can account for how an overconfident swimmer — and overconfident swimmers in general — made the move from accurately thinking himself capable of swimming in the environments in which he was trained, to erroneously thinking himself capable of tackling more formidable bodies of water. Persson must at the minimum furnish this in order to claim that his disambiguated mechanisms offer more fine-grained explanation than ordinary mechanisms. He must do this in order to show that

Elster, through his mechanisms, is committed to the existence of explanatory entities that are neither mechanisms nor laws.

Persson falls short of being able to give a mechanistic account of why best swimmers drown in terms of methodological individualism. The distinctions he makes between outcome and epistemic risk are productive advances in trying to build a causal narrative about why best swimmers drown, but they do not tell us anything about why a *particular* swimmer made an erroneous judgment of epistemic risk because of overconfidence, or why that swimmer cultivated overconfidence rather than skill.

But even assuming that he can, Persson is no closer to showing that he has ensnared Elster in the second respect: that Elster's disambiguated mechanisms, while no longer being mechanisms, also fail to meet the explanatory criteria for laws. Elster never claims that more fine-grained mechanistic explanations must be law-like, especially if these more fine-grained explanations are achieved by articulation in terms of methodological individualism. Intuitively, this makes sense. We do not come closer to positing laws by making our explanations more specific. Insofar as wide scope and invariance remain explanatory virtues of laws, then the explanatory strength of laws is achieved through generalization. What Persson shows is that resolving the indeterminacies of *applications* of particular mechanisms makes them more specifically explanatory. While a fine-grained explanation may be more powerfully explanatory, it may not exhibit improved explanatory strength in a *law-like way*.

What Elster must mean, then, when he says that a mechanism is "replaced by a law [albeit usually a weak one]" when we manage to identify their "triggering conditions" is not that we have managed to identify the triggering conditions of a mechanism for a particular context; rather, that we have managed to identify some consistent feature that triggering conditions must have in order for a particular mechanism to be brought about (Elster 2007, 44). Persson briefly considers the possibility that Elster could be talking about what generally triggers mechanisms rather than what locally does, but quickly dismisses the idea (Persson 2012, 112).

Persson conflates the types of explanatory strength available. Something can be explanatorily powerful because it is law-like: it accommodates a number of counterfactual situations and has relevance across a wide scope. On the other hand, something can be explanatorily powerful because it comprehensively accounts for a specific situation. Persson has overlooked this distinction in arguing that the augmented explanatory strength — in terms of specificity — mechanisms pick up whenever we resolve their local indeterminacies are short of the *generalized* explanatory strength of laws. The two kinds of explanatory strength are to be distinguished, and, as I see it, are fairly incommensurable. Explanatory strength that covers something specific and local — usually a sort of descriptive explanation — should not be compared to the explanatory strength of laws, which should be of wide scope and invariant.

Here, I will consolidate my earlier argument that Persson has conflated his resolution of the indeterminacy of *applications* of particular mechanisms with the resolution of indeterminacies in the *mechanisms themselves,* with the current argument that explanatory strength can be augmented in different, distinguishable ways. Generally, if the indeterminacies of a mechanism are cleared up in terms of methodological individualism, then we have succeeded in improving the explanatory strength in an *application* of a particular mechanism. This might turn the applied

mechanism into a fact, or description. As should be evident, if we succeed in resolving the indeterminacies of a mechanism *in general,* then we have improved explanatory strength in another capacity — we have made the mechanism more generally explanatory, and therefore more law-like. While this second possibility was what Persson was aiming for, he instead achieves the first possibility. This outcome does not allow Persson to argue that mechanisms elucidated locally, in the ways that he has shown, fall short of being laws.

Ultimately, I think Persson's criticisms reflect an inflexible reading of Elster's mechanisms. Elster's account does accommodate and acknowledge the distinctions between applications of mechanisms, rather than mechanisms themselves, and explanatory strength achieved through specificity as opposed to explanatory strength achieved through generalization. Only by overlooking and conflating these differences that Persson manages to argue his case.

Contact Details: kimberly.chuang@gmail.com

References

Cartwright, Nancy. 1983. *How the laws of physics lie*. Oxford: Oxford University Press.
Elster, Jon. 2005. "A plea for mechanisms" in *Social mechanisms: an analytical approach to social theory*. Cambridge: Cambridge University Press.
Elster, Jon. 2007. *Explaining social behavior: more nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
Little, Daniel. 2012. Social mechanisms and scientific realism discussion of, "Mechanistic Explanation in Social Contexts," by Johannes Persson. *Social Epistemology Review and Reply Collective* 1 (3): 1-5.
McAdam, Doug, Sidney Tarrow and Charles Tilly, eds. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.
Persson, Johannes. 2007. The best swimmers drown — Mechanisms and epistemic risks: A constructive critique of Elster. In *Hommage à Wlodek: 60 Philosophical papers dedicated to Wlodek Rabinowicz* Toni Rønnow-Rasmussen, Björn Petersson, Jonas Josefsson, Dan Egonsson, eds., 1-34. http://www.fil.lu.se/hommageawlodek/site/abstra.htm
Persson, Johannes. 2012. Mechanistic explanation in social contexts: Elster and the problem of local scientific growth. *Social Epistemology* 26 (1): 105-114.
Woodward, Jim. 2002. What is a mechanism? A counterfactual account. *Philosophy of Science* 69: S366-S377.