Optimality Theory and the Problem of Constraint Aggregation

1. The Issue Of recent changes in linguistic theory, the shift from derivations to constraints is among the most substantive. Central to this is Optimality Theory (OT; Prince & Smolenksy 1993), which – unlike, e.g., Chomsky & Lasnik's Filters or Pāṇini's Aṣṭādhyāyī – claims that structural well-formedness (grammaticality) is enforced by sharply conflicting, mutually incompatible constraints. As it is impossible to satisfy all constraints simultaneously, grammars must resolve conflicts to determine an input's 'surface representation', the form 'most harmonic' with the constraints. Prince & Smolensky, and most subsequent work, resolve conflicts by 'rank[ing] constraints in a *strict dominance hierarchy*'.

We question this position and develop a formal mathematical approach to constraint-conflict resolution. This facilitates a recharacterisation of OT's difficulties into problems for constraint-based grammar *per se* versus problems that are artefacts of conflict-resolution via strict dominance hierarchies. The formalism also shows that OT occupies a well-defined position in the space of logically possible constraint-based grammars and that modifications to OT (e.g., Sympathy Theory (McCarthy 1998); Boolean operations on constraints (Crowhurst & Hewitt 1999)) diverge from OT uniformly, on one formal 'dimension'. So, we may classify these modifications according to how greatly they increase the power of the theory. Further, we can determine what phenomena fall fully beyond the logical remit of constraint-based theories, and can provide a condition that a theory must satisfy in order to count as being constraint-based. This establishes two falsifiability criteria for Prince & Smolensky's claim that 'Universal Grammar consists largely of a set of constraints on representational well-formedness' and that grammaticality is a matter of structural well-formedness.

2. The Formal Methods of the paper are drawn from social choice theory, exploiting the formal similarity between the problems that it and OT address.

Committee decisions (e.g., selection of abstracts) or elections exemplify social-choice-theoretic problems. Such problems consist of two steps: to determine, for every voter and every candidate, a measure of that voter's (dis)preference for that candidate; then, using the measurements, to rank the candidates from 'best' to 'worst'. Similarly, in OT, Gen produces a candidate set for a given input, each member of which violates each constraint (voter) to differing degrees. These violations are used to determine the optimal candidate. (Selecting a single winner and ranking all candidates are mathematically equivalent.)

Social choice theory formalises three aspects of this process. *Measurability* concerns the significance attached to the preference measurements. Are they significant on an absolute scale, or is how they rank-order the candidates alone meaningful? *Comparability* asks if one can compare the preference measurements of different voters for the same or different candidates. Is it meaningful to ask if Voter_1 more strongly prefers Candidate_1 than Voter_2 does Candidate_2? *Aggregation* concerns how the overall ranking of candidates is determined given voters' preference measurements. Are all voters equal, or do some have priority?

In OT, measurement (of constraint violation) and aggregation (of violations to determine the actual output) are familiar, but comparability is less so. However, appropriate comparability assumptions appear necessary for dissolving several cases of opacity and we point out that some proposed modifications to OT deviate from Prince & Smolensky with regard to assumptions about comparability. So, the problems of OT and social choice theory are clearly parallel.

A key insight of social choice theory (which won K. Arrow the Nobel Prize) is that measurability and comparability assumptions strongly constrain the space of aggregation procedures. We argue that the concept of constraint-based grammar forces three minimal conditions on aggregation (cf. Arrow 1951):

Universal Domain The aggregation function is defined for all possible inputs.
Strong Pareto Principle For x, y candidate outputs, if x violates no constraint more than y, then y cannot be ranked higher than x. If, in addition, x violates one constraint less than y, then x is higher ranked than y (and x is *Pareto-superior* to y.)
Relevance Principle To rank any two candidate outputs, the aggregation function needs only the candidates' scores on all constraints; other candidates' are irrelevant.

Under varying measurability and comparability assumptions, we state theorems about the nature of aggregation in different constraint-based grammars. In particular, we show that OT makes the most minimal measurability and comparability assumptions, and that strict dominance hierachies are the only aggregation functions satisfying these assumptions and the three minimal conditions above. Changing assumptions, however, yields a larger class of aggregation functions, allowing us to explore the issues outlined in **1** and amplified below.

3. Applications and Advantages of Different Aggregation Procedures range from empirical and theoretical matters in phonology, to OT syntax and to the nature of economy of derivation in minimalist syntax (Chomsky 1995).

Not all OT-inspired work uses strict dominance hierarchies. Alternatives include weighted combinations of constraints (Flemming 1997, Burzio 1999), constraint co-ranking (Itô & Mester 1997), constraint conjunction (Smolensky 1995, Kirchner 1996, Crowhurst & Hewitt 1999). We show that in all these cases the nature of measurability and comparability is really at stake. Moreover, we debug Itô & Mester's account by making these underlying assumptions explicit.

On the basis of the Strong Pareto Principle, we distinguish two types of opacity. Pareto opacity, where the selected candidate is Pareto-superior to the real output, presents grave problems for most constraint-based phonologies, including OT. NonPareto opacity, where the selected candidate performs worse than the real output on at least one constraint, can in some instances be solved by modifying measurability and cross-constraint comparability assumptions (e.g., Kirchner 1996). We underline the usefulness of the concept of Pareto superiority by showing that it can often be used to solve cases of opacity (e.g., Tiberian segholates). However, some cases of NonPareto opacity are impervious to changes in assumptions: in Bedouin Arabic *a*-raising, the set of violation scores for the optimal candidate of one selection process can be too similar (in a technical sense) to the set of violation scores for a suboptimal candidate of another selection process. So, it impossible for any aggregation function using only violation scores to determine the right outcome in both cases and thus such cases are not accommodable within any constraint-based grammar.

Pure constraint-based grammars are committed to *evaluationism*, the claim that the constraint violation scores of any two candidates contain sufficient information to rank them in a global harmony ordering. Our formalism permits us to prove a theorem relating satisfaction of evaluationism to a simple criterion based on the linguistic correlates of Arrow's postulates. So, modifications of OT not satisfying this criterion violate evaluationism, the essence of constraint-based grammar. We show McCarthy's Sympathy Theory to suffer this deficiency.